Targeted Mutagenesis of Soybean β-amyrin Synthase Genes via the iPB-RNP Method and Characterization of the Soyasaponin-Deficient Mutants

（iPB-RNP 法を介したダイズ β-アミリン合成酵素遺伝子の標的変異誘発及びソヤサポニン欠損変異体の特徴付け）

Hokkaido University
Graduate School of Agriculture
Frontiers in Production Sciences
Unit of Agricultural Plant Science
Doctoral Course
Hiroki Asa

Contents

CHAPTER 1
General Introduction

Soyasaponins are triterpenoid saponins predominantly present in soybean (*Glycine max*) and other legumes such as peas (*Pisum sativum*) and lentils (*Lens culinaris*). Triterpenoid saponins have been extensively investigated because of their diverse pharmacological activities (Augustin et al. 2011; Moses et al. 2014; Guang et al. 2014; Yates et al. 2021). These phytocompounds, therefore, have been considered attractive candidates that offer medical effects with fewer safety concerns. Recently, the complete biosynthetic pathway of QS-21, a triterpenoid saponin from *Quillaja saponaria*, has been elucidated and its heterologous production in engineered yeast has been successfully achieved (Reed et al. 2023; Martin et al. 2024; Liu et al. 2024). QS-21 is a potent vaccine adjuvant that has been clinically approved for human use (Lövgren Bengtsson et al. 2011; Didierlaurent et al. 2017). These achievements underscore the high expectations for these phytocompounds in biological applications. Although the hemolytic and cytotoxic activities of soyasaponins are substantially lower than those of QS-21, they are capable of eliciting moderate immunogenic responses (Yates et al. 2021). Therefore, soyasaponins represent promising alternatives that overcome the safety concerns associated with the relatively high toxicity of QS-21. Furthermore, their anti-inflammatory, antiviral, and antimicrobial properties further suggest potential roles in reducing the risk of coinfection and secondary infections (Yates et al. 2021).

On the other hand, soyasaponins are also responsible for the astringent aftertaste, one of the undesirable tastes, of products derived from soybean and pea (Okubo et al. 1992; Heng et al. 2006). This undesirable flavor can limit the use of these crops as plant-based protein sources. Driven by global population growth and economic development, meeting the surging demand for dietary protein has become an urgent priority (Henchion et al. 2017). Soyfoods are considered a critical source of vegetable protein because, based on the amino acid composition and quantities, their protein quality is equivalent to that of animal proteins (FAO/WHO 1991; Hughes et al. 2011). Consequently, it is becoming increasingly important to remove undesirable tastes from soyfoods. Because soyasaponins are confined to the seed hypocotyls, their mechanical elimination of hypocotyls during processing poses technical difficulties and economical inefficiencies. Therefore, genetic improvement strategies have been undertaken to deplete soyasaponins in soybean seeds (Kato et al. 2007).

Compounds that are uniquely produced by certain plants are referred to as specialized metabolites (or secondary metabolites), in contrast to ubiquitous compounds such as sugars and amino acids that are produced across all organism. The production of specialized metabolites is generally considered as the consequence of evolutionary adaptation, serving functions such as attracting pollinators or providing defense against biotic and abiotic stresses (Pichersky and Gang 2000; Pichersky and Lewinsohn 2011). In soybean, a prominent example is isoflavones. Isoflavones are well understood to initiate symbiosis with rhizobia by inducing the expression of *nod* genes of *Bradyrhizobium japonicum* (Kosslak et al. 1987; Subramanian et al. 2006). Isoflavones also influence the composition of the soybean rhizosphere-microbiome (White et al. 2017; Okutani et al. 2020), and act as phytoalexins to inhibit the proliferation of pathogenic organisms (Zhou et al. 2011). In soybean plants, soyasaponins are constitutively accumulated at levels comparable to isoflavones. The soyasaponin content is approximately 0.5 % (w/w) in whole seeds and up to 6% (w/w) in hypocotyls (Shiraiwa et al. 1991a; Tsukamoto et al. 1995). Despite the considerable metabolic investment implied by this accumulation, the precise biological role of soyasaponins in soybean plants remains largely unresolved.

Understanding the role of the soyasaponins in plants is challenging because their chemical structures are considerably diverse. Soyasaponins biosynthesized in soybean are broadly classified into two groups based on their chemical structure: group A soyasaponins and 2,3-dihydro-2, 25-dihydroxy-6-methyl-4 *H*-pyran-4-one (DDMP) soyasaponins. Group A soyasaponins have two sugar chains at the C-3 and C-22 hydroxyl groups of the aglycone moiety, soyasapogenol A (3β, 21β, 22β, 24-tetrahydroxyolean-12-ene) (Shiraiwa et al. 1991b). DDMP soyasaponins contain a sugar chain at the C-3-hydroxyl group of the aglycone moiety in which a DDMP residue forms a hemiacetal link with soyasapogenol B (3β, 22β, 24-trihydroxyolean-12-ene) at the C-22-hydroxyl group (Kudou et al. 1992, 1993a). Moreover, DDMP soyasaponins can be degraded into group B and group E soyasaponins (Kudou et al. 1993a). Variations in the number and composition of sugar moieties result in 50 types of soyasaponins. This extensive structural diversity, along with its resulting functions, complicates efforts to define the precise biological roles of soyasaponins.

Examination of soyasaponin distribution in soybean plants indicates a functional differentiation among these compounds. A previous study demonstrated that one of the group B saponins was detected in all organs examined, whereas group A saponins were confined to the seed hypocotyls (Shimoyamada et al. 1990). Subsequent investigations revealed that DDMP saponins are present in nearly all organs, whereas group A and group B saponins predominantly accumulate in the roots

(Tsuno et al. 2018; Ma et al. 2024). Furthermore, group A and group B saponins were secreted from soybean roots and occur in the rhizosphere, whereas DDMP saponins were absent from both root exudates and the rhizosphere (Tsuno et al. 2018; Fujimatsu et al. 2020). These distinct distribution patterns suggest that each class of soyasaponins fulfills specific physiological roles within particular organs and tissues of the plants. In barrel medicine (*Medicago truncatula*), bayogenin and hederagenin glycosides, the major classes of saponins, accumulate predominantly in the roots, whereas zanhic acid glycosides, another class of saponins, are synthesized exclusively in aerial organs (Lei et al. 2019). Recently, medicagenic acid glycosides, present in both aerial and root tissues of barrel medics, have been shown to form a two-component system that contributes to resistance against fungal pathogens in the roots (Lei et al. 2019; Lacchini et al. 2023). In contrast, zanhic acid glycosides exhibit weaker antimicrobial activity than medicagenic acid glycosides and are therefore considered to function primarily as antifeedant agents against herbivores (Avato et al. 2006; Lei et al. 2019). These findings support the functional differentiation among soyasaponins in legume plants.

The production of saponin-deficient mutant is essential not only to improve flavor profiles of soybean but also to elucidate the physiological role of soyasaponins. The biosynthetic pathway of soyasaponins has been extensively studied. The initial step in soyasaponin biosynthesis is the cyclization of 2,3-oxidosqualene, catalyzed by β-amyrin synthase (Kushiro et al. 1998). Subsequent oxidation of β-amyrin by cytochrome P450 enzymes produces soyasapogenol A and B, the aglycones for soyasaponins. Glycosylation of these aglycones further diversifies the structural and functional characteristics of soyasaponins (Seki et al. 2015). In soybean, several genes involved in soyasaponin biosynthesis have been identified in soybean, including those encoding β-amyrin synthase (BAS) (Takagi et al. 2011), three cytochrome P450 enzymes (Shibuya et al. 2006; Yano et al. 2017), and eight glycosyltransferases (Shibuya et al. 2010; Sayama et al. 2012; Yano et al. 2018; Takagi et al. 2018; Sundaramoorthy et al. 2019; Chung et al. 2020).

Genetic modifications targeting genes involved in soyasaponin biosynthesis have been performed to reduce soyasaponin contents in soybean seeds. For example, knockdown of the *GmBAS* genes resulted in a marked reduction of soyasaponin levels in soybean seeds (Takagi et al. 2011). Soyasaponin-deficient mutants identified in a high-density soybean mutant library, generated through ethyl methanesulfonate treatment, were found to harbor mutations in the *GmBAS1* locus (Krishnamurthy et al. 2019). Together, these findings underscore that functional disruption of *GmBAS* genes represents a particularly effective means of reducing soyasaponin content in soybean seeds.

The objective of this study is to obtain soyasaponin-deficient mutant for developing cultivars with improved flavor profiles and to elucidate the biological functions of soyasaponins from multiple perspectives. In chapter 2, targeted mutagenesis targeting *GmBAS* genes was performed. In chapter 3, using a *GmBAS1* knockout mutant, the impact of saponin deficiency on metabolic flux in seeds and the composition of the rhizosphere bacterial community was analyzed. In chapter 4, the possibility of seed-specific regulation of soyasaponin biosynthesis was discussed. Moreover, the spatial distribution of soyasaponins in soybean seeds was examined using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry imaging (MALDI-TOF MSI).

CHAPTER 2

Simultaneous targeted Mutagenesis for Soybean β-amyrin Synthase Genes via DNA-free CRISPR/Cas9 System Using a Single Guide RNA

2.1 Background and Purpose

Two homologs coding for β-amyrin synthase (*GmBAS1* and *GmBAS2*) were reported in a previous molecular biological study of soybeans (Takagi et al. 2011). These homologs share over 95% similarity in their amino acid sequences and likely retain key motifs essential for enzymatic function (Takagi et al. 2011). While the high similarity between their amino acid sequences suggests potential functional redundancy, their expression patterns differ across tissues. The expression level of *GmBAS1* is significantly higher than that of *GmBAS2* in various tissues (Takagi et al. 2011). In contrast, *GmBAS2* is expressed at a lower level than *GmBAS1* and specific expressions in flowers and pods, as indicated by RNA-seq data retrieved from a gene expression database (Almeida-Silva et al. 2023). These differences may reflect the spatial and temporal regulation of soyasaponin biosynthesis in soybean tissues. However, the functional roles of these homologs remain unclear. Therefore, generating mutants that lack one or both functions of the *BAS* homologs in the same genetic background, followed by their physiological characterization, will be essential for understanding the distinct roles of these genes.

Targeted mutagenesis is an effective technique for generating mutations in target genes. The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated endonuclease 9 (Cas9) system is one of the most widely used targeted mutagenesis systems for higher plants. In soybeans, the CRISPR/Cas9 system is conventionally applied through two transformation methods: Agrobacterium-mediated and particle bombardment (Sim et al. 2023). Both methods typically require the production of transgenic plants via regeneration from explants. However, the efficiency of plant regeneration is genotype-dependent, and soybean transformation is only successful in a limited range of genotypes (Yamada et al. 2012). This limitation also restricts the application of targeted mutagenesis in soybean. Furthermore, the integration of the CRISPR/Cas9 expression module as foreign DNA into the soybean genome poses a significant challenge for commercial use, as the removal of foreign genes through genetic segregation is a time-consuming and labor-intensive process. Recently, Kuwabara et al. (2024) optimized the in planta bombardment-ribonucleoprotein (iPB-RNP)

method, initially developed as a DNA-free targeted mutagenesis system in wheat (*Triticum aestivum*) (Kumagai et al. 2022), for soybean. In the iPB-RNP system, a complex of single guide RNA (gRNA) and Cas9 protein is directly introduced into the shoot apical meristems (SAMs) of soybean embryonic axes via particle bombardment (Kuwabara et al. 2024). This method circumvents the need for regeneration to obtain mutants, thus expanding the range of genotypes available for targeted mutagenesis. Additionally, the direct delivery of the ribonucleoprotein (RNP) complex reduces the risk of off-target mutations due to probable rapid degradation of the RNP, and it eliminates the need for genetic segregation to remove foreign DNA.

In this chapter, as an advanced application of the iPB-RNP method, targeted mutagenesis was performed at each individual locus and simultaneously at the two loci using a single RNP targeting a common sequence in two homologous *GmBAS* genes. Subsequently, DNA sequencing and high-performance liquid chromatography (HPLC) analyses were carried out to assess the impact of loss of function in each *GmBAS* homolog on soyasaponin accumulation.

2.2 Materials and Methods

2.2.1 Preparation of soybean shoot apical meristem

Mature soybean seeds (*G. max* L. cv. GL3510) were sterilized using 1.5% sodium hypochlorite for 2 min and thoroughly rinsed. After being washed ten times with sterile distilled water, the seeds were soaked in sterile distilled water for 15–16 h in the dark at 25°C. Embryonic axes were carefully excised from the cotyledons using a scalpel, and unifoliate leaves and primary stipules were delicately removed under a stereomicroscope. A total of 16 explants were arranged in a circular pattern, approximately 1 cm in diameter, with the shoot apical meristem regions oriented upward on a bombardment medium. This medium consisted of Murashige-Skoog (MS) medium (pH 5.7, Sigma-Aldrich), 20 g/L sucrose, 36.4 g/L mannitol, 36.4 g/L sorbitol (Fujifilm Wako Pure Chemical), 0.5 mg/L 2-Morpholinoethanesulfonic acid (MES) (Dojindo Laboratories), 0.7 mg/L L-proline (Fujifilm Wako Pure Chemical), 20 mg/L meropenem (Sumitomo Dainippon Pharma), and 4.1 g/L phytagel (Millipore Sigma).

2.2.2 Preparation of gold particles and particle bombardment

In this study, only crispr RNA (crRNA) and trans-activating crRNA (tracrRNA) were used as guide sequences for the Cas9 protein to perform DNA-free targeted mutagenesis. Both the crRNA and

tracrRNA were obtained from FASMAC Co., Ltd. and dissolved in nuclease-free water to a concentration of 1 μg/μL, stored at –80°C. To form a single gRNA, guide crRNA and tracrRNA were mixed in equal amounts and incubated on ice for 10 min. Subsequently, *Streptococcus pyogenes* Cas9 (SpCas9) protein (NEB), 1x CutSmart buffer (NEB), and RNase inhibitor (Thermo Fisher Scientific) were added, and the mixture was incubated at room temperature for an additional 10 min. The gRNA/SpCas9 complex solution was then combined with TransIT transfection reagent (TaKaRa) and incubated at room temperature for 5 min. Gold particles (1.0 μm; BioRad) were added to the mixture and incubated on ice for 10 min. The suspension was centrifuged at 1,200 ×g for 1 s, and the resulting pellet was resuspended in nuclease-free water. The gRNA/SpCas9-coated gold particle suspension was distributed onto a hydrophilic membrane (Scotchtint, 3M) attached to a microcarrier and allowed to dry naturally prior to bombardment.

Bombardment was performed using a Biolistic PDS-1000/He Particle Delivery System (BioRad) according to a slightly modified version of the manufacturer's instructions. The distance between the target plantlets and the stopping screen was approximately 3.5 cm. Each plate containing the explants was bombarded at a helium pressure of 1,800 psi. Following bombardment, the explants were incubated at 25°C in darkness for 24 h.

2.2.3 Caring of explants

Twenty-four hours after bombardment, the explants were transferred to MS medium (pH 5.7) supplemented with 30 g/L sucrose, 0.5 g/L MES, 20 mg/L meropenem, 3 mL/L Plant Preservative Mixture (Plant Cell Technology), 1.12 mg/L 6-Benzyladenine (Fujifilm Wako Pure Chemical), 7 g/L charcoal (Nacalai Tesque), and 2.05 g/L phytagel. These explants were maintained at 25°C under an 18-hour light photoperiod. Approximately two weeks later, explants with successfully elongated shoots and roots were individually transferred to seedling trays filled with potting soil (Sumitomo Forestry Landscaping). To prevent desiccation of the young leaves, each tray was covered with a plastic sheet. The plantlets were kept at 25°C under a 16-hour light photoperiod. The E_0 plants were subsequently transferred to plastic pots filled with a mixture of potting soil and culture soil (Katakura & Co-op Agri Corp) in a 1:1 ratio. These plants were maintained in a closed greenhouse under a 12-hour photoperiod at 25°C until the E_1 seeds were collected.

2.2.4 Cleaved amplified polymorphic sequence (CAPS) analysis

To extract genomic DNA, leaf pieces were collected in bulk from bombarded seedlings, with 3–4 trifoliate leaf pieces obtained separately from each plant. Genomic DNA extraction from leaf pieces and mature seeds was conducted as described by Sugano et al. (2020). The target regions in the *GmBAS1* and *GmBAS2* loci were amplified by PCR using specific primers (Table 2.1). PCR conditions were as follows: 30 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 30 s. As the targeted region contained the recognition site of restriction enzyme EcoNI (NEB), successful mutagenesis should abolish the site, conferring the resistance to EcoNI to the PCR product. The PCR products were digested with the EcoNI and separated by electrophoresis in 2.0% agarose gels. DNA fragments of the expected sizes (i. e., resistant to EcoNI) were identified as mutants.

2.2.5 DNA sequencing

Fragments amplified using specific primers for sequencing were purified using ExoSAP-IT reagent (Affymetrix) and sequenced directly or after cloning into the pGEM-T-Easy vector (Promega). Sequencing of the amplified products was performed on an ABI3130 sequencer (Applied Biosystems). This sequence analysis was carried out by the Instrumental Analysis Division at the Graduate School of Agriculture, Hokkaido University, and occasionally by FASMAC Co., Ltd. through their contract analysis service.

2.2.6 Expression analysis of the *GmBAS* genes by quantitative RT-PCR (qRT-PCR)

Total RNA was extracted from the leaves, stems, and roots of juvenile seedlings, as well as from immature seeds, using the LiCl precipitation procedure (Adachi et al. 2021). Each cDNA was synthesized from 500 ng of total RNA using the M-MLV reverse transcriptase system (Invitrogen) with random hexamer primers and an oligo(dT)20 primer, following the manufacturer's instructions.

The qRT-PCR reaction was carried out in a 20-μL mixture containing 1 μL of cDNA, 0.8 μL of each primer (10 μM), and 10 μL of SYBR Premix ExTaq II (Tli RNaseH plus) (TaKaRa). Reactions were performed using a CFX 96 Real-Time System (BioRad) under the following conditions: 95°C for 4 min, followed by 40 cycles of 95°C for 10 s, 59.5°C for 20 s, 72°C for 20 s, and 78°C for 2 s. Fluorescence was quantified before and after incubation at 78°C to monitor primer dimer formation. The expression levels of *GmBAS1* (Glyma.07G001300) and *GmBAS2* (Glyma.08G225800) were normalized against the expression of the *Bic-C2* gene (Glyma.03G064800). The genome of *G. max* Wm82.a2.v1 (Phytozome genome ID: 275) were used as a reference. Amplification of a single DNA

fragment for each transcript was confirmed by melting curve analysis. Averages and standard errors of relative expression levels were calculated from three independent plants. Primers used for expression analysis are listed in Table 2.1.

2.2.7 Extraction and HPLC analysis of soyasaponins

Soyasaponins were extracted from mature seeds, leaves, stems, and roots of juvenile seedlings. Dried seeds and lyophilized leaves, stems, and roots were ground using a Multi-Beads Shocker (MB1200; Yasui Kikai Co.). Thirteen mg of fine powder was thoroughly mixed with 260 µL of 70% aqueous ethanol containing 0.1% (v/v) acetic acid. The mixture was incubated at 4°C for 1 hour in darkness, with vortexing every 15 min. After centrifugation at 17,900 ×g for 10 min, 100 µL of the supernatant was transferred into a 0.2 mL tube. Leaf, stem, and root samples were filtered using 0.45 µm GL chromatodisk filters (GL Sciences). Ten µL of HCl was added to the samples, and the mixture was incubated at 85°C for 3 h. The hydrolyzed solution was then transferred into an HPLC vial insert.

Analysis was performed using an HPLC system (Hitachi LaChrom Elite, Hitachi High-Technologies Corp.) with a Unison UK-C18 reverse-phase column (3.0 mm × 150 mm, Imtakt Corp.). The column temperature was maintained at 40°C, and separation was performed under gradient conditions. Solvent A was acetonitrile (0.1% acetic acid), and solvent B was water (0.1% acetic acid). The gradient program for mature seeds was as follows: linear 35–25% B from 0–12.5 min; isocratic at 25% B from 12.3–13 min; isocratic at 0% B from 13.1–15.5 min; and isocratic at 35% B from 15.6–24 min. For leaves, stems, and roots of juvenile seedlings, the gradient was linear 50–46% B from 0–40 min; linear 46–0% B from 40–42 min; isocratic at 0% B from 42–50 min; linear 0–50% B from 50–52 min; and isocratic at 50% B from 52–60 min. The injection volume for each sample was 10 µL, and the flow rate was 0.44 mL/min. Soyasapogenols were detected at a wavelength of 210 nm. Soyasapogenol A and B, purchased from Fujifilm Wako Pure Chemicals, were dissolved in 70% aqueous ethanol. The content of soyasapogenols in mature seeds was determined based on the peak area ratio of the standards. In leaves, stems, and roots, the content was calculated based on the peak height ratio of the standards.

2.2.8 Statistical analysis

Tests for significance among the soyasapogenol contents were performed using Student's *t*-test. *P*-values < 0.05 were considered statistically significant. Test of significance among means of

transcript abundances was performed using Tukey-Kramer test. *P*-values < 0.01 were considered statistically significant.

2.3 Results

2.3.1 Induction of mutations at soybean β-amyrin synthase (*GmBAS*) loci in the E_0 generation

GmBAS homologs are located on chromosomes 7 and 8 and consist of 16 exons (Figure 2.1). To investigate the expression patterns of *GmBAS* homologs in GL3510, qRT-PCR was performed on juvenile seedlings and immature seeds using homolog-specific primer sets. *GmBAS1* was expressed in all examined tissues, with the highest levels in the roots (Figure 2.1.b). The expression levels of *GmBAS2* were similar across all tissues (Figure 2.1.c).

A single gRNA targeting sequences conserved in both *GmBAS* homologs was designed to induce targeted mutagenesis (Figure 2.2.a). Mutation induction at the target regions was assessed by CAPS analysis one month after bombardment. DNA fragments were classified as wild-type or mutant-type based on the expected fragment size (Figure 2.2.b). Among the 1,467 seedlings screened, 80 showed mutations at the *GmBAS1* locus, 10 at the *GmBAS2* locus, and 48 at both loci (Table 2.2). In this study, the post-bombarded generation was designated as E_0. To obtain mutants at either the *GmBAS1* or *GmBAS2* loci, or both, in the E_1 generation, E_0 plants were analyzed by CAPS at the 3–4 trifoliate leaf development stage and in the E_1 seeds. A total of 36 E_0 plants were selected for further analysis in the next generation (Table 2.2).

2.3.2 Detection of inheritable mutations to the E_1 generation

The inheritance of mutations to the next generation was evaluated by CAPS and sequencing analyses in the E_1 generation (Table 2.3). A total of 16 E_0 plants displayed inheritable mutations in the E_1 generation: seven at the *GmBAS1* locus, two at the *GmBAS2* locus, and seven at both loci (Table 2.2). E_1 seeds possessing homozygous mutant alleles or biallelic mutations, as confirmed by CAPS analysis, were subjected to sequencing. Sequencing analysis revealed a range of mutations, including single-nucleotide insertions and deletions of 1, 3, 4, 5, 6, or 7 nucleotides at the *GmBAS1* locus, and single-nucleotide insertions and deletions of 1, 2, 4, 5, 7, or 10 nucleotides at the *GmBAS2* locus (Figure 2.3 and Table 2.4). Based on predicted amino acid sequences, all mutant alleles—except those with deletions of 3 or 6 nucleotides—were highly likely to generate premature stop codons through frameshift mutations, resulting in nonfunctional mutant proteins (Figure 2.4). Moreover, no off-target mutations were detected in the E_1 generation (Table 2.5).

2.3.3 Determination of soyasaponin in mutant seeds and juvenile seedlings

To assess the impact of loss of function in each *GmBAS* homolog on soyasaponin accumulation in soybean seeds, the saponin content of mutant seeds was measured using HPLC. Mutant lines used in HPLC analysis are listed in Table 2.6. These lines were categorized into three haplotypes (*bas1*, *bas2*, and *double*) based on the genotypes of the *GmBAS* loci. Three lines (27-10-1, 50-3-2, and 104-4-2) were selected as representative *bas1* mutants. Another three lines (32-10-2-1, 52-11-2-13, and 99-3-2) were chosen as representative *bas2* mutants. The remaining four lines, 52-11-1-5, 113-7-4, 113-7-6, and 113-7-10 were classified as representative *double* mutants. Two plants, 52-11-2-13 and 52-11-1-5, were separated from the E_1 plants that were judged to be *double* mutants and were subsequently divided as *bas2* and *double* mutant haplotypes, respectively (Table 2.6).

Due to the structural diversity of soyasaponins, which vary in the number and composition of sugar moieties, identifying all soyasaponins via HPLC is challenging. Therefore, extracts from mature seeds were hydrolyzed to detect soyasaponin aglycons (soyasapogenols), and the total soyasapogenols content was estimated. In the extracts from wild-type seeds, peaks corresponding to soyasapogenol A and B were detected (Figure 2.5). The contents of soyasapogenol A and B were 88.9 ± 29.9 mg/100 g DW and 166.2 ± 42.0 mg/100 g DW, respectively (Table 2.7). In the extracts from mutant-type seeds, peaks corresponding to both soyasapogenols were detected only in the *bas2* mutant lines, but not in the *bas1* and *double* mutant seeds (Figure 2.5). Therefore, the *bas1* and *double* mutant seeds were deficient in soyasapogenols. The concentrations of soyasapogenol A and B in seeds were 92.6 ± 8.5 and 163.6 ± 18.4 mg/100 g DW in the 32-10-2-1 line, 75.4 ± 11.0 and 150.9 ± 20.2 mg/100 g DW in the 52-11-2-13 line, and 76.2 ± 6.2 and 155.1 ± 15.5 mg/100 g DW in the 99-3-2 line, respectively (Table 2.7). No statistically significant difference in soyasapogenols content was observed between the *bas2* mutants and control seeds. Collectively, the soyasapogenol content of *bas2* mutant seeds was compatible to that of control seeds.

To assess the effects of *GmBAS* mutations in other plant tissues, soyasapogenols content in the roots, stems, and leaves of juvenile seedlings was measured. In wild-type plant extracts, peaks for soyasapogenol A and B were detected in the roots and stems, while only soyasapogenol B was detected in the leaves (Figure 2.6). In wild-type roots, the concentrations of soyasapogenol A and B were 105.4 ± 16.5 and 377.4 ± 33.7 mg/100 g DW, respectively. In wild-type stems, the concentrations of soyasapogenol A and B were 14.3 ± 10.1 and 124.2 ± 41.1 mg/100 g DW, respectively. In both tissues, the content of soyasapogenol B was higher than that of soyasapogenol A. In wild-type leaves, the

concentration of soyasapogenol B was 150.0 ± 30.0 mg/100 g DW (Table 2.8). The total soyasapogenol content in the roots was the highest among all tissues examined (Table 2.8).

For the HPLC analysis in juvenile seedlings, two mutant lines from each haplotype were selected (Table 2.8). Peaks for soyasapogenols were detected in the roots, stems, and leaves of only the *bas2* mutant plants, and not in the *bas1* or double mutant plants. Therefore, the roots, stems, and leaves of the *bas1* and *double* mutants were deficient in soyasapogenols. In the roots of the 32-10-2-1 line, the concentrations of soyasapogenol A and B were 92.4 ± 20.6 and 360.9 ± 61.1 mg/100 g DW, respectively, and in the 52-11-2-13 line, they were 102.8 ± 18.7 and 386.6 ± 42.7 mg/100 g DW (Table 2.8). The soyasapogenol content in the *bas2* mutant roots were compatible to that of control roots. In the stems, the concentrations of soyasapogenol A and B were 26.5 ± 3.2 and 245.1 ± 114.8 mg/100 g DW in the 32-10-2-1 line, and 25.1 ± 15.7 and 278.9 ± 93.5 mg/100 g DW in the 52-11-2-13 line (Table 2.8). Although soyasapogenol content appeared to increase in *bas2* mutant stems, no significant difference was observed except for the content of soyasapogenol B in the 52-11-2-13 line. The concentrations of soyasapogenol B in leaves were 225.6 ± 46.4 mg/100 g DW in the 32-10-2-1 line, and 237.6 ± 33.2 mg/100 g DW in the 52-11-2-13 line (Table 2.8). The content of soyasapogenol B was significantly higher in the *bas2* mutant leaves than that in control leaves.

2.4 Discussion

Targeted mutagenesis of *GmBAS* homologs involved in the first step of soyasaponin biosynthesis in soybeans was performed using the iPB-RNP system. In this study, the iPB-RNP method was advanced by executing simultaneous mutagenesis at two target loci using a single gRNA without off-target mutations. Although the conserved sequences were targeted to the *GmBAS1* and *GmBAS2* loci in this study, the mutagenesis frequency of the *GmBAS1* locus was higher than that of the *GmBAS2* locus (Table 2.2). In *Arabidopsis thaliana*, heterochromatic features were strongly associated with the low mutagenesis frequencies using the CRISPR/Cas9 system (Weiss et al. 2022). Epigenetic resources may therefore influence the differences in targeted mutagenesis frequencies between *GmBAS1* and *GmBAS2* loci.

Conventional targeted mutagenesis through Agrobacterium transformation with a single gRNA has achieved simultaneous mutagenesis at multiple target regions in soybean (Kanazashi et al. 2018; Wang et al. 2019; Li et al. 2019; Jia et al. 2020; Fu et al. 2022; Shen et al. 2022). However, in the iPB-RNP method, opportunities for simultaneous mutagenesis at multiple loci are limited, as RNPs introduced into cells are rapidly degraded during cell division. In this study, mutant alleles at both target loci were detected in the E_1 seeds from seven E_0 plants (Table 2.2). These findings indicate that simultaneous mutagenesis was induced in a single cell where RNPs had been delivered. Furthermore, the introduction of RNPs into cells also generated different mutant alleles in one target locus. One- or five-nucleotide deletions at the *GmBAS1* locus, and two- or four-nucleotide deletions at the *GmBAS2* locus, were detected in the progeny of the 113-7 E_0 plant (Table 2.6). In conventional targeted mutagenesis via Agrobacterium transformation, four mutant alleles were identified at the *Gly m Bd 30K* locus, one of the allergenic genes in soybean, in the progeny of one T_0 plant (Sugano et al. 2020). This suggests that the constant presence of the CRISPR/Cas9 module via exogenous genes may result in the production of multiple mutant alleles at the target locus. Interestingly, in addition to the two mutant alleles, a wild-type allele was detected at the *GmBAS1* locus in the progeny of the 113-7 E_0 plant (Figure 2.7), suggesting that three alleles were produced at this locus. In the iPB-RNP system, the delivery of RNPs to cells that will differentiate into germ cells is crucial for inducing inheritable mutations in subsequent generations (Kuwabara et al. 2024). Thus, the three alleles detected in the progeny of the 113-7 plant likely originated from two or more cells that later differentiated into germ

cells. Enhancing the precision of RNP delivery to target cells in the iPB-RNP system could increase the likelihood of producing multiple mutant alleles at target loci.

In this study, deletion mutations ranging from one to ten nucleotides were observed in the target regions (Table 2.4). In the previous study using the iPB-RNP system, deletion mutations of up to 13 nucleotides have been observed in the *Gly m Bd 30K* gene, another allergenic gene (Kuwabara et al. 2024). Although there are few reports of iPB-RNP system-based targeted mutagenesis in soybeans, the deletion size induced by this system appears to be approximately ten nucleotides. By contrast, deletion mutations exceeding ten nucleotides are sometimes found in conventional targeted mutagenesis via Agrobacterium transformation in soybeans (Li et al. 2019; Shen et al. 2022). These results indicate that the number of nucleotides removed by the iPB-RNP system is generally lower compared to conventional targeted mutagenesis via transformation. In site-directed mutagenesis, nucleotide deletions primarily result from errors during the non-homologous end joining (NHEJ) repair process (Li et al. 2015). The availability of the CRISPR/Cas9 modules at target regions differs significantly between conventional targeted mutagenesis and iPB-RNP systems. The iPB-RNP system provides these modules transiently, while conventional methods allow prolonged exposure. Therefore, to achieve large deletion exceeding ten nucleotides using iPB-RNP systems, it is necessary to enhance the stability of CRISPR/Cas9 modules in the plant cells. Furthermore, it is considered important to increase the frequency with which these modules encounter cellular conditions prone to large deletions before they are degraded. For example, if large nucleotide deletions due to NHEJ repair errors occur more frequently during certain stages of cell division, the enhancement in CRISPR/Cas9 module availability during those stages may increase the probability of generating large deletions using the iPB-RNP systems.

Frameshift mutations at the *GmBAS1* locus resulted in the deficiency of soyasaponins in mature seeds, roots, stems, and leaves of mutant plants (Figures 2.5 and 2.6). These results suggest that *GmBAS1* is predominant in soyasaponin biosynthesis in these tissues. The amino acid sequence similarity between *GmBAS* homologs is 95.9% (Takagi et al. 2011), and their enzymatic activities are likely redundant, as they conserve the DCTAE and QW motifs, which are responsible for substrate binding and stabilization of carbocationic intermediates, respectively (Poralla et al. 1994; Abe and Prestwich 1995). The expression level of *GmBAS1* is significantly higher than that of *GmBAS2* in various tissues (Takagi et al. 2011). In *Glycyrrhiza uralensis*, the expression of the *BAS* gene was upregulated in response to methyl jasmonate, a compound associated with plant defense responses

(Tamura et al. 2018). Although *GmBAS2* expression was low in this study, the expression of *GmBAS* genes may increase in response to biotic stress.

Previous studies have shown that DDMP soyasaponins and group B soyasaponins, which contain soyasapogenol B as an aglycon, are the major types of soyasaponins in aerial tissues and roots, respectively (Tsuno et al. 2018; Ma et al. 2024). Consistent with these findings, quantitative analysis of soyasaponins showed that soyasapogenol B levels were higher than those of soyasapogenol A in all tissues examined (Tables 2.7 and 2.8). Moreover, the soyasaponin content in roots was the highest among all tissues examined (Tables 2.7 and 2.8). Various soyasaponin compounds accumulate in roots during the vegetative growth stage, and some of these compounds are secreted externally to regulate the rhizosphere environment (Tsuno et al. 2018). Therefore, the role of soyasaponin in soybean roots would extremely important.

No significant differences in morphological characteristics were observed between mutant and control plants under the controlled growth conditions in this study (Figure 2.8). The soyasaponin-deficient mutants generated in this study will provide valuable tools to investigate the biological and physiological roles of soyasaponins in soybeans. To address this challenge, the detailed physiological analysis of the soyasaponin-deficient mutants was provided in the subsequent chapter.

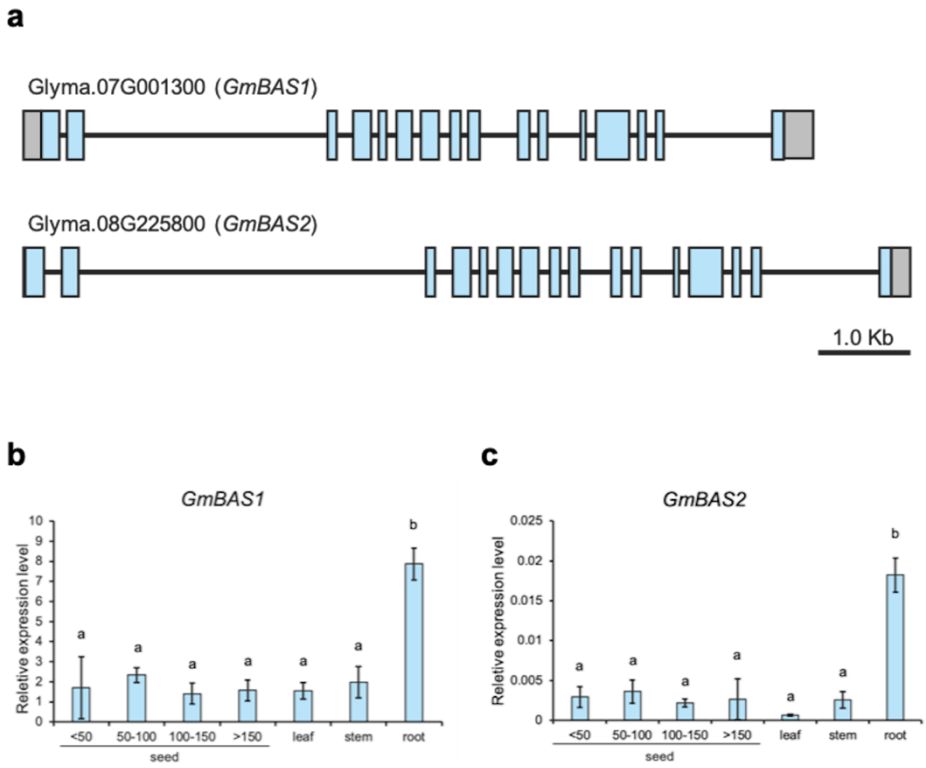


Figure 2.1 Structure and expression properties of *GmBAS* genes.

a Schematic diagrams of *GmBAS1* and *GmBAS2* genes. Blue and grey boxes represent translated and untranslated regions, respectively. Bold lines represent introns.

b, c Expression analysis of *GmBAS* genes in soybean tissues. Relative expression levels were normalized to the internal control gene *Bic-C2* (Glyma.03G064800). Seed developmental stages were classified based on fresh seed weight (mg per seed). Leaf, stem, and root samples were collected from young seedlings. Error bars represent the standard deviation from three biological replicates. Different alphabets indicate significant differences based on Tukey-Kramer test, ($p < 0.01$).

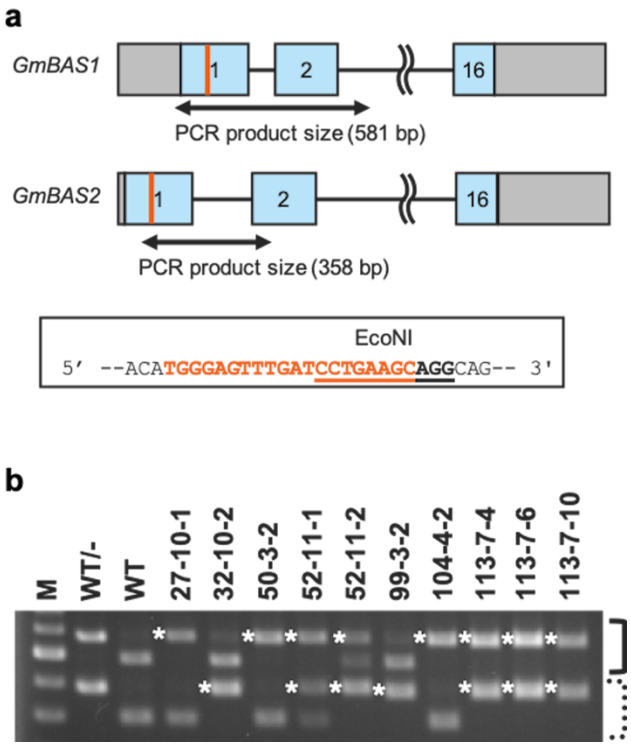


Figure 2.2 Target regions of a single gRNA and detection of mutations by CAPS analysis.

a Target sequences of a single gRNA conserved between *GmBAS* genes. The single gRNA was designed within conserved sequences between the two genes and is represented by vertical lines in the schematic diagrams. Bold sequences represent the target sequences of a single gRNA and the protospacer adjacent motif (PAM) region. Numbers in the box indicate exon numbers, and arrows show PCR products from each gene used in this study. Underlined sequences indicate the EcoNI recognition site.

b CAPS analysis of amplified products from the target loci. Solid and dotted square brackets on the left represent amplified and digested products from *GmBAS1* and *GmBAS2*, respectively. Asterisks denote mutant-type fragments. Numbers above each lane indicate E_1 plant numbers. 'M' represents the 100 bp ladder marker. 'WT/-' indicates undigested products amplified from the control plant.

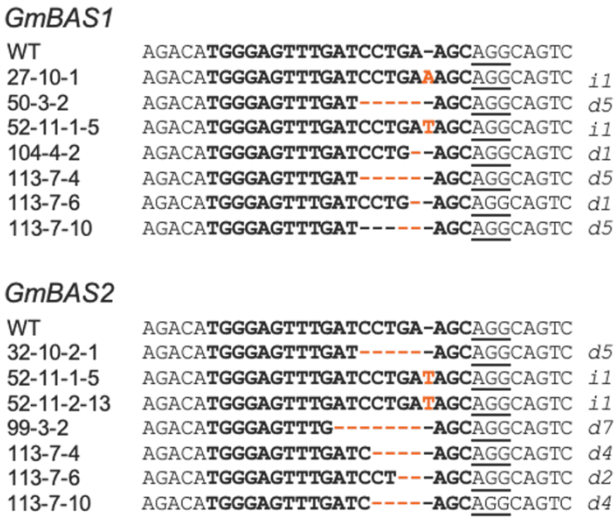
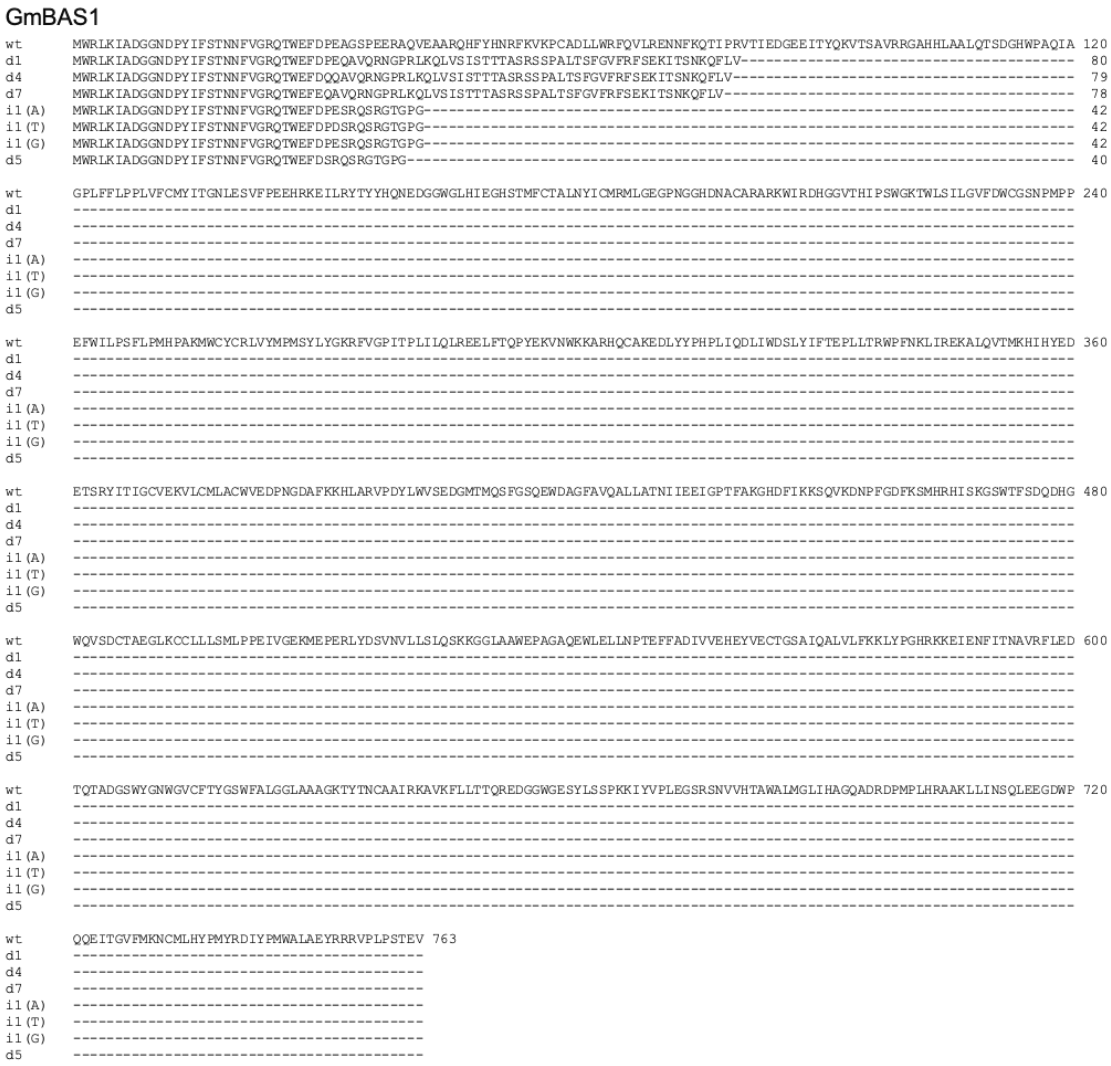


Figure 2.3 Mutations detected within the target loci in E_1 seeds.

'WT' indicates the target sequences and flanking regions of *GmBAS1* and *GmBAS2* genes in the control plant. Numbers on the left side of the sequences correspond to E_1 and E_2 seed numbers. Bold sequences represent the gRNA-targeted sequences; underlined sequences denote the PAM region. Italic letters on the right side indicate mutation types, e.g., *d1* (a single-nucleotide deletion) and *i1* (a single-nucleotide insertion).

a


GmBAS1
wt MWRLKIADGGNDPYIFSTNNFVGRQTWEFDPEAGSPEERAQVEAARQHFYHNRFKVKPCADLLWRFQVLRENNFKQTIPRVTIEDGEEITYQKVTSAVRRGAHHLAALQTSDGHWPAQIA 120
d1 MWRLKIADGGNDPYIFSTNNFVGRQTWEFDPEQAVQRNGPRLKQLVSISTTTASRSSPALTSFGVFRFSEKITSNKQFLV---------- 80
d4 MWRLKIADGGNDPYIFSTNNFVGRQTWEFDQQAVQRNGPRLKQLVSISTTTASRSSPALTSFGVFRFSEKITSNKQFLV---------- 79
d7 MWRLKIADGGNDPYIFSTNNFVGRQTWEFEQAVQRNGPRLKQLVSISTTTASRSSPALTSFGVFRFSEKITSNKQFLV---------- 78
i1(A) MWRLKIADGGNDPYIFSTNNFVGRQTWEFDPESRQSRGTGPG---------- 42
i1(T) MWRLKIADGGNDPYIFSTNNFVGRQTWEFDPDSRQSRGTGPG---------- 42
i1(G) MWRLKIADGGNDPYIFSTNNFVGRQTWEFDPESRQSRGTGPG---------- 42
d5 MWRLKIADGGNDPYIFSTNNFVGRQTWEFDSRQSRGTGPG---------- 40
wt GPLFFLPPLVFCMYITGNLESVFPEEHRKEILRYTYYHQNEDGGWGLHIEGHSTMFCTALNYICMRMLGEGPNGGHDNACARARKWIRDHGGVTHIPSWGKTWLSILGVFDWCGSNPMPP 240
d1 ----------
d4 ----------
d7 ----------
i1(A) ----------
i1(T) ----------
i1(G) ----------
d5 ----------
wt EFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRFVGPITPLILQLREELFTQPYEKVNWKKARHQCAKEDLYYPHPLIQDLIWDSLYIFTEPLLTRWPFNKLIREKALQVTMKHIHYED 360
d1 ----------
d4 ----------
d7 ----------
i1(A) ----------
i1(T) ----------
i1(G) ----------
d5 ----------
wt ETSRYITIGCVEKVLCMLACWVEDPNGDAFKKHLARVPDYLWVSEDGMTMQSFGSQEWDAGFAVQALLATNIIEEIGPTFAKGHDFIKKSQVKDNPFGDFKSMHRHISKGSWTFSDQDHG 480
d1 ----------
d4 ----------
d7 ----------
i1(A) ----------
i1(T) ----------
i1(G) ----------
d5 ----------
wt WQVSDCTAEGLKCCLLLSMLPPEIVGEKMEPERLYDSVNVLLSLQSKKGGLAAWEPAGAQEWLELLNPTEFFADIVVEHEYVECTGSAIQALVLFKKLYPGHRKKEIENFITNAVRFLED 600
d1 ----------
d4 ----------
d7 ----------
i1(A) ----------
i1(T) ----------
i1(G) ----------
d5 ----------
wt TQTADGSWYGNWGVCFTYGSWFALGGLAAAGKTYTNCAAIRKAVKFLLTTQREDGGWGESYLSSPKKIYVPLEGSRSNVVHTAWALMGLIHAGQADRDPMPLHRAAKLLINSQLEEGDWP 720
d1 ----------
d4 ----------
d7 ----------
i1(A) ----------
i1(T) ----------
i1(G) ----------
d5 ----------
wt QQEITGVFMKNCMLHYPMYRDIYPMWALAEYRRRVPLPSTEV 763
d1 ----------
d4 ----------
d7 ----------
i1(A) ----------
i1(T) ----------
i1(G) ----------
d5 ----------

b

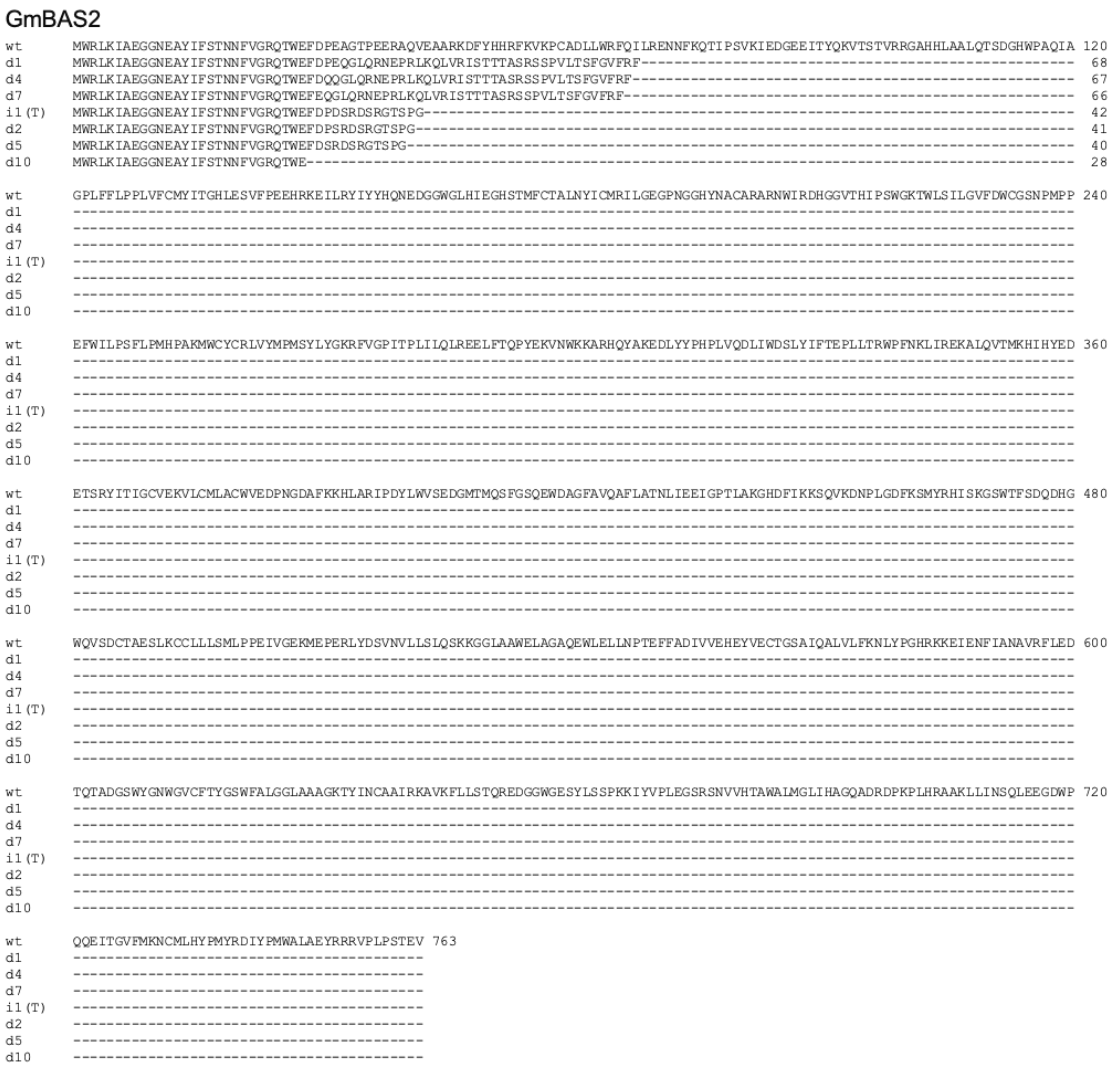


Figure 2.4 Predicted amino acid sequences of mutant GmBAS proteins in mutant and control plants. 'WT' indicates the amino acid sequences of *GmBAS1* and *GmBAS2* genes in the control plant. Letters and numbers on the left side of the alignment indicate mutation types, e.g., d1 (a single-nucleotide deletion) and i1 (a single-nucleotide insertion). All frameshift mutations detected in this study result in premature stop codon.

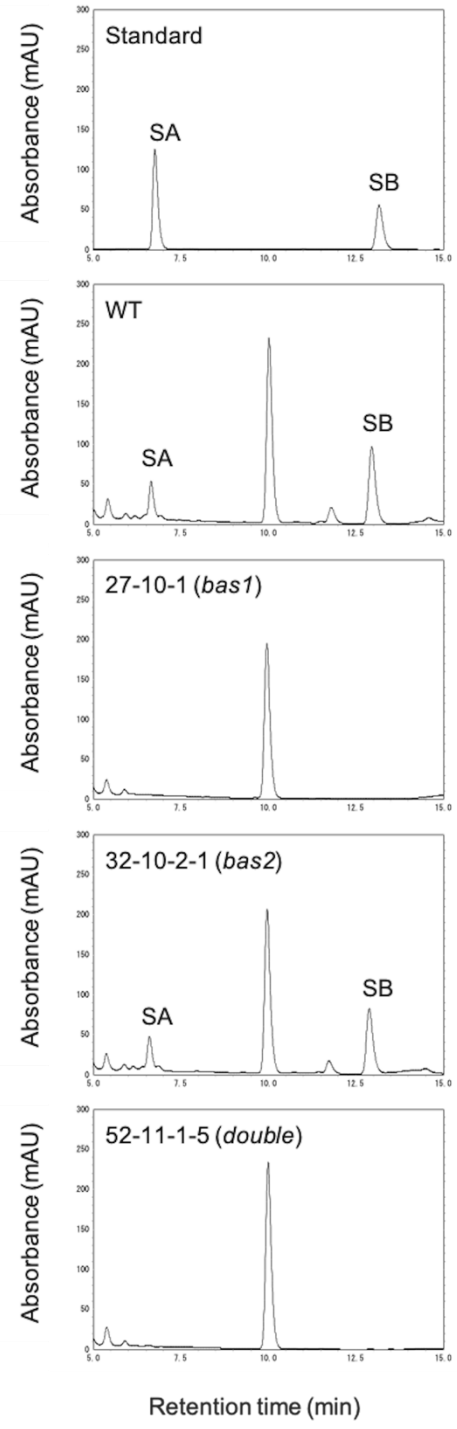


Figure 2.5 HPLC analysis of soyasapogenols extracted from mutant and control seeds. Hydrolyzed seed extracts were compared to authentic standards. 'WT' represents the chromatogram of a control seed. Numbers in the upper left of each panel represent mutant line numbers. SA and SB denote the peaks for soyasapogenol A and B, respectively.

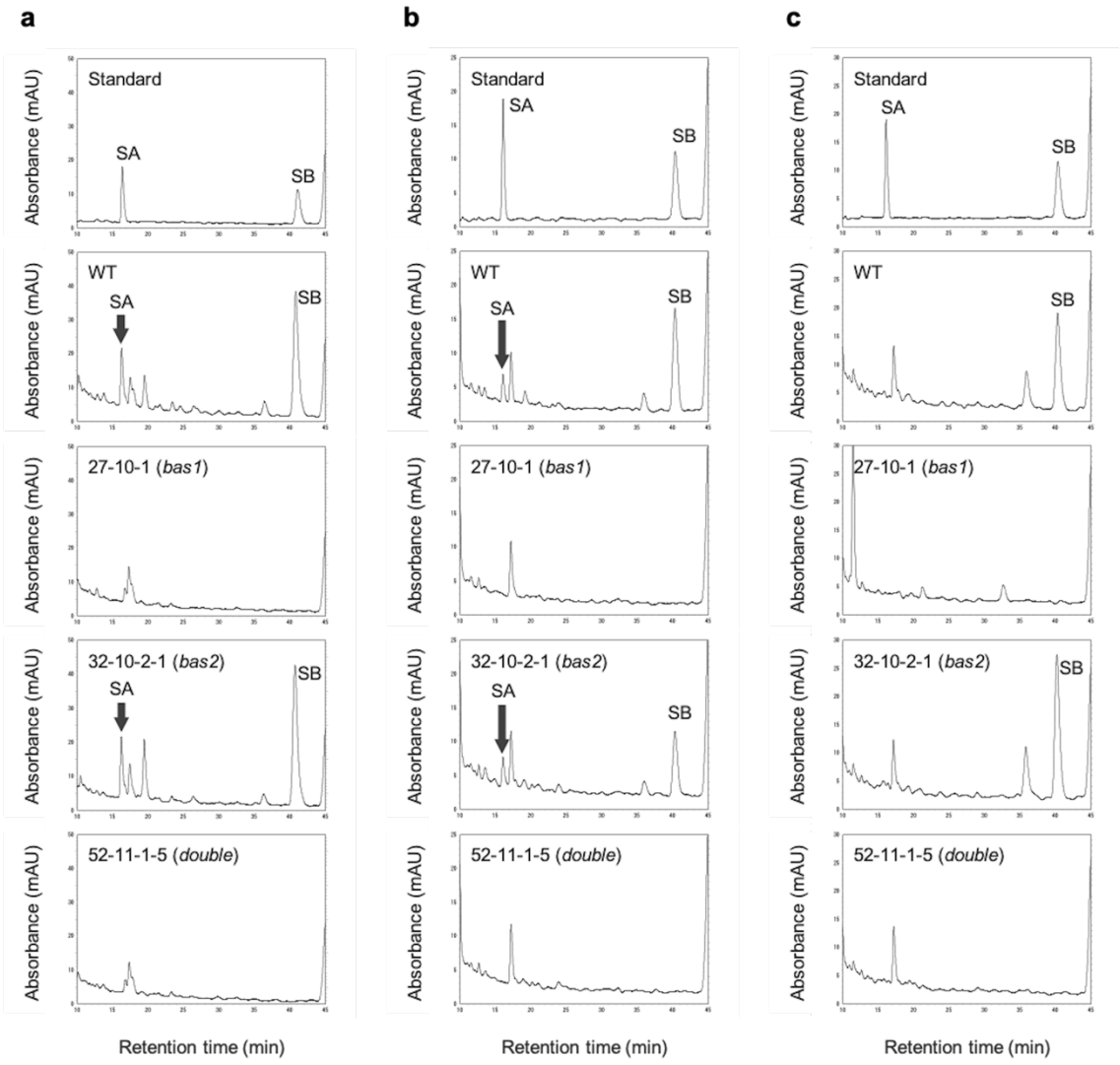


Figure 2.6 HPLC analysis of soyasapogenols extracted from mutant and control seedlings. Hydrolyzed extracts from root (**a**), stem (**b**) and leaf (**c**) samples were compared to authentic standards. 'WT' represents the chromatogram of a control seedling. Numbers in the upper left of each panel represent mutant line numbers. SA and SB denote the peaks for soyasapogenol A and B, respectively.

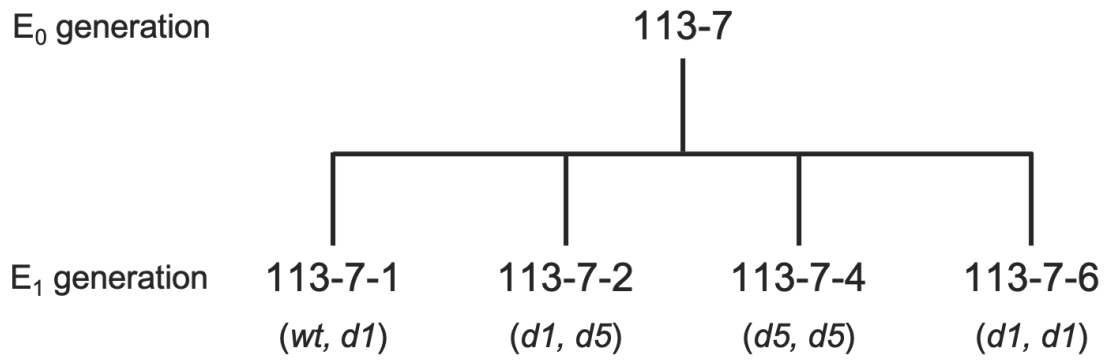


Figure 2.7 Mutant alleles in the *GmBAS1* loci in the E_1 generation derived from 113-7.

Italic letters in parentheses denote mutation types: e. g., *d1* (a single-nucleotide deletion).

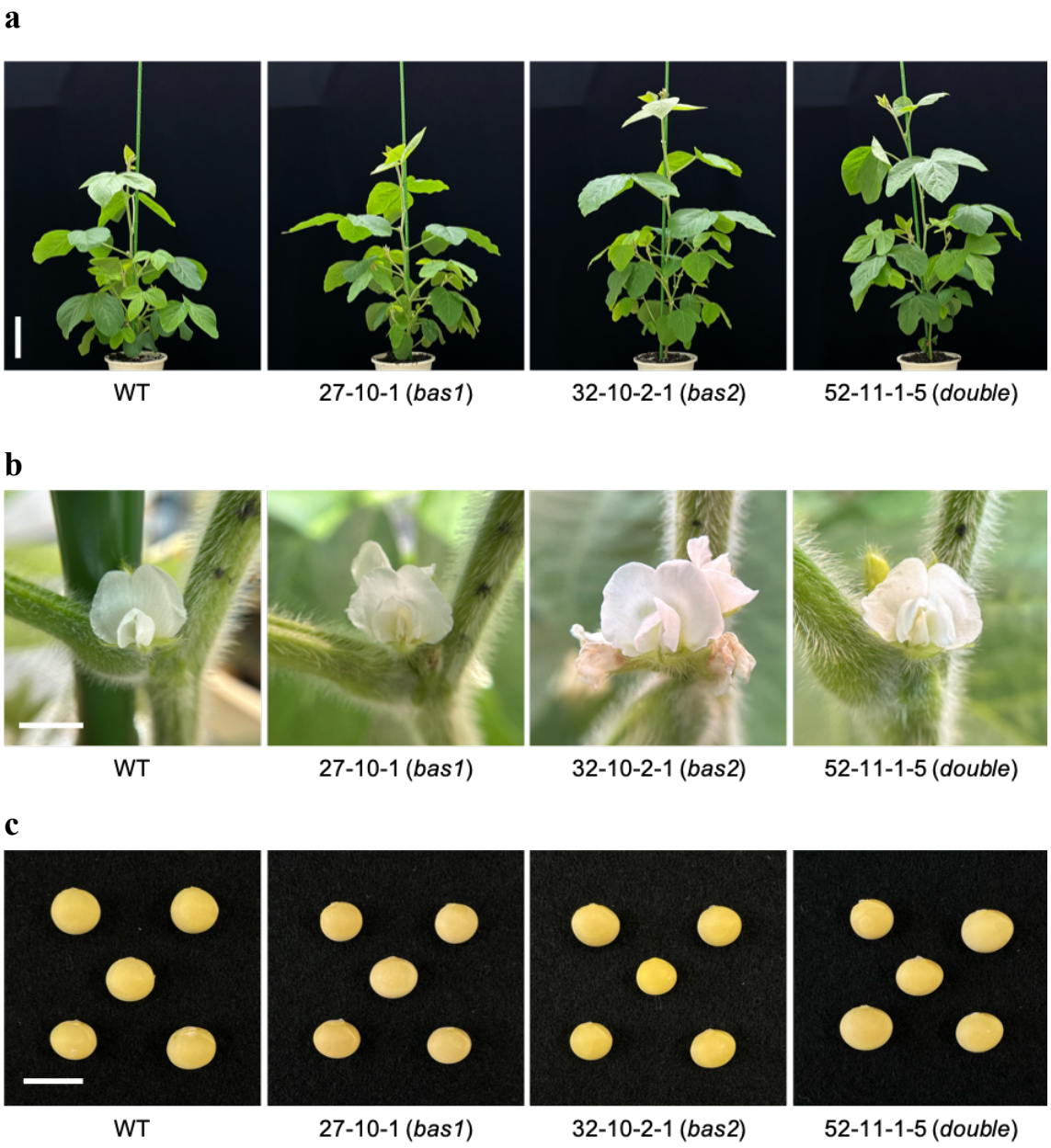


Figure 2.8 Morphological characteristics of representative mutant and control plants.

a Whole plants (1 month after sowing). Scale bar = 10 cm

b Flowers. Scale bar = 0.5 cm

c Mature seeds. Scale bar = 1 cm

'WT' indicates the image of a control plant. Numbers bottom of each panel indicate mutant line number.

Table 2.1 Primer sequences used in this study.

Primer name	Primer sequence (5'-3')	Use of amplicon
GmBAS1_qRT_Fw1	AGGGATGTGGAGGCTGAAGA	qRT-PCR
GmBAS1_qRT_Rv1	TCTGCCTCCCCAACGAAATTGT	qRT-PCR
GmBAS2_qRT_Fw1	AAGCGAAGGAGAGGATGTGGA	qRT-PCR
GmBAS2_qRT_Rv1	GTTTGTGCTGAATATGTATGCCTC	qRT-PCR
Bic-C2_Fw	GCTATCAGGGACAATGGCAC	qRT-PCR
Bic-C2_Rv	CCATGACTTTTTAGGAGCTGGG	qRT-PCR
GmBAS1 Fw	GATCAGGGATGTGGAGGCTG	CAPS and sequencing analysis
GmBAS1 Rv	ACTCCTAAGCAAAACTGGAATCAA	CAPS and sequencing analysis
GmBAS2 Fw	GTGGAGGCTGAAGATAGCAGA	CAPS and sequencing analysis
GmBAS2 Rv	ATGTGTCGGGAAACAAGCAT	CAPS and sequencing analysis
off1_Fw1	TAGCAGATGGTGGAAAAGACCC	Sequencing analysis of off-target
off1_Rv1	TGGCCAATGGCTTGATGTCAC	Sequencing analysis of off-target
off2_Fw1	GTGGAGGCTGAAGATAGCAGA	Sequencing analysis of off-target
off2_Rv1	GGTTTGATGCAAGTATGTCACAA	Sequencing analysis of off-target
off3_Fw1	CGAGGAATGGTGAACCTCGT	Sequencing analysis of off-target
off3_Rv1	AGTAGGTCACCGCTTTGCTT	Sequencing analysis of off-target
off4_Fw1	TACGTCGGAGAGTGGAAGGGT	Sequencing analysis of off-target
off4_Rv1	CCACAAGGACTCTTCTGCAACT	Sequencing analysis of off-target

Table 2.2 Mutation frequency at the target loci in E_o and E_1 generations

No. of bombarded embryonic axes	No. of analyzed E_0 plants	No. of mutants in E_0 generation				No. of plants analyzed in E_1 generation	No. of mutants in E_1 generation			
		GmBAS1	*GmBAS2*	Both loci	Total		*GmBAS1*	*GmBAS2*	Both loci	Total
1697	1467	80	10	48	138	36	7	2	7	16

Table 2.3 Mutation frequency in the E_1 generation.

E_0 plant No.	Mutated loci in the E_0 generation	No. of analyzed E_1 seeds	No. of mutants in the E_1 generation
23-13	*GmBAS1*	20	0
27-10	*GmBAS1*	40	40
32-10	*GmBAS2*	30	23
39-2	*GmBAS1*	5	0
39-7	Both loci	4	0
48-3	*GmBAS1*	1	0
50-3	*GmBAS1*	2	2
52-9	Both loci	1	1
52-11	Both loci	2	2
53-5	*GmBAS1*	1	0
54-9	*GmBAS2*	1	0
56-4	*GmBAS1*	12	7
57-3	Both loci	3	0
58-2	*GmBAS1*	1	1
59-11	*GmBAS1*	2	0
77-15	Both loci	18	0
89-2	*GmBAS1*	8	0
95-11	*GmBAS1*	8	0
99-3	*GmBAS2*	8	8
101-9	*GmBAS1*	8	0
102-10	*GmBAS1*	16	1
103-9	*GmBAS1*	8	0
104-4	Both loci	18	18
106-5	Both loci	16	2
106-7	Both loci	2	0
108-3	Both loci	8	0
110-14	*GmBAS1*	8	0
110-15	Both loci	15	8
113-7	Both loci	15	15
114-8	Both loci	2	1
117-2	*GmBAS1*	8	0
118-4	*GmBAS1*	8	0

120-7	*GmBAS1*	8	0
121-8	*GmBAS1*	7	3
125-10	Both loci	7	0
129-6	*GmBAS1*	2	2

Table 2.4 Mutation patterns detected in this study.

Genotype	Mutations detected in targeted loci	
	GmBAS1	*GmBAS2*
bas1	*i1(A), i1(T), i1(G), d1, d3, d4, d5*	-
bas2	-	*d5, d7*
double	*i1(A), i1(T), d1, d4, d5, d6, d7*	*i1(A), i1(T), d1, d2, d4, d10*

Italic letters denote mutation types: e. g., *d1* (a single-nucleotide deletion); *i1* (a single-nucleotide insertion); *(A)* (adenine insertion).

Table 2.5 Evaluation of off-target mutations in this study.

Locus name	Putative off-target locus	Putative off-target sequence[a]	No. of mismatch bases	No. of examined mutants[b]	No. of mutant with off-target mutations
Glyma.03G121300.4	exon	TGGGAGTTTGATCCTGAGGCAGG	1	10	0
Glyma.03G121600.1	exon	TGGGAATTTGATCCTGAGGCAGG	2	10	0
Glyma.08G027000.2	exon	TGGGAGTTTGATCCTAATGCTGG	2	10	0
Glyma.04G206100.1	exon	TGGGAGCTTGATGGTGAAGCTGG	3	10	0

The off-target loci were selected by the gRNA design program of CRISPR-P (http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/CRISPR), CHOPCHOP (https://chopchop.cbu.uib.no/), CCTop (https://cctop.cos.uni-heidelberg.de:8043/).

[a]Red and underlined nucleotides denote mismatch nucleotides to the target site of gRNA and PAM sequence, respectively.

[b]Representative E_1 mutants (27-10-1, 32-10-2, 50-3-2, 52-11-1, 52-11-2, 99-3-2, 104-4-2, 113-7-4, 113-7-6, and 113-7-10) were selected as plant materials for this study.

Table 2.6 Genotypes of representative mutants used in HPLC analysis

Line No.	Genotypes		Haplotype	Analyzed tissue in HPLC analysis
	GmBAS1 locus	*GmBAS2* locus		
WT	*wt*	*wt*	*wt*	seed, root, stem, leaf
27-10-1	*i1(A)*	*wt*	*bas1*	seed, root, stem, leaf
50-3-2	*d5*	*wt*	*bas1*	seed, root, stem, leaf
104-4-2	*d1*	*wt*	*bas1*	seed
32-10-2-1	*wt*	*d5*	*bas2*	seed, root, stem, leaf
52-11-2-13	*wt*	*i1(T)*	*bas2*	seed, root, stem, leaf
99-3-2	*wt*	*d7*	*bas2*	seed
52-11-1-5	*i1(T)*	*i1(T)*	*double*	seed, root, stem, leaf
113-7-4	*d5*	*d4*	*double*	seed
113-7-6	*d1*	*d2*	*double*	seed
113-7-10	*d5*	*d4*	*double*	root, stem, leaf

Italic letters denote mutation types: e. g., *d1* (a single-nucleotide deletion); *i1* (a single-nucleotide insertion); *(A)* (adenine insertion); *wt* (no mutation).

Table 2.7 Soyasapogenols content in mutant and control seeds

Line No.	Haplotype	SA content (mg/100 g DW)	SB content (mg/100 g DW)
WT	*wt*	88.9 ± 29.9	166.2 ± 42.0
27-10-1	*bas1*	nd	nd
50-3-2	*bas1*	nd	nd
104-4-2	*bas1*	nd	nd
32-10-2-1	*bas2*	92.6 ± 8.5	163.6 ± 18.4
52-11-2-13	*bas2*	75.4 ± 11.0	150.9 ± 20.2
99-3-2	*bas2*	76.2 ± 6.2	155.1 ± 15.5
52-11-1-5	*double*	nd	nd
113-7-4	*double*	nd	nd
113-7-6	*double*	nd	nd

SA and SB denote soyasapogenol A and B, respectively. Italic letters denote haplotype. Values represents the mean ± SD (n = 5). nd, not detected.

Table 2.8 Soyasapogenols content in mutant and control seedlings

Plant tissue	Line No.	Haplotype	SA content (mg/100 g DW)	SB content (mg/100 g DW)
root	WT	*wt*	105.4 ± 16.5	377.4 ± 33.7
	27-10-1	*bas1*	nd	nd
	50-3-2	*bas1*	nd	nd
	32-10-2-1	*bas2*	92.4 ± 20.6	360.9 ± 61.1
	52-11-2-13	*bas2*	102.8 ± 18.7	386.6 ± 42.7
	52-11-1-5	*double*	nd	nd
	113-7-10	*double*	nd	nd
stem	WT	*wt*	14.3 ± 10.1	124.2 ± 41.1
	27-10-1	*bas1*	nd	nd
	50-3-2	*bas1*	nd	nd
	32-10-2-1	*bas2*	26.5 ± 3.2	245.1 ± 114.8
	52-11-2-13	*bas2*	25.1 ± 15.7	278.9 ± 93.5*
	52-11-1-5	*double*	nd	nd
	113-7-10	*double*	nd	nd
leaf	WT	*wt*	nd	150.0 ± 30.0
	27-10-1	*bas1*	nd	nd
	50-3-2	*bas1*	nd	nd
	32-10-2-1	*bas2*	nd	225.6 ± 46.4**
	52-11-2-13	*bas2*	nd	237.6 ± 33.2**
	52-11-1-5	*double*	nd	nd
	113-7-10	*double*	nd	nd

SA and SB denote soyasapogenol A and B, respectively. Italic letters denote haplotype. Values are presented as mean ± SD ($n = 4$). "nd" denotes values that were not detected. Asterisks indicate significant differences between the mutant lines and the control plant, based on an unpaired two-tailed Student's *t*-test, (*$p < 0.05$, **$p < 0.01$).

CHAPTER 3

Functional Characterization of the Soybean Knockout Mutant of a β-Amyrin Synthase Gene Demonstrates the Role of Soyasaponins in Antioxidant Capacity of Seeds and Rhizosphere Bacterial Interactions

3.1 Background and Purpose

Triterpenoid saponins are widely recognized as defense compounds against pathogens and herbivores because of their amphipathic properties, which contribute to membrane perturbation (Augustin et al. 2011). Avenacin, a triterpenoid saponin of *Avena strigosa*, is known as an antimicrobial phytoprotectant (Papadopoulou et al. 1999). However, the remarkable diversity in their accumulation patterns across species, tissues and developmental stages in higher plants suggests that each compound has distinct and finely tuned physiological roles (Moses et al. 2014). Previous studies employing mutants have demonstrated that endogenous triterpenoids are linked to diverse physiological functions. For example, overaccumulation of triterpene scaffold biosynthesis results in stunted growth of the root in *Ar. thaliana* and *Av. strigosa* (Field and Osbourn 2008; Field et al. 2011; Kemen et al. 2014). Knockout mutant of the *marneral synthase 1* gene, encoding an oxidesqualene cyclase in *Ar. thaliana*, show round-shaped leaves, late flowering, and delayed embryogenesis (Go et al. 2012). In *L. japonicus*, silencing of oxidesqualene cyclase genes affects rapid nodulation, late flowering, and stunted root phenotype (Delis et al. 2011; Krokida et al. 2013). Mutations of genes involved in glycosylation of triterpenoid saponin biosynthesis resulted in reduced germination rates in *M. truncatula*, severe morphological defects in leaves, flowers and fruits in *Solanum lycopersicum* (Naoumkina et al. 2010; Carelli et al. 2011; Itkin et al. 2011). Moreover, the accumulation of incompletely glycosylated saponins are associated with abnormal development of the root epidermal cell layers in *Av. strigosa* and *M. truncatula* (Mylona et al. 2008; Pollier et al. 2013)

Diverse and unique functions of soyasaponins have been documented. For example, DDMP saponins exhibit stronger antioxidant activity than group B saponins, which possess the same sugar moieties (Yoshiki and Okubo 1995; Yoshiki et al. 2001). DDMP saponins stimulate root growth in *Lactuca sativa*, *Chrysanthemum coronarium*, *Brassica juncea*, *Phleum pratense*, *Lolium multiflorum*, *Trifolium repens*, *Brassica campestris*, *Medicago sativa*, *Astragalus sinicus* and *Cryptotaenia japonica* (Tsurumi and Tsujino 1995; Tsurumi and Wada 1995). In contrast, soil treatment with group

B saponins altered the bacterial community and markedly enriched plant growth-promoting rhizobacteria, such as *Novosphingobium* (Fujimatsu et al. 2020). Thus, DDMP saponins may function internally within plants as components of the antioxidant system, whereas group A and/or group B saponins may act externally as communication mediators between soybeans and microbes. Nevertheless, these complex physiological and ecological processes are likely influenced by numerous other metabolites, and the quantitative contribution of soyasaponins remains largely unclear.

In this chapter, to determine the impact of soyasaponin deficiency to seed oxidation level, the 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical-scavenging capacity was assessed in mutant soybean plants generated in chapter 2. In addition, to investigate the rebalance with other antioxidants, transcriptome analysis was performed using RNA-seq and with quantification of total polyphenols and isoflavones in hypocotyls. Furthermore, the impact of *GmBAS1* mutation on the rhizosphere bacterial communities was evaluated through 16S rRNA gene amplicon sequencing.

3.2 Materials and Methods

3.2.1 Plant materials and growth conditions.

Soybean *bas1* mutant lines were generated using the iPB-RNP methods via the CRISPR/Cas9 system as described in chapter 2 (Asa et al. 2025). Three independent mutant lines, designated *bas1-1*, *bas1-2*, and *bas1-3*, each carrying a different mutant allele, were used in this study. The soybean variety GL3510, a commercial line from Kanematsu Corp. was used as the control. For germination, seeds were placed on wet paper towels and incubated for 3–4 d at 25°C in the dark. The seedlings were transferred to plastic pots filled with a mixture of potting soil (Sumitomo Forestry Landscaping Co., Ltd.) for sowing and culture soil (Katakura & Co-op Agri Corp.) in a 1:3 ratio and grown in a closed greenhouse under a 12-hour photoperiod at 25°C. For rhizosphere analysis using 16S rRNA gene amplicon sequencing, seedlings were grown on a mixture of potting and culture soils in a 1:7 ratio without fertilizer.

3.2.2 Evaluation of DPPH radical-scavenging capacity in mature seeds.

Three dried seeds from each individual were pooled and ground using a Multi-Beads Shocker (MB1200, Yasui Kikai Co.). Fine powder (100 mg) was mixed with 500 µL of 80% (v/v) aqueous ethanol and sonicated for 1 min. The mixture was vortexed for 3 h at 20°C in the dark. After centrifugation at 21,500 ×g for 10 min at 10°C, the supernatants were transferred into 1.5 mL tubes. After centrifugation at 21,500 ×g for 5 min at 10°C, the supernatant was used for subsequent serial

dilutions. Extracts from the *bas1* mutant and control seeds were diluted 1-, 2-, and 4-fold and 2-, 4-, and 8-fold with 80% (v/v) aqueous ethanol, respectively. Diluted solutions were used to assess the radical-scavenging capacity.

Analysis was performed using a DPPH Antioxidant Assay Kit (DOJINDO LABORATORIES Inc.) following the manufacturer's instructions. A series of standard solutions of Trolox (0, 40, 60, and 80 μg/mL) were prepared by diluting a 100 μg/mL solution with ethanol. The DPPH working solution was prepared by dissolving the supplied DPPH Reagent in 10 mL of ethanol. One hundred microliters of DPPH working solution was mixed thoroughly with 20 μL of sample or standard solution and 80 μL of assay buffer. After incubation for 30 min at 25°C in the dark, DPPH absorbance at 517 nm was measured using a Multi-Skan SkyHigh (Thermo Fisher Scientific Inc.). The DPPH radical-scavenging percentage was calculated using the mean absorbance of the three technical replicates. Trolox Equivalent Antioxidant Capacity (TEAC) was determined by comparing the concentration of Trolox standard to that of each sample required to scavenge 50% of DPPH radicals. The values are expressed as micrograms of Trolox equivalents (TE). Tests of significance among the means of the data were performed using Dunnett test. Statistical significance was set at $p < 0.05$.

3.2.3 RNA sequencing (RNA-seq) analysis.

Total RNA was extracted from the hypocotyls and cotyledons of developing seeds using the LiCl precipitation procedure described previously (Adachi et al. 2021). For each biological replicate, hypocotyls and cotyledons were collected from three individual seeds and pooled separately. Extraction was performed using three biological replicates for both the *bas1* mutant line and control plants.

Sample preparation and paired-end sequencing using a NovaSeq 6000 (Illumina, Inc.) were performed by Rhelixa, Inc. Raw FASTQ reads were processed using the RaNA-seq web server (Prieto and Barrios 2020). Briefly, the reads were filtered based on quality, and the remaining reads were mapped to the soybean reference genome (*G. max* Wm82.a2.v1). Gene expression levels were quantified as transcripts per million (TPM). Differential expression analysis between the *bas1* mutant and control plants was performed using DESeq2 within the RaNA-seq pipeline. Genes with an adjusted *p*-value (False Discovery Rate (FDR)) < 0.05 were considered differentially expressed genes (DEGs).

Functional enrichment analysis of DEGs was performed using g:Profiler (version e113_eg59_p19_f6a03c19) via its web interface (Raudvere et al. 2019). The list of DEGs was

submitted to the g:GOSt tool to identify enriched pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The significance threshold was set at FDR < 0.05, and the *G. max* genome was used as the background for the analysis.

Descriptions of the DEGs were obtained using PhytoMine, a data-mining tool within the Phytozome database (ver. 13) (https://phytozome-next.jgi.doe.gov/phytomine/begin.do). The gene IDs of the selected DEGs were used as queries after replacing the prefix 'GLYMA_' with 'Glyma.' (e.g., GLYMA_08G350800 to Glyma.08G350800) to retrieve the corresponding functional descriptions of the genes. Gene descriptions were manually curated by integrating the results from the queries against both the Wm82.a2.v1 and Wm82.a4.v1 reference annotations, because the Wm82.a4.v1 annotation provided more comprehensive functional information.

3.2.4 Extraction and quantification of total polyphenols by the Fast Blue BB methods.

Dried hypocotyls were dissected from three to seven seeds and ground in bulk, as described above. Ten milligrams of fine powder were mixed with 1 mL of 80% (v/v) aqueous methanol and vortexed for 3 h at room temperature in the dark. After centrifugation at 23,200 ×g for 10 min at 4°C, 250 µL of the supernatants were transferred into 1.5 mL tubes. The supernatant was then dried by evaporation. The pellets were dissolved in 50 µL of 80% (v/v) aqueous methanol and centrifuged. The supernatant was used for subsequent analysis.

The total polyphenol content was analyzed using the Fast Blue BB (FBBB) method (Medina 2011; Lester et al. 2012) with some modifications. Serial dilutions of the gallic acid standard (0, 10, 20, 30, 40, and 50 mg/L) were prepared in distilled water to create a calibration curve. Next, 20 µL of 0.1% (w/v) FBBB [4-benzoylamino-2,5-dimethoxybenzenediazonium chloride hemi (zinc chloride) salt was mixed with 200 µL of 20-fold diluted samples or standard solutions. After adding 20 µL of 5% (w/v) NaOH, the mixture was incubated for 2:30 h at 30°C in the dark. Absorbance was measured at 420 nm using PowerScan HT (BioTek Instruments Inc.). The total polyphenol content of each sample was calculated from the mean absorbance of the two replicates. The values are expressed as milligrams of gallic acid equivalent (GAE). Tests of significance among the means of the data were performed using Dunnett test. Statistical significance was set at $p < 0.05$.

3.2.5 Extraction and HPLC analysis of isoflavones.

Isoflavones were extracted from the dried hypocotyls. Twenty milligrams of fine powder prepared as described above was mixed with 600 µL of 70% (v/v) aqueous ethanol. The mixture was vortexed in the dark for 30 min at room temperature. After centrifugation at 21,500 ×g for 10 min at 20°C, the supernatants were transferred into 1.5 mL tubes. This procedure was performed twice, and the two supernatants were combined. After centrifugation at 21,500 ×g for 10 min at 20°C, 100 µL of the solution was transferred into an HPLC vial insert.

The analysis was performed using an HPLC system (Primaide, Hitachi High-Tech Corp.) with an Inertsil ODS-3 C18 reverse-phase column (4.6 mm × 250 mm, 5µm, GL Sciences Inc.). The column temperature was maintained at 40°C, and separation was performed under gradient conditions. Solvent A was acetonitrile (0.1% acetic acid), and solvent B was water (0.1% acetic acid). The gradient program was as follows: linear 85–65% B, 0–30 min; linear 65–85% B, 30–31 min; isocratic at 85% B, 31–35 min. The injection volume of each sample was 10 µL, and the flow rate was 1 mL/min. Isoflavones were detected at 254 nm. Daizin, malonyl-daidzin, genistin, malonyl-genistin, glycitin, and malonyl-glycitin were purchased from Nagara Science Co., Ltd. and dissolved in 70% aqueous ethanol. The isoflavone content was determined based on the ratio of the area of the peaks to that of the standards. Tests of significance among the means of the data were performed using Dunnett test. Statistical significance was set at $p < 0.05$.

3.2.6 DNA extraction from soil samples, 16S rRNA gene amplicon sequencing, and bioinformatic analysis.

Soil samples were collected 50 d after sowing during the early pod-filling stages. Soil loosely attached to the plant roots was removed using vigorous shaking. The remaining root-adhering soil was then passed through a 1.5-mm sieve to remove small root fragments and collected in a sterile 15 mL centrifuge tube (Corning Inc.). Soil was also collected from pots in which no plants were grown. All samples were immediately placed on ice until they were transferred to 2 mL homogenizer tubes in the laboratory. The samples were stored at –30°C until analysis using 16S amplicon sequencing. In this study, root-adhering soil is referred to as rhizosphere soil, whereas soil collected from pots without plants is referred to as bulk soil.

DNA extraction from soil samples, amplicon sequencing of the 16S rRNA gene, and subsequent bioinformatics analysis were performed by the Bioengineering Lab. Co., Ltd. The vendor's protocol was as follows: Each soil sample was lyophilized in a freeze dryer (VD-250R, TAITEC Co., Ltd.) and

ground into a fine powder using a Multi-Beads Shocker. DNA was extracted from the powder using Lysis Solution F (Nippon Gene Co., Ltd.) and a Lab-Aid824s DNA Extraction Kit (Zessan Biotech Co., Ltd.). The concentration of the extracted DNA was quantified using a Synergy LX microplate reader (Agilent Technologies, Santa Clara) and the QuantiFluor dsDNA System (Promega Corp.).

A DNA library was prepared using a two-step-tailed PCR method. In the first PCR, the V4 region was amplified in a 10-μL reaction mixture containing template DNA, 2 × PCR buffer for KOD FX Neo (TOYOBO, Osaka, Japan), 400 nM forward and reverse primers, 400 μM dNTPs, and 0.2 U of KOD FX Neo (TOYOBO). Non-barcoded V4 forward and reverse primer mixtures were used (Table 3.1). The conditions for the first PCR were as follows: denaturation at 94°C for 2 min; 30 cycles of 98°C for 10 s, 50°C for 30 s, and 68°C for 30 s; and a final extension at 68°C for 7 min. The PCR products were purified using VAHTS DNA Clean Beads (Vazyme Biotech Co., Ltd.).

In the second PCR, products from the first PCR were amplified and purified using the same procedure as the first PCR, except that the reaction mixture contained the first PCR product and 250 nM barcoded forward and reverse primers (Table 3.1). The conditions for the second PCR were as follows: denaturation at 94°C for 2 min; 10 cycles at 98°C for 10 s, 60°C for 30 s, and 68°C for 30 s; and a final extension at 68°C for 2 min. The quality and quantity of the purified PCR products were checked on a Fragment Analyzer using the dsDNA 915 Reagent Kit (Agilent Technologies). Paired-end sequencing (2 × 300 bp) was performed using an Illumina NextSeq 1000 system with a NextSeq 1000/2000 P1 XLEAP-SBS Reagent kit.

Sequence data analysis was performed using a pipeline combining the FASTX Toolkit (v0.0.14), Sickle (v1.33), FLASH (v1.2.11), and QIIME 2 (v2024.10). First, raw reads were processed using the FASTX Toolkit to extract reads that matched the primer sequences, after which the primer sequences were removed. The reads were then quality-trimmed using Sickle. Bases with a quality score below 20 were removed, and reads shorter than 130 bp were discarded. Paired-end reads were subsequently merged using FLASH. Merged quality-filtered reads were processed using QIIME 2. The `dada2` plugin was used for denoising and chimera removal to construct amplicon sequence variants (ASVs). Taxonomic assignment for each ASV was performed using a pre-trained naïve Bayes classifier against the EzBioCloud 16S database.

Alpha and beta diversity analyses were performed using the diversity plugin in QIIME 2. Alpha diversity was assessed using the Shannon index, and statistical significance among groups was tested using the Kruskal–Wallis test. Beta diversity was assessed using weighted UniFrac distances and

visualized using principal coordinate analysis (PCoA). Permutational multivariate analysis of variance (PERMANOVA) was used to test for significant differences in community composition between groups. P-values for pairwise PERMANOVA were adjusted for multiple comparisons using the Benjamini-Hochberg (FDR) method, and these adjusted p-values are referred to as *q*-values. Linear discriminant analysis (LDA) effect size (LEfSe) (v1.0.8) was performed to identify differentially abundant taxa between the *bas1* and control rhizospheres. A significant difference was determined using the default parameters at a *q*-value of < 0.05 for the Kruskal–Wallis test and a logarithmic LDA score > 2.0.

3.2.7 Extraction and GC-MS analysis of phytosterols.

Ten milligrams of fine seed powder prepared as described above was mixed with 1 mL of methanol and 20 µL of 100 ppm uvaol (Extrasynthese S. A.; added as the internal standard). The mixture was vortexed for 15 s and sonicated for 60 min. After centrifugation at 13,040 ×g for 5 min, the supernatant was evaporated at 37°C for 60 min. The residue was hydrolyzed with 1 mL of methanol and 1 mL of 4 M HCl at 80°C for 1 h. After cooling, 2 mL of n-hexane/ethyl acetate (1:1) was added, followed by vortexing and centrifugation at 20,380 ×g for 3 min. The organic layer was then collected and evaporated. The residue was dissolved in 300 µL of methanol, and 50 µL was transferred to a GC-MS vial and evaporated. The final residue was derivatized with 50 µL of N-methyl-N-(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich Co. LLC) at 80°C for 30 min.

GC-MS analysis was performed using a 7890B GC system coupled with a 5977A MSD (Agilent Technologies) and HP-5MS capillary column (30 m × 0.25 mm, 0.25 µm, Agilent Technologies). The Inlet, MS source, and MS quadrupole temperatures were set at 250°C, 230°C, and 150°C, respectively. The oven temperature was set to 80°C for 1 min, increased to 300°C at 20 °C/min, and maintained at 300°C for 18 min. Helium was used as the carrier gas at a flow rate of 1.0 mL/min. The injection volume was 1 L. Mass spectra were recorded over an m/z range of 50–750. The compounds were identified by comparing their retention times and mass spectra with those of authentic standards (e.g., campesterol, stigmasterol, and β-sitosterol; Sigma-Aldrich Co. LLC). The sterol peak areas were normalized to that of the internal standard, uvaol. Relative quantification was performed by comparing each normalized value to that of the reference sample (WT1).

3.3 Result

3.3.1 Evaluation of DPPH antioxidant capacity in *bas1* mutant and control seeds.

To determine the effect of the loss of *GmBAS1* function on the antioxidant capacity of soybean seeds, DPPH radical-scavenging capacity was evaluated. The DPPH assay is a simple and widely used method for evaluating antioxidant capacity, and DDMP soyasaponins have been reported to exhibit DPPH radical-scavenging activity (Yoshiki et al. 2001). Thus, we examined the radical-scavenging capacities of extracts from the three mutant lines and control seeds. The control extracts exhibited 1.49 ± 0.09 μg TE/g dry weight (DW), whereas the *bas1-1*, *bas1-2*, and *bas1-3* extracts showed 0.58 ± 0.07 μg TE/g DW, 0.66 ± 0.04 μg TE/g DW, and 0.57 ± 0.05 μg TE/g DW, respectively (Figure 3.1). The DPPH antioxidant capacity of *bas1* mutant seeds was significantly lower than that of the control seeds (Dunnett test; $p < 0.01$).

3.3.2 Gene expression analysis of metabolic pathways in *bas1* mutant and control seeds.

RNA-seq was used to analyze transcriptome profiles and assess the differential expression of genes associated with the metabolic pathways disrupted by the loss of *GmBAS1* function in soybean seeds. The composition and content of soyasaponins differ between hypocotyl and cotyledon (Tsuno et al. 2018). Therefore, total RNA was extracted separately from hypocotyls and cotyledons. In the hypocotyls, a total of 43,056 genes were subjected to statistical tests. Among these, 2,211 DEGs were identified, of which 690 were upregulated and 1,521 were downregulated compared with the control (Figure 3.2.a). In the cotyledons, a total of 41,298 genes were subjected to statistical tests and 204 DEGs were detected, including 87 upregulated and 117 downregulated DEGs relative to the control (Figure 3.2.b).

Functional enrichment analysis revealed that the DEGs in the hypocotyls were enriched in various metabolic pathways (Table 3.2). Although the analysis revealed significant transcriptional changes in genes related to motor proteins and photosynthesis, the primary objective of this study was to elucidate the function of soyasaponins as specialized metabolites. Therefore, I focused on the changes observed in other metabolic pathways. All DEGs associated with stilbenoid, diarylheptanoid, and gingerol biosynthesis were included in the flavonoid biosynthesis pathway. Similarly, all DEGs enriched in sesquiterpenoid and triterpenoid biosynthesis, except for GLYMA_08G350800, which encodes β-amyrin 24-hydroxylase (CYP93E1), were also involved in steroid biosynthesis. Given

these results, DEGs that are predicted to encode enzymes involved in flavonoid and steroid biosynthetic pathways were manually curated (Tables 3.3 and 3.4). In the flavonoid biosynthetic pathway, 25 DEGs were listed (Table 3.3), of which 23 were downregulated and two (GLYMA_17G173200 and GLYMA_19G254600) were upregulated in the *bas1* mutant compared to the control. The expression of GLYMA_17G173200, which encodes dihydrokaempferol 4-reductase (DFR), was activated in the *bas1* mutant hypocotyls but was absent in the control hypocotyls. Eleven DEGs were listed in the steroid biosynthesis pathway (Table 3.5). GLYMA_06G020700, which encodes sterol 24-C-methyltransferase (SMT2), was one of the few genes upregulated in *bas1* mutant hypocotyls compared to the control.

3.3.3 Determination of total polyphenol and isoflavones in *bas1* mutant and control seeds.

To further examine the effect of the loss of *GmBAS1* function on flavonoid biosynthesis, I determined the total polyphenol content in the hypocotyls of the two mutant lines and control seeds using the FBBB method. This method uses an azo-coupling reaction and enables the quantification of phenolic compounds possessing at least one unsubstituted ortho- or para-position on the hydroxyl group (Medina 2011). Unlike the Folin–Ciocalteu method, the FBBB method allows the determination of the total polyphenol content without the influence of non-phenolic antioxidants. The concentration of total polyphenols in the control hypocotyls was 5.52 ± 1.39 µg GAE/mg DW (Figure 3.3). The concentrations of total polyphenols in *bas1-1* and *bas1-2* hypocotyls were 8.36 ± 1.04 µg GAE/mg DW and 8.72 ± 0.80 µg GAE/mg DW, respectively (Figure 3.3). Accordingly, the total polyphenol content in *bas1* mutant hypocotyls was significantly higher than that in control hypocotyls (Dunnett test, $p < 0.05$).

Transcriptome analysis indicated that the biosynthetic pathway from naringenin to flavanols was activated by the upregulation of DFR and ANR genes in the *bas1* mutant hypocotyl. Isoflavones constitute an important class of flavonoids in soybean seeds, and their biosynthesis competes with the pathway activated in the *bas1* mutant, using naringenin as a substrate. Therefore, to assess whether these transcriptional changes influenced isoflavonoid biosynthesis in the *bas1* mutant, the isoflavone content in the hypocotyls of the two mutant lines and control seeds was measured using HPLC. Peaks corresponding to daidzin, malonyl-daidzin, genistin, malonyl-genistin, and glycitin were detected in the extracts from the control hypocotyls. Because the peak for malonyl-glycitin overlapped with that for malonyl-daidzin and could not be reliably separated, the peak for malonyl-daidzin was considered

to represent the combined amount of both compounds. The contents of daizin, malonyl-daizin, genistin, malonyl-genistin, and glycitin were 1.88 ± 0.12 mg/g DW, 3.57 ± 0.34 mg/g DW, 0.42 ± 0.03 mg/g DW, 1.78 ± 0.11 mg/g DW, and 2.08 ± 0.11 mg/g DW, respectively (Figure 3.4). In the extracts from *bas1* mutant hypocotyls, peaks for the five isoflavones were also detected. The contents of each isoflavone in *bas1-1* hypocotyls were 1.19 ± 0.26 mg/g DW, 3.36 ± 0.55 mg/g DW, 0.26 ± 0.05 mg/g DW, 1.02 ± 0.68 mg/g DW, and 2.24 ± 0.46 mg/g DW, respectively (Figure 3.4). The contents of each isoflavone in *bas1-2* hypocotyls were 1.04 ± 0.07 mg/g DW, 3.10 ± 0.17 mg/g DW, 0.21 ± 0.03 mg/g DW, 0.9 ± 0.47 mg/g DW, and 2.24 ± 0.27 mg/g DW, respectively (Figure 3.4). In both *bas1* mutant lines, the levels of daidzin and genistin in the hypocotyls were significantly reduced compared with those in the control hypocotyls (Dunnett test, $p < 0.01$; Figure 3.4.a and b). In addition, the total isoflavone content in the control, *bas1-1*, and *bas1-2* hypocotyls were 9.73 ± 0.15 mg/g DW, 8.07 ± 1.09 mg/g DW, and 7.49 ± 0.33 mg/g DW, respectively. The total isoflavone content in *bas1-2* hypocotyls was also significantly lower than that in the control hypocotyls (Dunnett test, $p < 0.05$; Figure 3.4.c).

3.3.4 Evaluation of the bacterial community in the rhizosphere of *bas1* mutant and control plants.

To assess the influence of the loss of *GmBAS1* function on the soybean rhizosphere, I analyzed the bacterial communities in the rhizosphere of *bas1* mutants and control plants using amplicon sequencing of the 16S rRNA gene. Soyasaponins are secreted throughout the plant life cycle, and their concentration in root exudates is maximized at the V3 stage of hydroponic culture (Tsuno et al. 2018). Under soil cultivation, soyasaponin levels in the rhizosphere progressively increase during the vegetative stages and reach their highest levels during the flowering stage (Fujimatsu et al. 2020). Based on these findings, the impact of the loss of *GmBAS1* function is expected to be the greatest during the early pod-filling stage. Therefore, soil samples were collected 50 d after sowing.

Bacterial diversity in each sample was estimated using the Shannon index. The Shannon indices of bulk soils, control rhizosphere, and *bas1* rhizosphere soils were 8.55 ± 0.16, 8.70 ± 0.08, and 8.66 ± 0.19, respectively (Figure 3.5.a). No significant differences were observed in the Shannon index among the bulk, WT, and *bas1* groups (Kruskal–Wallis test, $p < 0.05$). Next, to visualize the differences in bacterial community composition among the sample groups, PCoA was performed based on the weighted UniFrac distance matrix. The PCoA plot showed that the bulk soil samples formed a distinct cluster, separated from the rhizosphere soil samples (Figure 3.5.b). PERMANOVA

revealed a significant difference in the overall bacterial community composition among the three groups (pseudo-F = 2.64, $p < 0.01$). Subsequent pairwise comparisons showed that the community composition of the bulk soil differed significantly from that of the *bas1* ($q = 0.0495$) and control ($q = 0.0495$) rhizosphere soils. In contrast, no significant differences were observed between the *bas1* and control groups ($q = 0.617$). To investigate whether the *bas1* mutation influenced specific soil bacterial taxa, differences in the enrichment of bacterial families and genera between the *bas1* and control rhizospheres were analyzed using LEfSe. A total of 289 and 628 taxa at the family and genus levels, respectively, were subjected to the analysis (Figure 3.6). The relative abundances of 10 families and 17 genera were significantly different between the *bas1* and control rhizospheres (Figure 3.7). The relative abundance of *Novosphingobium* was comparatively less abundant.

3.4 Discussion

Various compounds in soybean seeds, including soyasaponins, have been reported to possess antioxidant activity (Yoshiki and Okubo 1995; Yoshiki et al. 2001; Kähkönen and Heinonen 2003; Han et al. 2009; Carrera and Seguin 2016; Xu et al. 2017; Zilic et al. 2019; Kim et al. 2022). In this study, to determine the extent to which the antioxidant capacity of soybean seeds is affected by the loss of function of the *GmBAS1* gene and the resulting soyasaponin deficiency, I evaluated the DPPH radical-scavenging capacity of soybean seed extracts. The antioxidant capacity of extracts from *bas1* mutant seeds was significantly lower than that of the control seeds, suggesting that soyasaponins contribute to the antioxidant capacity of soybean seeds (Figure 3.1).

To further investigate how the *bas1* mutation influences metabolic flux, transcriptome analysis was performed using RNA-seq. Functional enrichment analysis of the DEGs using the KEGG pathway database revealed considerable enrichment in flavonoid and steroid biosynthetic pathways (Table 3.2). Soyasaponins share a common skeleton comprising β-amyrin, which is synthesized via the cyclization of 2,3-oxidosqualene (Kushiro et al. 1998). Steroids have also been synthesized from 2,3-oxidosqualene (Schaller 2003, 2004; Benveniste 2004; Spanova and Daum 2011). Transcriptome analysis revealed that the *SMT2* gene was upregulated in *bas1* hypocotyls whereas the *CYP710A* gene was downregulated (Table 3.4). This suggests that the loss of *GmBAS1* function results in an increase in β-sitosterol content. The β-sitosterol content in *bas1* seeds was higher than that in control seeds (Figure 3.8). In addition, the expression levels of genes involved in the terpenoid backbone biosynthetic pathway were altered in the *bas1* hypocotyls (Table 3.5). Mapping these gene expression changes onto the KEGG pathway showed the downregulation of genes involved in terpenoid backbone biosynthesis and the activation of β-sitosterol biosynthesis (Figure 3.9). These changes in steroid biosynthesis may be caused by unusual concentrations of intermediates in terpenoid backbone biosynthesis and are likely unrelated to the physiological role of soyasaponins. However, the side effects of these metabolic changes, as well as the transcriptomic changes in genes related to motor proteins and photosynthesis, should be considered when characterizing the physiological properties of the *bas1* mutants.

This study demonstrated that polyphenolic compounds were significantly increased in *bas1* hypocotyls. While altered expression of the flavonoid biosynthetic genes was observed, these results could not account for the increase in polyphenol levels (Table 3.3). Furthermore, the isoflavone

content in *bas1* hypocotyls was approximately 20% lower than that of the control hypocotyls (Figure 3.4). Out of 25 DEGs in the flavonoid pathway, only a single gene encoding isoflavone-7-beta-glucosidase 6-malonyltransferase (GLYMA_19G030800) was estimated to be involved in isoflavone biosynthesis (Table 3.3). This suggests that the reduction in total isoflavone levels in *bas1* hypocotyls is unlikely to be driven by the transcriptional regulation of isoflavone biosynthetic genes. Based on these findings, I considered how the *bas1* mutation influences the biosynthesis of polyphenol compounds, including flavonoids, from the perspective of metabolic flux. In the *bas1* mutant, the disruption of soyasaponin biosynthesis likely leads to a surplus of acetyl-CoA. I propose that this excess precursor was redirected toward polyphenol biosynthesis, thereby resulting in the observed accumulation of polyphenolic compounds. Although the mechanisms involved in the utilization of acetyl-CoA toward polyphenol biosynthesis in higher plants are not fully clarified, previous studies have demonstrated that the amount of the acetyl-CoA supply can influence the flavonoid contents using engineered *Saccharomyces cerevisiae*, which expresses plant-derived biosynthetic genes (Zhang et al. 2021). Polyphenolic compounds are widely recognized as antioxidants (Kähkönen and Heinonen 2003; Han et al. 2009; Xu et al. 2017; Zilic et al. 2019; Kim et al. 2022). Despite the increased accumulation of polyphenols, the *bas1* mutant exhibited significantly lower DPPH antioxidant capacity than the control (Figure 3.1 and 3.3). These findings indicate that soyasaponins, rather than other polyphenolic compounds, play a primary role as specialized metabolites in the antioxidant system of soybean hypocotyls. The disrupted pattern of transcriptional changes in flavonoid biosynthesis could be attributed to the cellular redox dysfunction caused by the loss of soyasaponins.

In the tomato (*Solanum lycopersicum*) rhizosphere, a steroidal glycoalkaloid, α-tomatine, has been reported to influence bacterial composition and promote the enrichment of *Sphingobium* (Nakayasu et al. 2021). A previous study demonstrated that the abundance of *Sphingobium* was positively correlated with α-tomatine content in soil, as shown using a pseudo-rhizosphere system (Takamatsu et al. 2023). Furthermore, *Sphingobium* sp. RC1, which specifically degrades steroid-type saponins, was isolated from tomato roots (Nakayasu et al. 2023). These findings suggest that α-tomatine plays an important role in establishing a close association between tomato roots and *Sphingobium*. In this study, the abundance of *Novosphingobium* in the *bas1* rhizosphere was significantly lower than that in the control rhizosphere, despite no detectable differences in the overall bacterial communities between the *bas1* and control rhizospheres (Figure 3.5, 3.6 and 3.7). Together with the enrichment of *Novosphingobium* in soyasaponin-treated soil (Fujimatsu et al. 2020), these

results support the conclusion that soyasaponins play a specific role in mediating the association between soybean roots and *Novosphingobium*. In the soyasaponin-treated soil, three families and three genera were enriched compared to bulk soil, while only one family was depleted (Fujimatsu et al. 2020). Given that the number of bacterial taxa affected by the treatment of soils with soyasaponins is limited, the relative abundance of bacterial taxa other than *Novosphingobium* could be attributed to the differences in the profile of root exudates between the *bas1* mutant and the control. Comprehensive transcriptome and metabolome profiling of the mutant and control plants, as well as their rhizospheres, may help clarify whether the difference in bacterial composition other than *Novosphingobium* is caused by metabolic changes or niche competition.

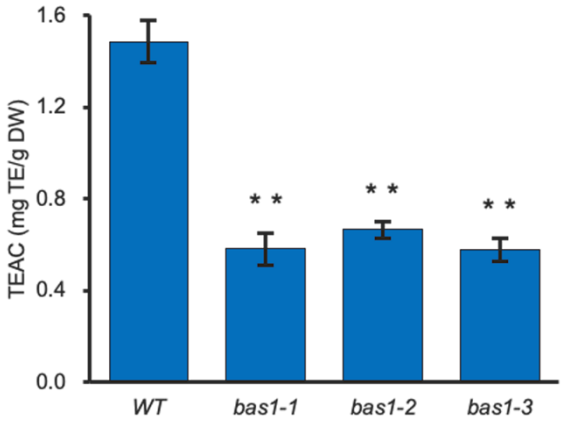


Figure 3.1 DPPH radical-scavenging capacity of extracts from mutant and control seeds. '*WT*' indicates the value for control seeds. '*bas1-1*', '*bas1-2*', and '*bas1-3*' indicate values for each of the three mutant lines. Antioxidant capacity was expressed as Trolox Equivalent (TE) per 1 g dry weight (DW). Values are presented as mean ± SD (n = 3). Asterisks indicate significant differences between the mutant lines and control seeds based on Dunnett test ($p < 0.01$).

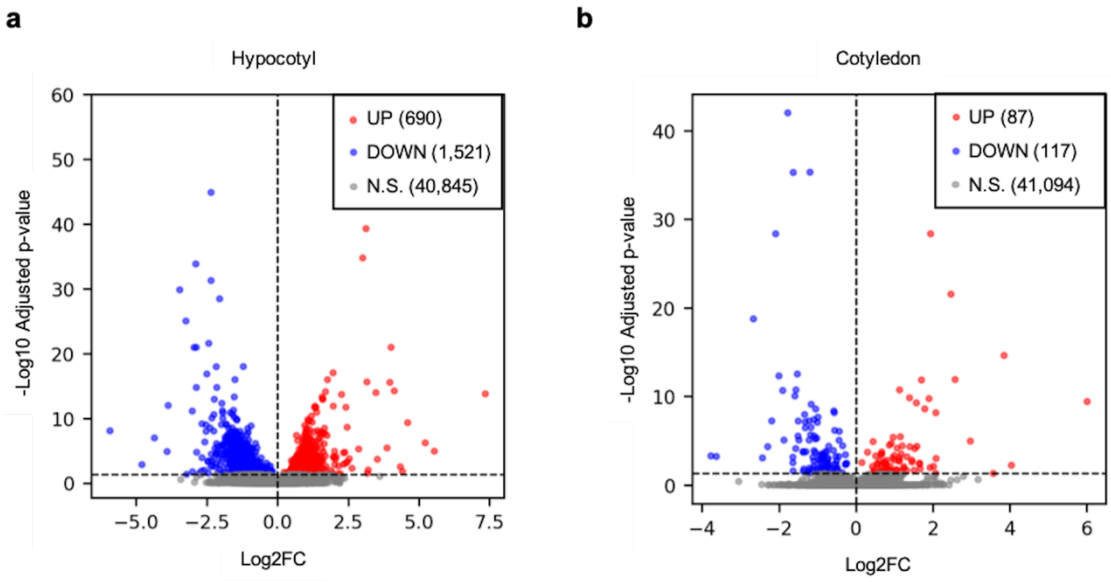


Figure 3.2 Volcano plot of DEGs in hypocotyls (a) and cotyledons (b) of the *bas1* mutant seeds. Red and blue dots indicate significantly upregulated and downregulated genes, respectively (adjusted *p*-value < 0.05). Grey dots indicate non-significant (N.S.) genes. Numbers in parentheses indicate the number of genes in each category.

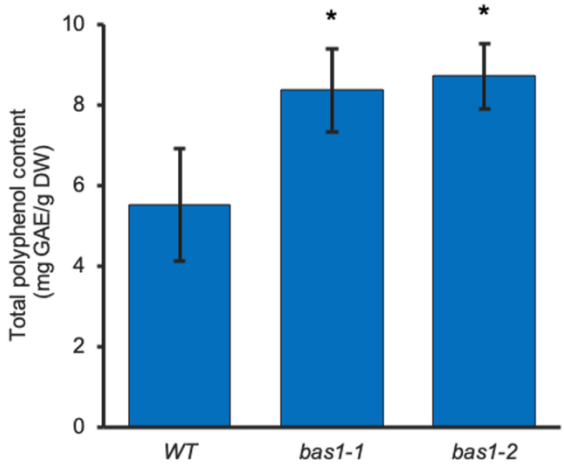


Figure 3.3 Total polyphenol content in mutant and control hypocotyls.

'*WT*' indicates the value of the control seeds. '*bas1-1*' and '*bas1-2*' indicate the values of each mutant line. The phenolic content was expressed as gallic acid equivalents (GAE) per 1 g dry weight (DW). Values are presented as mean ± SD (n = 3). Asterisks indicate significant differences between the mutant lines and control seeds based on Dunnett test ($p < 0.05$).

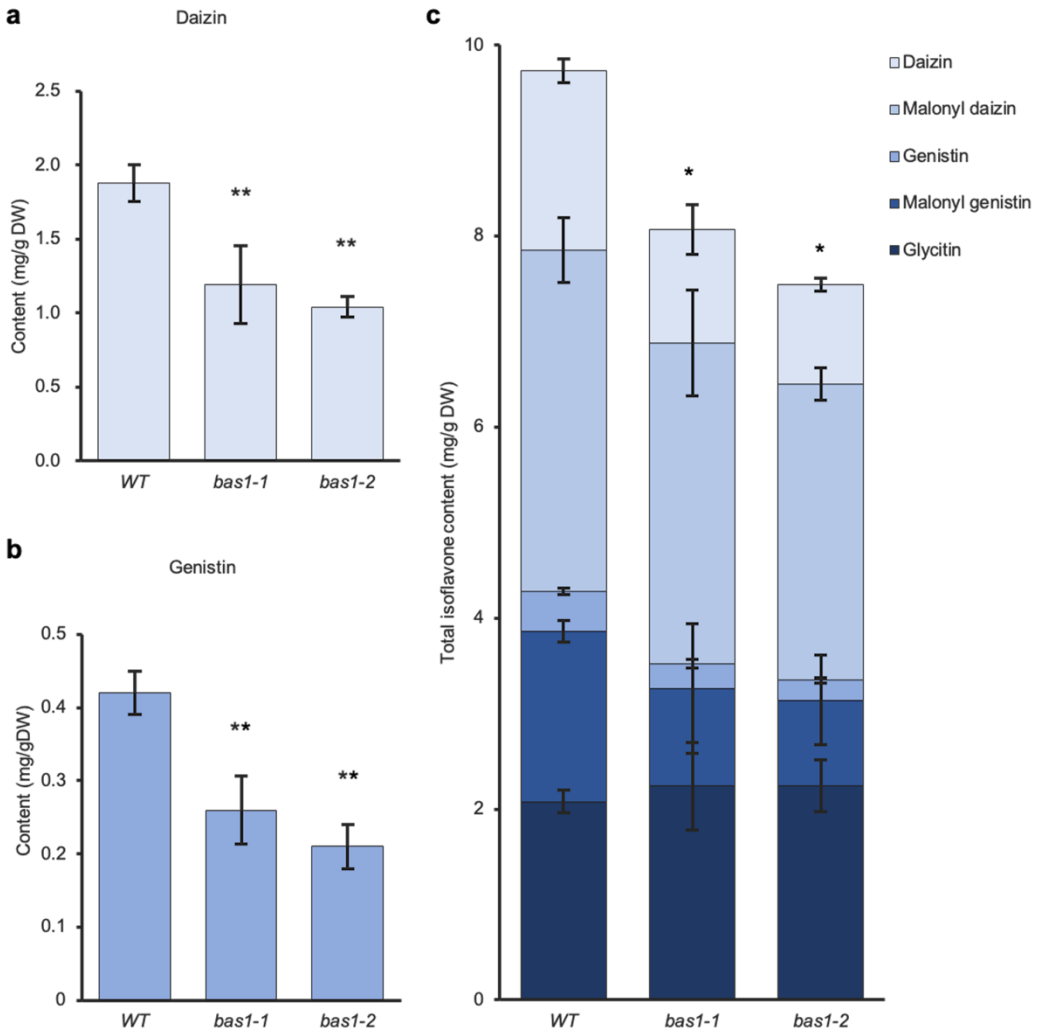


Figure 3.4 The contents of daizin (**a**), genistin (**b**), and total isoflavones (**c**) in the hypocotyls of the mutant and control plants.

'*WT*' indicates the value of the control hypocotyls. '*bas1-1*' and '*bas1-2*' indicate the values of each mutant line. Values are presented as mean ± SD (n = 3). Asterisks indicate significant differences between the mutant lines and control hypocotyls based on Dunnett test (*$p < 0.05$, ** $p < 0.01$).

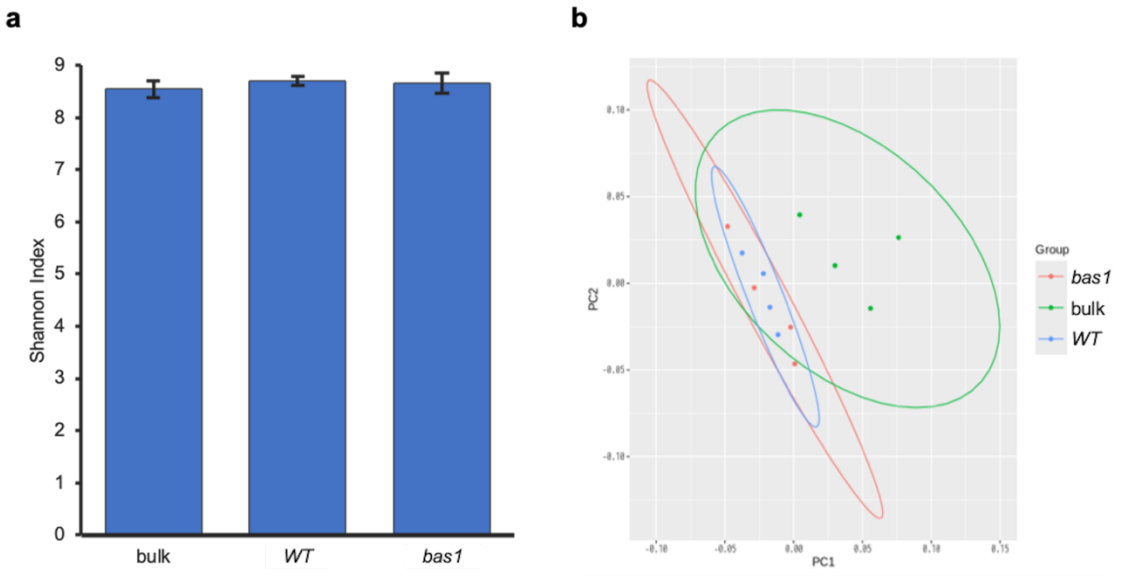


Figure 3.5 Analysis of bacterial communities in bulk and rhizosphere soils using amplicon sequencing of the 16S rRNA gene.

a Shannon diversity indices of bacterial communities in bulk soil and rhizosphere soil of mutant and control plants. 'bulk' indicates the value of the bulk soil. '*WT*" indicates the value of the control rhizosphere. '*bas1*' indicates the value of the *bas1* mutant rhizosphere. Values are presented as mean ± SD (n = 4).

b Weighted UniFrac-based principal coordinate analysis of the bacterial communities in the bulk and rhizosphere of mutant and control plants. Each ellipse corresponds to the 95% confidence interval of each group.

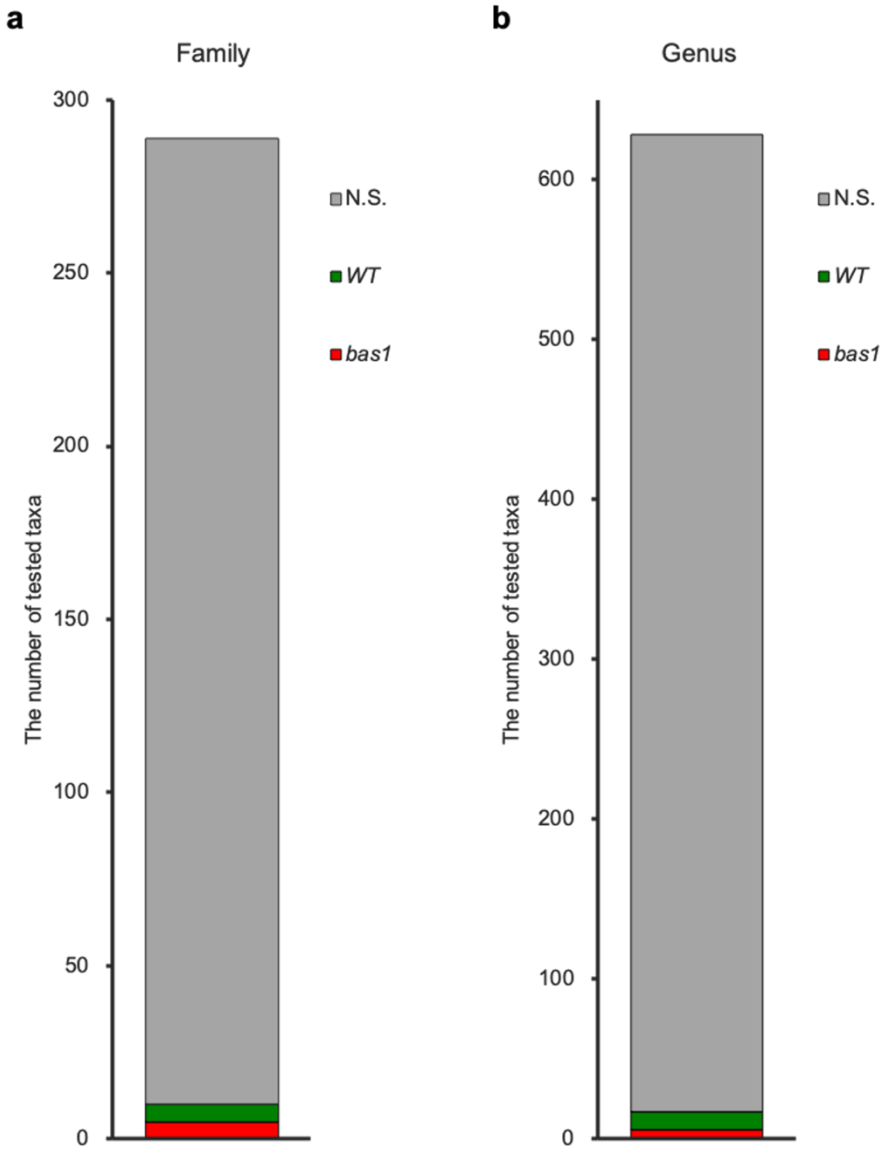


Figure 3.6 Numbers of tested families (a) and genera (b) in the *bas1* mutant and control rhizosphere. Red sections represent taxa enriched in the *bas1* mutant rhizosphere, whereas green sections represent those enriched in control rhizosphere, identified by LEfSe analysis (Kruskal–Wallis test, $q < 0.05$; LDA score > 2.0). Grey sections represent taxa with non-significant (N.S.) differences.

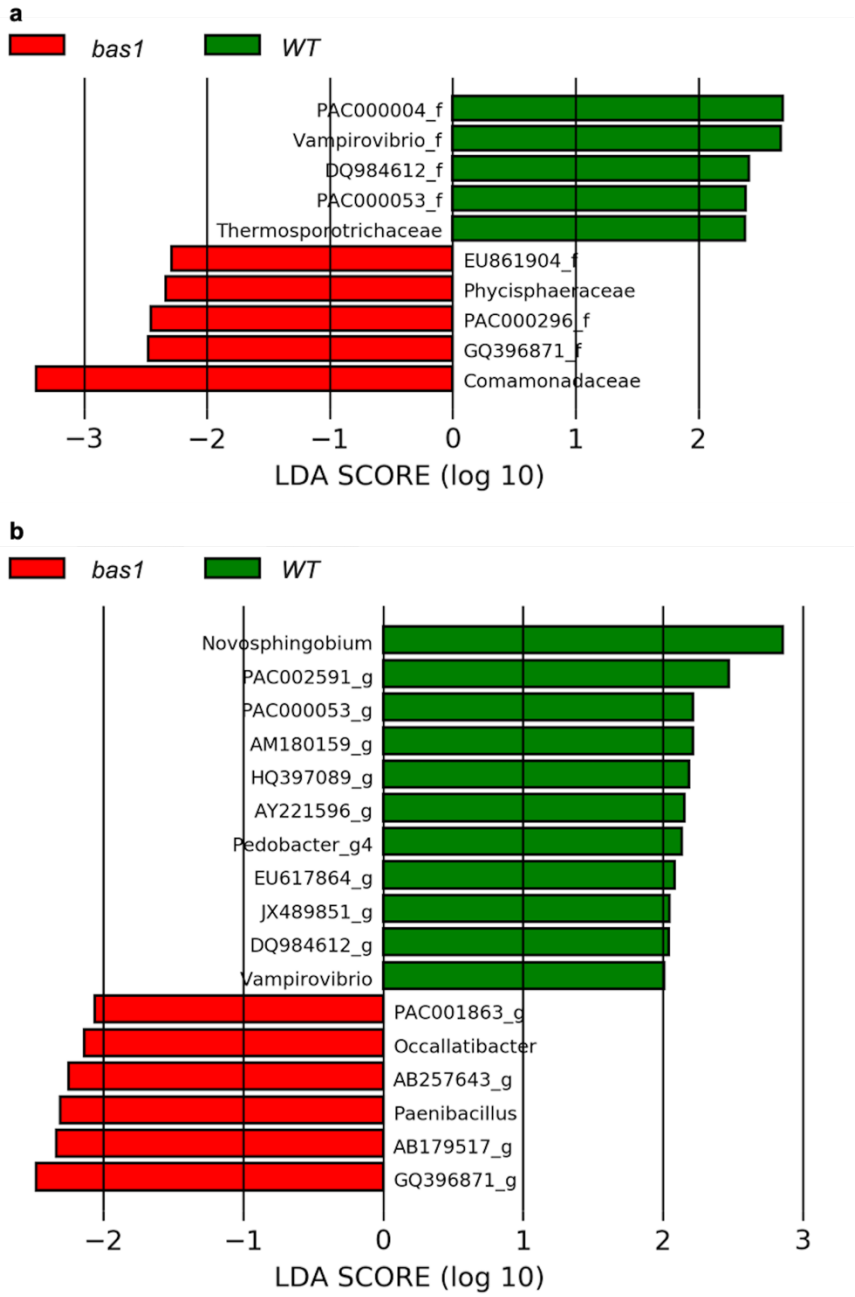


Figure 3.7 Differentially abundant bacterial taxa between mutant and control rhizospheres. Bar plots of bacterial families (**a**) and genera (**b**) identified as significant by LEfSe analysis.

The length of the bars represents the logarithmic LDA score for each taxon. Taxa enriched in the *bas1* mutant rhizosphere are shown in red, whereas those enriched in the control rhizosphere are shown in green, based on a Kruskal–Wallis test ($q < 0.05$; LDA score > 2.0).

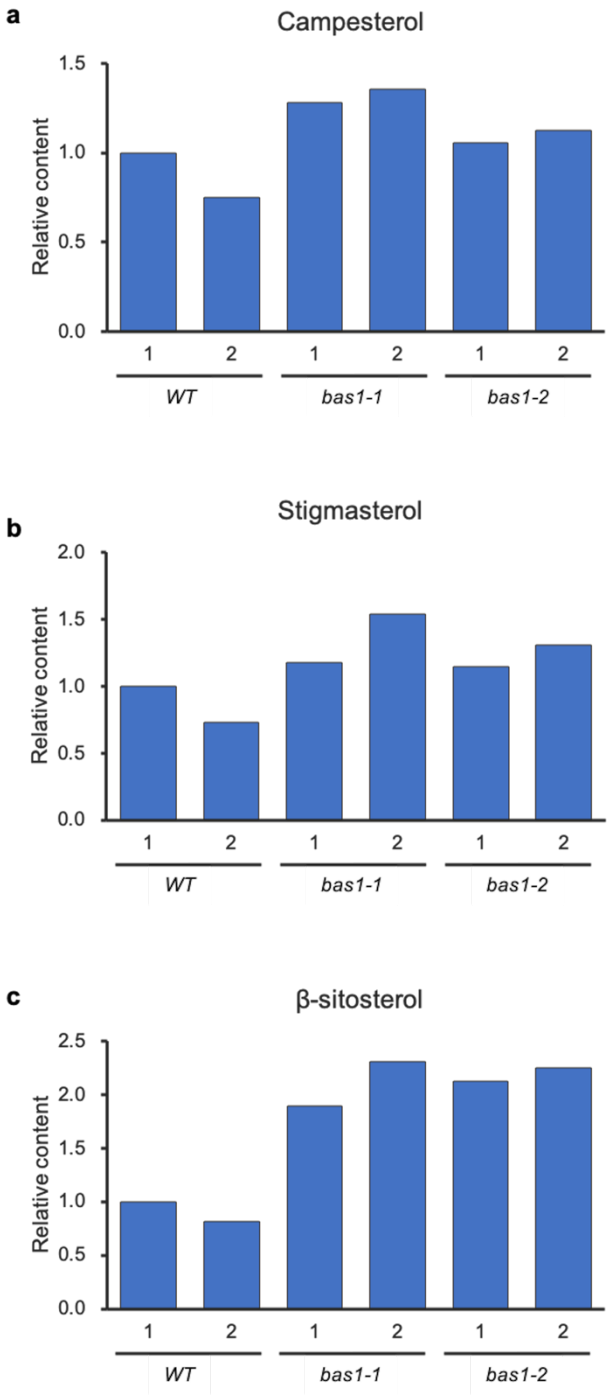


Figure 3.8 Relative contents of campesterol (**a**), stigmasterol (**b**), and β-sitosterol (**c**) in mutant and control seeds.

'*WT*' indicates the value of control seeds. '*bas1-1*' and '*bas1-2*' indicate values of each mutant lines. Each bar represents the relative content of a single biological replicate. All values were normalized to the reference sample (WT1).

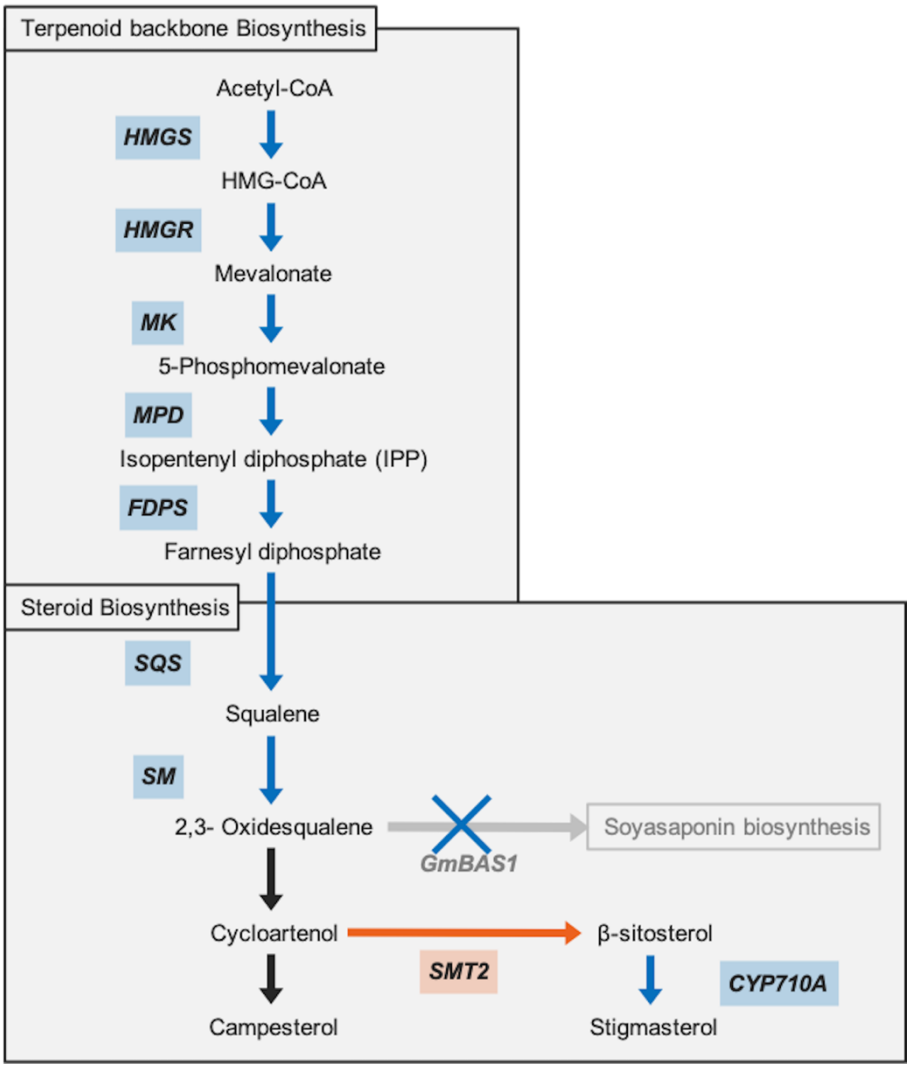


Figure 3.9 Schematic representation of gene expression changes in the terpenoid backbone and steroid/soyasaponin biosynthetic pathways in *bas1* mutant hypocotyls.

The gene expression changes listed in Table 3 were mapped onto a KEGG-based pathway diagram. Genes significantly upregulated in the *bas1* mutant compared to the control are highlighted in red, whereas downregulated genes are highlighted in blue. The red and blue arrows indicate potentially increased and decreased metabolic flows, respectively. A blue cross mark, gray arrow, and gray letters, '*GmBAS1*' and 'Soyasaponin biosynthesis' denote the deficiency of the pathway. Abbreviations: HMGS, hydroxymethylglutaryl-CoA synthase; HMGR, hydroxymethylglutaryl-CoA reductase; MK, mevalonate kinase; MPD, mevalonate pyrophosphate decarboxylase; FDPD, farnesyl diphosphate synthase; SQS, squalene synthase; SM, squalene monooxygenase; SMT2, sterol 24-C-methyltransferase.

Table 3.1 Primer sequences used in DNA library preparation for amplicon sequencing.

Primer name	Primer sequence (5'-3')	Use of amplicon
1st-515f_MIX	ACACTCTTTCCCTACACGACGCTCTTCCGATCT-NNNNN-GTGCCAGCMGCCGCGGTAA	1st PCR of a DNA library preparation
1st-806r_MIX	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-NNNNN-GGACTACHVGGGTWTCTAAT	1st PCR of a DNA library preparation
2ndF	AATGATACGGCGACCACCGAGATCTACAC-Index2-ACACTCTTTCCCTACACGACGC	2nd PCR of a DNA library preparation
2ndR	CAAGCAGAAGACGGCATACGAGAT-Index1-GTGACTGGAGTTCAGACGTGTG	2nd PCR of a DNA library preparation

Table 3.2 Results of functional enrichment analysis using the KEGG database.

Term name	Term ID	Adjusted *p*-value
Motor proteins	KEGG:04814	2.55E-20
Photosynthesis - antenna proteins	KEGG:00196	6.44E-11
Photosynthesis	KEGG:00195	4.35E-07
Metabolic pathways	KEGG:01100	3.81E-05
Stilbenoid, diarylheptanoid and gingerol biosynthesis	KEGG:00945	0.00252
Flavonoid biosynthesis	KEGG:00941	0.00294
Sesquiterpenoid and triterpenoid biosynthesis	KEGG:00909	0.01054
Steroid biosynthesis	KEGG:00100	0.03009

Table 3.3 List of DEGs related to the flavonoid pathway in hypocotyls.

Gene ID	Expression mean	Expression level		Log2FC	Adjusted *p*-value	Description
		WT	bas1			
GLYMA_02G236500	26.215	33.145	19.284	-0.751	0.018080	Trans-cinnamate 4-monooxygenase (CYP73A)
GLYMA_08G220200	0.868	1.374	0.362	-0.801	0.005324	Shikimate O-hydroxycinnamoyltransferase (HCT)
GLYMA_08G311600	5.062	7.006	3.118	-0.958	0.026081	Shikimate O-hydroxycinnamoyltransferase (HCT)
GLYMA_08G311900	10.633	15.273	5.993	-1.218	4.72E-05	Shikimate O-hydroxycinnamoyltransferase (HCT)
GLYMA_15G241300	23.985	31.649	16.321	-0.914	0.013260	Shikimate O-hydroxycinnamoyltransferase (HCT)
GLYMA_18G103400	21.504	30.931	12.076	-1.287	5.14E-06	Shikimate O-hydroxycinnamoyltransferase (HCT)
GLYMA_18G103500	36.001	46.047	25.955	-0.803	0.011089	Shikimate O-hydroxycinnamoyltransferase (HCT)
GLYMA_01G228700	213.832	270.086	157.579	-0.773	0.022343	Naringenin chalcone synthase (CHS)
GLYMA_08G109500	17.223	23.908	10.537	-1.110	0.000999	Naringenin chalcone synthase (CHS)
GLYMA_08G110500	7.441	10.211	4.670	-0.983	0.010633	Naringenin chalcone synthase (CHS)
GLYMA_11G011500	51.236	66.598	35.874	-0.874	0.004404	Naringenin chalcone synthase (CHS)
GLYMA_19G105100	5.281	7.682	2.880	-1.161	0.009216	Naringenin chalcone synthase (CHS)
GLYMA_20G241600	5.197	10.021	0.372	-3.005	0.018274	Chalcone-flavonone isomerase 1 (CHI)
GLYMA_02G119600	26.602	35.341	17.864	-0.945	0.001348	Flavonoid 3’-monooxygenase (F3’5’H)
GLYMA_10G204800	0.800	1.364	0.236	-0.93559	0.010607	Leucoanthocyanidin reductase (LAR)
GLYMA_05G181100	3.599	5.086	2.113	-0.966	0.001956	Anthocyanidin 3-O-glucosyltransferase (3GT)
GLYMA_12G132500	4.421	8.550	0.292	-2.885	1.48E-34	Anthocyanidin 3-O-glucosyltransferase (3GT)
GLYMA_13G254700	4.044	5.907	2.180	-1.118	0.014821	Anthocyanidin 3-O-glucosyltransferase (3GT)

GLYMA_13G302300	10.959	14.291	7.628	-0.825	0.014909	Anthocyanidin 3-O-glucosyltransferase (3GT)
GLYMA_12G132500	4.421	8.550	0.292	-2.885	1.48E-34	Anthocyanidin 3-O-glucoside 2"-O-glucosyltransferase (3GGT)
GLYMA_13G254700	4.044	5.907	2.180	-1.118	0.014821	Anthocyanidin 3-O-glucoside 2"-O-glucosyltransferase (3GGT)
GLYMA_04G227700	41.218	56.107	26.330	-1.063	0.000300	Flavonol 3-O-methyltransferase (3OMT)
GLYMA_19G030800	6.502	9.977	3.028	-1.446	0.000133	Isoflavone-7-beta-glucoside 6”-malonyltransferase (IF7MAT)
GLYMA_17G173200	11.677	0	23.354	4.606	4.79E-10	Dihydrokaempferol 4-reductase (DFR)
GLYMA_19G254600	1.069	0.397	1.740	0.971	0.019256	Anthocyanidin reductase (ANR)

Table 3.4 List of DEGs related to the steroid pathway in hypocotyls.

Gene ID	Expression mean	Expression level		Log2FC	Adjusted *p*-value	Description
		WT	bas1			
GLYMA_11G112000	15.47	21.70	9.24	-1.149	0.000116	Squalene synthase (SQS)
GLYMA_12G038200	9.58	14.06	5.10	-1.303	0.001092	Squalene synthase (SQS)
GLYMA_07G185500	17.36	25.90	8.81	-1.454	1.59E-05	Squalene monooxygenase (SM)
GLYMA_08G063700	35.90	45.61	26.19	-0.777	0.021325	Squalene monooxygenase (SM)
GLYMA_14G090400	4.99	8.47	1.52	-1.910	8.20E-07	Squalene monooxygenase (SM)
GLYMA_09G186900	10.41	16.22	4.60	-1.619	2.15E-06	Cholestenol delta-isomerase (EBP)
GLYMA_13G217400	5.52	7.73	3.32	-1.014	0.004429	Cytochrome P450 family 710 subfamily A (CYP710A)
GLYMA_11G103500	21.41	28.14	14.68	-0.893	0.000270	Delta (24)-sterol reductase (SR)
GLYMA_12G028300	12.63	19.66	5.60	-1.645	1.14E-08	Delta (24)-sterol reductase (SR)
GLYMA_15G061800	5.09	2.93	7.24	1.067	0.000424	Squalene monooxygenase (SM)
GLYMA_06G020700	35.99	20.00	51.98	1.334	1.33E-12	Sterol 24-C-methyltransferase (SMT2)

Table 3.5 DEGs list related to terpenoid backbone pathway in hypocotyls.

Gene ID	Expression Mean	Expression level		Log2FC	Adjusted *p*-value	Description
		WT	bas1			
GLYMA_01G215500	94.0	118.2	69.82	-0.751	0.026044	Hydroxymethylglutaryl-CoA synthase (HMGS)
GLYMA_11G027000	19.2	25.7	12.80	-0.954	0.000547	Hydroxymethylglutaryl-CoA synthase (HMGS)
GLYMA_01G156700	10.1	13.4	6.84	-0.881	0.011893	Hydroxymethylglutaryl-CoA reductase (HMGR)
GLYMA_03G239000	16.3	23.2	9.52	-1.201	0.000381	Mevalonate kinase (MK)
GLYMA_20G109900	27.7	38.0	17.49	-1.080	0.000525	Mevalonate pyrophosphate decarboxylase (MPD)
GLYMA_09G015600	19.0	26.1	11.94	-1.070	0.003210	Farnesyl diphosphate synthase (FDPS)
GLYMA_15G121400	22.2	29.0	15.36	-0.878	0.031178	Farnesyl diphosphate synthase (FDPS)

CHAPTER 4
General Discussion

An objective of this study was to improve the flavor of soybean seeds by eliminating soyasaponins, which are the predominant agents underlying bitterness of soyfoods. By employing CRISPR/Cas9-mediated targeted mutagenesis of the *GmBAS1* gene, soyasaponins were effectively eliminated from soybean seeds. Disruption of the *GmBAS1* gene resulted in the inhibition of soyasaponin biosynthesis throughout the soybean plant, rather than being confined to seed tissues.

Characterization of the *bas1* mutant demonstrated that soyasaponins serve critical functions in soybean plants as specialized metabolites indispensable for the antioxidant system in hypocotyls. This provides the first evidence that the antioxidant capacity of soyasaponins contributes to that of soybean seeds. In general, flavonoids are often regarded as antioxidants (Desmet et al. 2021; Marone et al. 2022). Since flavonoids are aromatic specialized metabolites synthesized from phenylalanine, their accumulation level may be restricted by the pool size of aromatic acids (Vogt 2010). In soybean hypocotyls, isoflavones accumulate approximately 1 %(w/w), suggesting that they are the major end products in the biosynthetic pathways of aromatic specialized metabolites (Tsukamoto et al. 1995). However, the antioxidant capacity of isoflavone glycosides is not particularly high (Kim et al. 2022). Therefore, flavonoids are not thought to primarily contribute to the antioxidant capacity in soybean hypocotyls. As mentioned in chapter 3, DDMP soyasaponins are known to possess strong antioxidant activity, and their contents in soybean hypocotyls are comparable to those of isoflavones (Shiraiwa et al. 1991a; Yoshiki and Okubo 1995; Yoshiki et al. 2001). Together with the fact that the pyran ring of DDMP moieties is structurally similar to a phenol ring, these findings suggest that the role of DDMP soyasaponins likely overlap with that of flavonoids as antioxidants (Kudou et al. 1992, 1993). The findings obtained in this study strongly support this idea.

Furthermore, I verified that soyasaponins mediate specific interactions between soybean roots and *Novosphingobium* by the analysis of the *bas1* rhizosphere microbiome. To my knowledge, this is the first example where a specific relationship between soybean specialized metabolites and bacteria has been validated by both soil treatment with purified standards and analysis of the mutant rhizosphere. Treatment with *Novosphingobium* elicits induced systemic resistance in pepper (*Capsicium annuum*) (Hahm et al. 2012). Similarly, inoculation with *Novosphingobium* reduces salt stress in *Citrus macrophylla* (Vives-Peris et al. 2018). These findings suggest that these bacteria may be involved in stress control for the plants with which they form symbiotic relationships. However,

secreting soyasaponins solely to recruit specific bacteria before stress exposure appears to cost too much carbon. The function of triterpenoid saponins as defense compounds are likely to be conserved in various plant species (Augustin et al. 2011). Given that triterpenoid saponins are primarily accumulated in underground organs in many plants (Moses et al. 2014), it is plausible to assume that secretion of soyasaponins from root as defense compounds preceded the establishment of the symbiotic relationship. The enhanced stress tolerance conferred by this interaction would have allowed these symbiotic populations to be evolutionary selected. Therefore, it is conceivable that soybeans plants have shifted the function of soyasaponins from defense compounds to symbiotic signaling molecules in roots.

The insights gained from this study indicate that the complete removal of soyasaponins from soybean plants could compromise physiological processes essential for normal growth. Therefore, efficient breeding strategies should aim not at the complete elimination of soyasaponins, but at the fine-tuned modulation of their biosynthesis, ensuring the reduction of bitterness in seeds while preserving their physiological functions in other tissues. The most promising approach to overcoming this challenge may be most effectively achieved by specifically reducing soyasaponins in seeds. Several transcription factors involved in the regulation of soyasaponin biosynthetic genes have been identified in *G. uralensis* and *M. truncatula* (Mertens et al. 2016; Tamura et al. 2018; Ribeiro et al. 2020). Among these, TRITERPENE SAPONIN ACTIVATION REGULATOR3 (TSAR3) is a seed-specific basic helix-loop-helix (bHLH) transcription factor that governs saponin biosynthesis during seed development in *M. truncatula* (Ribeiro et al. 2020). In soybean, at least 15 homologs encode bHLH transcription factors classified in the same subclade as MtTSARs, including Glyma13g32650, which is expressed in soybean seed, nodule, and root tissues (Suzuki et al. 2021). Identification and targeted mutagenesis of TSAR homologs that are strongly and specifically expressed in seed tissue may enable the selective suppression of soyasaponin accumulation in seeds.

Elucidating the spatial distribution of soyasaponins provides a critical basis of understanding their physiological functions. Soyasaponins are often evaluated both quantitatively and qualitatively through liquid chromatography analysis. However, this approach requires sample homogenization, resulting in the loss of spatial information regarding tissue architecture. MALDI-TOF MSI enables the visualization of the spatial distribution of biomolecules in a tissue specimen (Caprioli et al. 1997; Spengler and Hubert 2002). This study employed MALDI-TOF MS to visualize the distribution of soyasaponins in soybean seed tissues. DDMP soyasaponin βg and αg, representing major types of

soyasaponins in soybean hypocotyls, were successfully detected (Figure 4.1). Both types were mainly detected in the cortex and pith of soybean hypocotyls (Figure 4.1.b). In the roots of *A. strigose*, the *BAS* gene which is involved in the biosynthesis of triterpenoid saponin avenacin, is expressed in the epidermal cells of root tips (Wegel et al. 2009). In the roots of *Panax ginseng* and *P. quinquefolius*, high concentrations of triterpenoid saponins are present in the cork tissue, suggesting their role as defense compounds against animals and insect attacks (Sun et al. 2021). These findings suggest that the role of DDMP soyasaponins in soybeans might be different from other triterpenoid saponins. DDMP soyasaponins are known to possess strong antioxidant activity (Yoshiki and Okubo 1995; Yoshiki et al. 2001). These properties of DDMP soyasaponins suggest a possible contribution to the mitigation of oxidative stress occurring during intensive cell division in the initial phase of germination. The complete deficiency of soyasaponins from hypocotyls shows little impact under controlled experimental environment, but may affect growth when germination is subjected to external stress. Further investigation is required to elucidate the effect of soyasaponin deficiency on germination under stress conditions.

From these findings, it is clear that breeding strategies targeting on soyasaponin composition in soybean must be carefully designed to achieve flavor improvement without impairing the physiological processes. Genetic improvement targeting on transcriptional regulation of soyasaponin biosynthesis may facilitate the modification of soyasaponin contents in each tissue. Furthermore, enhancement of soyasaponin production in roots could contribute to the beneficial modulation of rhizosphere microbial communities. Therefore, a comprehensive understanding of the physiological roles of soyasaponins will crucial for future studies. With growing concerns about global food security driven by population growth and climate change, ensuring stable crop yield while reducing environmental impact has become a top priority in agriculture. The developing crops with tolerance to environmental stress and efficient resource utilization, including the use of symbiotic organisms, is an important goal in agriculture. The findings obtained in this study are expected to contribute progress in addressing these challenges.

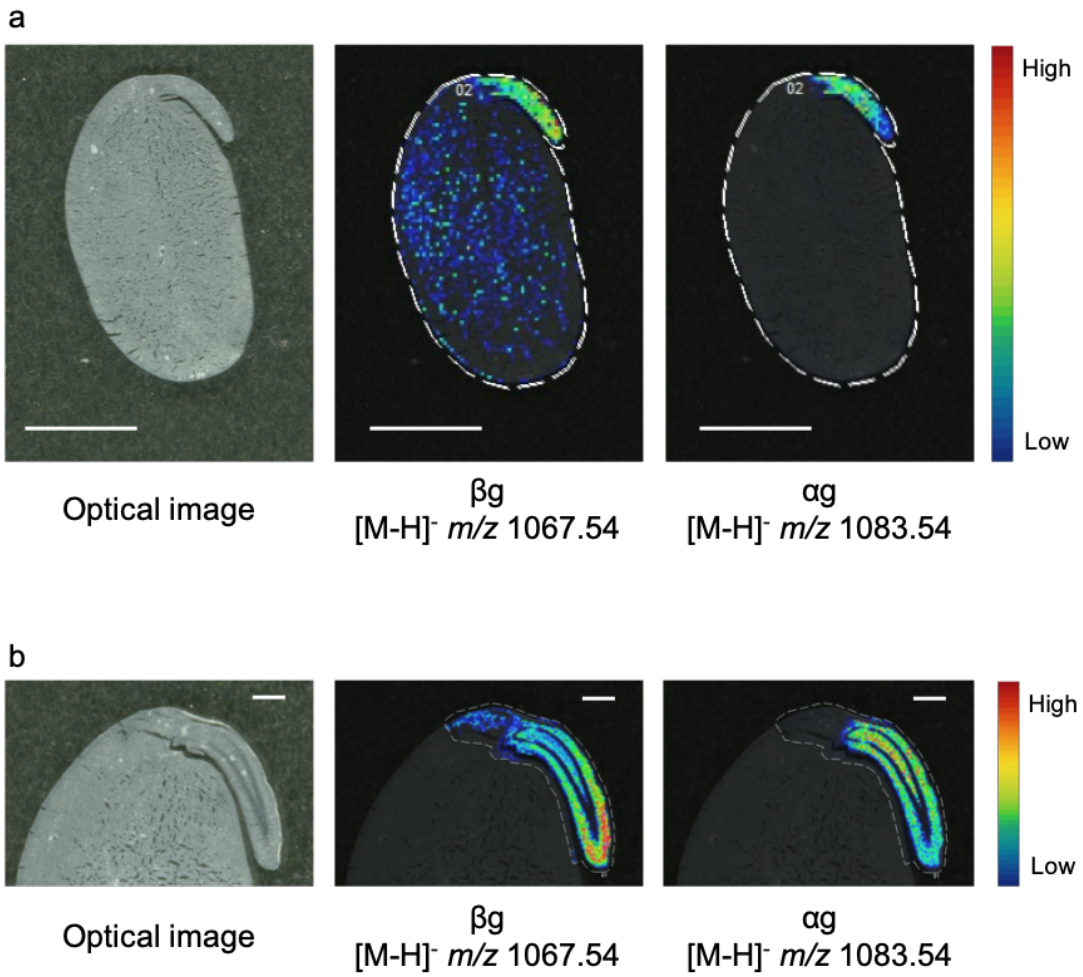


Figure 4.1 Spatial distributions of representative soyasaponins in soybean seeds.

a Whole seeds. Scale bars = 5 mm

b Hypocotyls. Scale bars = 100 µm

Longitudinal thin (40 µm) specimens were prepared from water-absorbed seeds. The specimens were sprayed with 1,5-Diaminonaphthalene (DAN) solution as a matrix. MALDI-TOF MSI analysis was performed in negative ion mode using ultrafleXtreme (Bruker). The laser spot size was set to approximately 50 µm, with the interval of the probing spots of 150 µm for whole seeds and 50 µm for hypocotyls. Data processing and ion map construction were performed by fleximaging 5.0 (Bruker). Color scales indicate relative intensity of each peak. (See appendix for detailed materials and methods).

Summary

Soyasaponins, which are triterpenoid saponins contained in soybean [*G. max* (L.) Merril], are responsible for the astringent aftertaste of soyfood, and their complete elimination from soybean seeds is a key challenge in the development of cultivars with improved taste. While the loss of function in the β-amyrin synthase genes (*GmBAS1* and *GmBAS2*) has proven effective in reducing soyasaponin content in soybean seeds, the specific functional roles of these genes remain unclear. In this study, targeted mutagenesis was performed on two *GmBAS* loci using a DNA-free CRISPR-Cas9 system. A complex of sgRNA targeting sequences conserved in the two loci and Cas9 protein was introduced into the shoot apical meristems of soybean embryonic axes via bombardment. CAPS analysis conducted one month post-bombardment revealed that 138 seedlings out of 1,467 screened exhibited mutations at one or both *GmBAS* loci. CAPS and sequencing analysis in the subsequent generation identified a total of 16 plants with inheritable mutations ranging from one to ten nucleotides. HPLC analysis showed that targeted mutagenesis in the *GmBAS1* locus resulted in the absence of soyasaponins in mature seeds, as well as in young roots, stems, and leaves. These findings demonstrate that *GmBAS1* is the predominant ß-amyrin synthase gene in soybean plants. Additionally, the DNA-free CRISPR/Cas9 system was shown to be highly efficient in inducing simultaneous mutagenesis at two target loci using a single gRNA.

Soybeans accumulate soyasaponins in various organs; however, their precise biological roles remain unclear. To elucidate the biological characteristics of soyasaponins in soybean plants, I examined their effects on seed metabolic flux and rhizosphere bacterial communities using the *GmBAS1* mutants mentioned above. The 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical-scavenging capacity of *bas1* mutant seeds was significantly lower than that of control seeds, suggesting that soyasaponins confer antioxidant capacity to soybean seeds. Transcriptome analysis using RNA sequencing indicated that flavanol synthesis was upregulated in the *bas1* mutant hypocotyls. Total polyphenol levels were consistently elevated, whereas the isoflavone content declined in *bas1* hypocotyls. Furthermore, examination of rhizosphere bacterial communities through 16S rRNA gene amplicon sequencing showed that the abundance of *Novosphingobium* significantly decreased in the *bas1* rhizosphere. Taken together, these findings highlight the dual and distinct roles of soyasaponins: they function as *in vivo* antioxidants and mediators of plant-microbe interactions that modulate the rhizosphere microbial community.

References

Abe I, Prestwich GD (1995) Identification of the active site of vertebrate oxidosqualene cyclase. Lipids 30:231–234. https://doi.org/10.1007/BF02537826

Adachi K, Hirose A, Kanazashi Y, Hibara M, Hirata T, Mikami M, Endo M, Hirose S, Maruyama N, Ishimoto M, Ave J, Yamada T (2021) Site-directed mutagenesis by biolistic transformation efficiently generates inheritable mutations in a targeted locus in soybean somatic embryos and transgene-free descendants in the T_1 generation. Transgenic Res 30:77–89. https://doi.org/10.1007/s11248-020-00229-4

Almeida-Silva F, Pedrosa-Silva F, Venancio TM (2023) The Soybean Expression Atlas v2: A comprehensive database of over 5000 RNA-seq samples. Plant J 116:1041–1051. https://doi.org/10.1111/tpj.16459

Asa H, Kuwabara C, Matsumoto K, Shigeta R, Yamamoto T, Masuda Y, Yamada T (2025) Simultaneous site-directed mutagenesis for soybean β-amyrin synthase genes via DNA-free CRISPR/Cas9 system using a single gRNA. Plant Cell Rep 44:40. https://doi.org/10.1007/s00299-025-03433-w

Augustin JM, Kuzina V, Andersen SB, Bak S (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry 72:435–457. https://doi.org/10.1016/j.phytochem.2011.01.015

Avato P, Bucci R, Tava A, Vitali C, Rosato A, Bialy Z, Jurzysta M (2006) Antimicrobial activity of saponins from *Medicago* sp.: structure-activity relationship. Phytotherapy Res 20:454–457. https://doi.org/10.1002/ptr.1876

Benveniste P (2004) Biosynthesis and accumulation of sterols. Annu Rev Plant Biol 55:429–457. https://doi.org/10.1146/annurev.arplant.55.031903.141616

Caprioli RM, Farmer TB, Gile J (1997) Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal Chem 69:4751–4760. https://doi.org/10.1021/ac970888i

Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, Graham N, Odoardi M, Piano E, Arcioni S, May S, Scotti C, Calderini O (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell 23:3070–3081. https://doi.org/10.1105/tpc.111.087312

Carrera CS, Seguin P (2016) Factors affecting tocopherol concentrations in soybean seeds. J Agric Food Chem 64:9465–9474. https://doi.org/10.1021/acs.jafc.6b03902

Chung SY, Seki H, Fujisawa Y, Shimoda Y, Hiraga S, Nomura Y, Saito K, Ishimoto M, Muranaka T (2020) A cellulose synthase-derived enzyme catalyses 3-O-glucuronosylation in saponin biosynthesis. Nat Commun 11:5664. https://doi.org/10.1038/s41467-020-19399-0

Delis C, Krokida A, Georgiou S, Peña-Rodríguez LM, Kavroulakis N, Ioannou E, Roussis V, Osbourn AE, Papadopoulou KK (2011) Role of lupeol synthase in *Lotus japonicus* nodule formation. New Phytol 189:335–346. https://doi.org/10.1111/j.1469-8137.2010.03463.x

Desmet S, Morreel K, Dauwe R, et al (2021) Origin and function of structural diversity in the plant specialized metabolome. Plants 10:11. https://doi.org/10.3390/plants10112393

Didierlaurent AM, Laupèze B, Di Pasquale A, Hergli N, Collignon C, Garçon N (2017) Adjuvant system AS01: helping to overcome the challenges of modern vaccines. Expert Rev Vaccines 16:55–63. https://doi.org/10.1080/14760584.2016.1213632

FAO/WHO (1991) Protein quality evaluation, 1st edn. FAO, Rome

Field B, Fiston-Lavier A-S, Kemen A, Geisler K, Quesneville H, Osbourn AE (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. Proc Natl Acad Sci U S A 108:16116–16121. https://doi.org/10.1073/pnas.1109273108

Field B, Osbourn AE (2008) Metabolic diversification—independent assembly of operon-like gene clusters in different plants. Science 320:543–547. https://doi.org/10.1126/science.1154990

Fu M, Chen L, Cai Y, Su Q, Chen Y, Hou W (2022) CRISPR/Cas9-mediated mutagenesis of *GmFAD2-1A* and/or *GmFAD2-1B* to create high-oleic-acid soybean. Agronomy (Basel) 12:3218. https://doi.org/10.3390/agronomy12123218

Fujimatsu T, Endo K, Yazaki K, Sugiyama A (2020) Secretion dynamics of soyasaponins in soybean roots and effects to modify the bacterial composition. Plant Direct 4:e00259. https://doi.org/10.1002/pld3.259

Go YS, Lee SB, Kim HJ, Kim J, Park H-Y, Kim J-K, Shibata K, Yokota T, Ohyama, K, Muranaka T, Arseniyadis S, Suh MC (2012) Identification of marneral synthase, which is critical for growth and development in Arabidopsis. Plant J 72:791–804. https://doi.org/10.1111/j.1365-313X.2012.05120.x

Guang C, Chen J, Sang S, Cheng S (2014) Biological functionality of soyasaponins and soyasapogenols. J Agric Food Chem 62:8247–8255. https://doi.org/10.1021/jf503047a

Hahm M-S, Sumayo M, Hwang Y-J, Jeon S-A, Park S-J, Lee JY, Ahn J-H, Kim B-S, Ryu C-M, Ghim S-Y (2012) Biological control and plant growth promoting capacity of rhizobacteria on pepper under greenhouse and field conditions. J Microbiol 50:380–385. https://doi.org/10.1007/s12275-012-1477-y

Han R-M, Tian Y-X, Liu Y, Chen C-H, Ai X-C, Zhang J-P, Skibsted LH (2009) Comparison of flavonoids and isoflavonoids as antioxidants. J Agric Food Chem 57:3780–3785. https://doi.org/10.1021/jf803850p

Henchion M, Hayes M, Mullen AM, Fenelon M, Tiwari B (2017) Future protein supply and demand: strategies and factors influencing a sustainable equilibrium. Foods 6:53. https://doi.org/10.3390/foods6070053

Heng L, Vincken J-P, van Koningsveld G, Legger A, Gruppen H, Boekel T, Roozen J, Voragen F (2006) Bitterness of saponins and their content in dry peas. J Sci Food Agric 86:1225–1231. https://doi.org/10.1002/jsfa.2473

Hughes GJ, Ryan DJ, Mukherjea R, Schasteen CS (2011) Protein digestibility-corrected amino acid scores (pdcaas) for soy protein isolates and concentrate: criteria for evaluation. J Agric Food Chem 59:12707–12712. https://doi.org/10.1021/jf203220v

Itkin M, Rogachev I, Alkan N, Rosenberg T, Malitsky S, Masini L, Meir S, Iijima Y, Aoki K, Vos R, Prusky D, Burdman S, Beekwilder J, Aharoni A (2011) GLYCOALKALOID

METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. Plant Cell 23:4507–4525. https://doi.org/10.1105/tpc.111.088732

Jia J, Ji R, Li Z, Yu Y, Nakano M, Long Y, Feng L, Qin C, Lu D, Zhan J, Xia R, Meyers BC, Liu B, Zhai J (2020) Soybean DICER-LIKE2 regulates seed coat color via production of primary 22-nucleotide small interfering RNAs from long inverted repeats. Plant Cell 32:3662–3673. https://doi.org/10.1105/tpc.20.00562

Kähkönen MP, Heinonen M (2003) Antioxidant Activity of Anthocyanins and Their Aglycons. J Agric Food Chem 51:628–633. https://doi.org/10.1021/jf025551i

Kanazashi Y, Hirose A, Takahashi I, Mikami M, Endo M, Hirose S, Toki S, Kaga A, Naito K, Ishimoto M, Abe J, Yamada T (2018) Simultaneous site-directed mutagenesis of duplicated loci in soybean using a single guide RNA. Plant Cell Rep 37:553–563. https://doi.org/10.1007/s00299-018-2251-3

Kemen AC, Honkanen S, Melton RE, Findlay KC, Mugford ST, Hayashi K, Haralampidis K, Rosser SJ, Osbourn A (2014) Investigation of triterpene synthesis and regulation in oats reveals a role for β-amyrin in determining root epidermal cell patterning. Proc Natl Acad Sci U S A 111:8679–8684. https://doi.org/10.1073/pnas.1401553111

Kim M-S, Jung YS, Jang D, Cho CH, Lee S-H, Han NS, Kim D-O (2022) Antioxidant capacity of 12 major soybean isoflavones and their bioavailability under simulated digestion and in human intestinal Caco-2 cells. Food Chem 374:131493. https://doi.org/10.1016/j.foodchem.2021.131493

Kosslak RM, Bookland R, Barkei J, Paaren HE, Appelbaum ER (1987) Induction of *Bradyrhizobium japonicum* common *nod* genes by isoflavones isolated from *Glycine max*. Proc Natl Acad Sci U S A 84:7428–7432. https://doi.org/10.1073/pnas.84.21.7428

Krishnamurthy P, Fujisawa Y, Takahashi Y, Abe H, Yamane K, Mukaiyama K, Son H-R, Hiraga S, Kaga A, Anai T, Tsukamoto C, Ishimoto M (2019) High-throughput screening and characterization of a high-density soybean mutant library elucidate the biosynthesis

pathway of triterpenoid saponins. Plant Cell Physiol 60:1082–1097. https://doi.org/10.1093/pcp/pcz025

Krokida A, Delis C, Geisler K, Garagounis C, Tsikou D, Peña-Rodríguez LM Katsarou D, Field B, Osbourn AE, Papadopoulou KK (2013) A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. New Phytol 200:675–690. https://doi.org/10.1111/nph.12414

Kudou S, Tonomura M, Tsukamoto C, Shimoyamada M, Uchida T, Okubo K (1992) Isolation and structural elucidation of the major genuine soybean saponin. Biosci Biotechnol Biochem 56:142–143. https://doi.org/10.1271/bbb.56.142

Kudou S, Tonomura M, Tsukamoto C, Uchida T, Sakabe T, Tamura N, Okubo K (1993) Isolation and structural elucidation of DDMP-conjugated soyasaponins as genuine saponins from soybean seeds. Biosci Biotechnol Biochem 57:546–550. https://doi.org/10.1271/bbb.57.546

Kumagai Y, Liu Y, Hamada H, Luo W, Zhu J, Kuroki M, Nagira Y, Taoka N, Katoh E, Imai R (2022) Introduction of a second "Green Revolution" mutation into wheat via in planta CRISPR/Cas9 delivery. Plant Physiol 188:1838–1842. https://doi.org/10.1093/plphys/kiab570

Kushiro T, Shibuya M, Ebizuka Y (1998) Beta-amyrin synthase–cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. Eur J Biochem 256:238–244. https://doi.org/10.1046/j.1432-1327.1998.2560238.x

Kuwabara C, Miki R, Maruyama N, Yasui M, Hamada H, Nagira Y, Hirayama Y, Ackley W, Li F, Imai R, Taoka N, Yamada T (2024) A DNA-free and genotype-independent CRISPR/Cas9 system in soybean. Plant Physiol kiae491. https://doi.org/10.1093/plphys/kiae491

Lacchini E, Erffelinck M-L, Mertens J, Marcou S, Molina-Hidalgo FJ, Tzfadia O, Venegas-Molina J, Cárdenas PD, Pollier J, Tava A, Bak S, Höfte M, Goossens A (2023) The saponin bomb: a nucleolar-localized β-glucosidase hydrolyzes triterpene saponins in *Medicago truncatula*. New Phytol 239:705–719. https://doi.org/10.1111/nph.18763

Lei Z, Watson BS, Huhman D, Yang DS, Sumner LW (2019) Large-scale profiling of saponins in different ecotypes of *Medicago truncatula*. Front Plant Sci 10:850. https://doi.org/10.3389/fpls.2019.00850

Lester GE, Lewers KS, Medina MB, Saftner RA (2012) Comparative analysis of strawberry total phenolics via Fast Blue BB vs. Folin–Ciocalteu: Assay interference by ascorbic acid. J Food Compos Anal 27:102–107. https://doi.org/10.1016/j.jfca.2012.05.003

Li C, Nguyen V, Liu J, Fu W, Chen C, Yu K, Cui Y (2019) Mutagenesis of seed storage protein genes in soybean using CRISPR/Cas9. BMC Res Notes 12:176. https://doi.org/10.1186/s13104-019-4207-2

Li Z, Liu Z-B, Xing A, Moon BP, Koellhoffer JP, Huang L, Ward T, Clifton E, Falco SC, Cigan AM (2015) Cas9-guide RNA directed genome editing in soybean. Plant Physiol 169:960–970. https://doi.org/10.1104/pp.15.00783

Liu Y, Zhao X, Gan F, Chen X, Deng K, Crowe SA, Hudson GA, Belcher MS, Schmidt M, Astolfi MCT, Kosina SM, Pang B, Shao M, Yin J, Sirirungruang S, Iavarone AT, Reed J, Martin LBB, El-Demerdash A, Kikuchi S, Misra RC, Liang X, Cronce MJ, Chen X, Zhan C, Kakumanu R, Baidoo EEK, Chen Y, Petzold CJ, Northen TR, Osbourn A, Scheller H, Keasling JD (2024) Complete biosynthesis of QS-21 in engineered yeast. Nature 629:937–944. https://doi.org/10.1038/s41586-024-07345-9

Lövgren Bengtsson K, Morein B, Osterhaus AD (2011) ISCOM technology-based Matrix M™ adjuvant: success in future vaccines relies on formulation. Expert Rev Vaccines 10:401–403. https://doi.org/10.1586/erv.11.25

Ma L, Yuan J, Qin H, Zhang M, Zhang F, Yu F, Tian Z, Wang G (2024) *GmMATE100* is involved in the import of soyasaponins A and B into vacuoles in soybean plants (*Glycine max* L.). J Agric Food Chem 72:9994–10004. https://doi.org/10.1021/acs.jafc.4c01774

Marone D, Mastrangelo AM, Borrelli GM, Mores A, Laidò G, Russo MA, Ficco DBM (2022) Specialized metabolites: physiological and biochemical role in stress resistance, strategies to

improve their accumulation, and new applications in crop breeding and management. Plant Physiol Biochem 172:48–55. https://doi.org/10.1016/j.plaphy.2021.12.037

Martin LBB, Kikuchi S, Rejzek M, Owen C, Reed J, Orme A, Misra RC, El-Demerdash A, Hill L, Hodgson H, Liu Y, Keasling JD, Field RA, Truman AW, Osbourn A (2024) Complete biosynthesis of the potent vaccine adjuvant QS-21. Nat Chem Biol 20:493–502. https://doi.org/10.1038/s41589-023-01538-5

Medina MB (2011) Simple and rapid method for the analysis of phenolic compounds in beverages and grains. J Agric Food Chem 59:1565–1571. https://doi.org/10.1021/jf103711c

Mertens J, Pollier J, Vanden Bossche R, Lopez-Vidriero I, Manuel Franco-Zorrilla J, Goossens A (2016) The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. Plant Physiol 170:194–210. https://doi.org/10.1104/pp.15.01645

Moses T, Papadopoulou KK, Osbourn A (2014) Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. Crit Rev Biochem Mol Biol 49:439–462. https://doi.org/10.3109/10409238.2014.953628

Mylona P, Owatworakit A, Papadopoulou K, Jenner H, Qin B, Findlay K, Hill L, Qi X, Bakht S, Melton R, Osbourn A (2008) *Sad3* and *sad4* are required for saponin biosynthesis and root development in oat. Plant Cell 20:201–212. https://doi.org/10.1105/tpc.107.056531

Nakayasu M, Ohno K, Takamatsu K, Aoki Y, Yamazaki S, Takase H, Shoji T, Yazaki K, Sugiyama A (2021) Tomato roots secrete tomatine to modulate the bacterial assemblage of the rhizosphere. Plant Physiol 186:270–284. https://doi.org/10.1093/plphys/kiab069

Nakayasu M, Takamatsu K, Kanai K, Masuda S, Yamazaki S, Aoki Y, Shibata A, Suda W, Shirasu K, Yazaki K, Sugiyama A (2023) Tomato root-associated *Sphingobium* harbors genes for catabolizing toxic steroidal glycoalkaloids. mBio 14:e00599-23. https://doi.org/10.1128/mbio.00599-23

Naoumkina MA, Modolo LV, Huhman DV, Urbanczyk-Wochniak E, Tang Y, Sumner LW, Dixon RA (2010) Genomic and coexpression analyses predict multiple genes involved in triterpene

saponin biosynthesis in *Medicago truncatula*. Plant Cell 22:850–866. https://doi.org/10.1105/tpc.109.073270

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

Okutani F, Hamamoto S, Aoki Y, Nakayasu M, Nihei N, Nishimura T, Yazaki K, Sugiyama A (2020) Rhizosphere modelling reveals spatiotemporal distribution of daidzein shaping soybean rhizosphere bacterial community. Plant Cell Environ 43:1036–1046. https://doi.org/10.1111/pce.13708

Papadopoulou K, Melton RE, Leggett M, Daniels MJ, Osbourn AE (1999) Compromised disease resistance in saponin-deficient plants. Proc Natl Acad Sci U S A 96:12923–12928. https://doi.org/10.1073/pnas.96.22.12923

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci 5:439–445. https://doi.org/10.1016/S1360-1385(00)01741-6

Pichersky E, Lewinsohn E (2011) Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol 62:549–566. https://doi.org/10.1146/annurev-arplant-042110-103814

Pollier J, Moses T, González-Guzmán M, De Geyter N, Lippens S, Vanden Bossche R, Marhavý P, Kremer A, Morreel K, Guérin CJ, Tava A, Oleszek W, Thevelein JM, Campos N, Goormachtig S, Goossens A (2013) The protein quality control system manages plant defence compound synthesis. Nature 504:148–152. https://doi.org/10.1038/nature12685

Poralla K, Hewelt A, Prestwich GD, Abe I, Reipen I, Sprenger G (1994) A specific amino acid repeat in squalene and oxidosqualene cyclases. Trends Biochem Sci 19:157–158. https://doi.org/10.1016/0968-0004(94)90276-3

Prieto C, Barrios D (2020) RaNA-Seq: interactive RNA-Seq analysis from FASTQ files to functional analysis. Bioinformatics 36:1955–1956. https://doi.org/10.1093/bioinformatics/btz854

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vil J (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47:W191–W198. https://doi.org/10.1093/nar/gkz369

Reed J, Orme A, El-Demerdash A, Owen C, Martin LBB, Misra RC, Kikuchi S, Rejzek M, Martin AC, Harkess A, Leebens-Mack J, Louveau T, Stephenson MJ, Osbourn A (2023) Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. Science 379:1252–1264. https://doi.org/10.1126/science.adf3727

Ribeiro B, Lacchini E, Bicalho KU, Mertens J, Arendt P, Vanden Bossche R, Calegario G, Gryffroy L, Ceulemans E, Buitink J, Goossens A, Pollier J (2020) A seed-specific regulator of triterpene saponin biosynthesis in *Medicago truncatula*. Plant Cell 32:2020–2042. https://doi.org/10.1105/tpc.19.00609

Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S, Tatsuzaki N, Tsukamoto C, Ishimoto M (2012) The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell 24:2123–2138. https://doi.org/10.1105/tpc.111.095174

Schaller H (2003) The role of sterols in plant growth and development. Prog Lipid Res 42:163–175. https://doi.org/10.1016/S0163-7827(02)00047-4

Schaller H (2004) New aspects of sterol biosynthesis in growth and development of higher plants. Plant Physiol Biochem 42:465–476. https://doi.org/10.1016/j.plaphy.2004.05.012

Seki H, Tamura K, Muranaka T (2015) P450s and UGTs: Key players in the structural diversity of triterpenoid saponins. Plant Cell Physiol 56:1463–1471. https://doi.org/10.1093/pcp/pcv062

Shen B, Schmidt MA, Collet KH, Liu Z-B, Coy M, Abbitt S, Molloy L, Frank M, Everard JD, Booth R, Samadar PP, He Y, Kinney A, Herman EM (2022) RNAi and CRISPR-Cas silencing E3-RING ubiquitin ligase AIP2 enhances soybean seed protein content. J Exp Bot 73:7285–7297. https://doi.org/10.1093/jxb/erac376

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

Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. FEBS Lett 584:2258–2264. https://doi.org/10.1016/j.febslet.2010.03.037

Shimoyamada M, Kudo S, Okubo K, Yamauchi F, Harada K (1990) Distributions of saponin constituents in some varieties of soybean plant. Agric Biol Chem 54:77–81. https://doi.org/10.1080/00021369.1990.10869887

Shiraiwa M, Harada K, Okubo K (1991a) Composition and content of saponins in soybean seed according to variety, cultivation year and maturity. Agric Biol Chem 55:323–331. https://doi.org/10.1271/bbb1961.55.323

Shiraiwa M, Kudo S, Shimoyamada M, Harada K, Okubo K (1991b) Composition and structure of "group A saponin" in soybean seed. Agric Biol Chem 55:315–322. https://doi.org/10.1271/bbb1961.55.315

Sim J, Kuwabara C, Sugano S, Adachi K, Yamada T (2023) Recent advances in the improvement of soybean seed traits by genome editing. Plant Biotechnol (Tsukuba) 40:193–200. https://doi.org/10.5511/plantbiotechnology.23.0610a

Spanova M, Daum G (2011) Squalene – biochemistry, molecular biology, process biotechnology, and applications. Eur J Lipid Sci Technol 113:1299–1320. https://doi.org/10.1002/ejlt.201100203

Spengler B, Hubert M (2002) Scanning microprobe matrix-assisted laser desorption ionization (SMALDI) mass spectrometry: instrumentation for sub-micrometer resolved LDI and MALDI surface analysis. J Am Soc Mass Spectrom 13:735–748. https://doi.org/10.1016/S1044-0305(02)00376-8

Subramanian S, Stacey G, Yu O (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. Plant J 48:261–273. https://doi.org/10.1111/j.1365-313X.2006.02874.x

Sugano S, Hirose A, Kanazashi Y, Adachi K, Hibara M, Itoh T, Mikami M, Endo M, Hirose S, Maruyama N, Abe J, Yamada T (2020) Simultaneous induction of mutant alleles of two allergenic genes in soybean by using site-directed mutagenesis. BMC Plant Biol 20:513. https://doi.org/10.1186/s12870-020-02708-6

Sun C, Ma S, Li L, Wang D, Liu W, Liu F, Guo L, Wang X (2021) Visualizing the distributions and spatiotemporal changes of metabolites in *Panax notoginseng* by MALDI mass spectrometry imaging. J Ginseng Res 45:726–733. https://doi.org/10.1016/j.jgr.2021.04.001

Sundaramoorthy J, Park GT, Komagamine K, Tsukamoto C, Chang JH, Lee J-D, Kim JH, Seo HS, Song JT (2019) Biosynthesis of DDMP saponins in soybean is regulated by a distinct UDP-glycosyltransferase. New Phytol 222:261–274. https://doi.org/10.1111/nph.15588

Suzuki H, Seki H, Muranaka T (2021) Insights into the diversification of subclade IVa bHLH transcription factors in Fabaceae. BMC Plant Biol 21:109. https://doi.org/10.1186/s12870-021-02887-w

Takagi K, Nishizawa K, Hirose A, Kita A, Ishimoto M (2011) Manipulation of saponin biosynthesis by RNA interference-mediated silencing of β-amyrin synthase gene expression in soybean. Plant Cell Rep 30:1835–1846. https://doi.org/10.1007/s00299-011-1091-1

Takagi K, Yano R, Tochigi S, Fujisawa Y, Tsuchinaga H, Takahashi Y, Takada Y, Kaga A, Anai T, Tsukamoto C, Seki H, Muranaka T, Ishimoto M (2018) 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 156:96–105. https://doi.org/10.1016/j.phytochem.2018.09.002

Takamatsu K, Toyofuku M, Okutani F, Yamazaki S, Nakayasu M, Aoki Y, Kobayashi M, Ifuku K, Yazaki K, Sugiyama A (2023) α-Tomatine gradient across artificial roots recreates the

recruitment of tomato root-associated *Sphingobium*. Plant Direct 7:e550. https://doi.org/10.1002/pld3.550

Tamura K, Yoshida K, Hiraoka Y, Sakaguchi D, Chikugo A, Mochida K, Kojoma M, Mitsuda N, Saito K, Muranaka T, Seki H (2018) The basic helix-loop-helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. Plant Cell Physiol 59:778–791. https://doi.org/10.1093/pcp/pcy046

Tsukamoto C, Shimada S, Igita K, Kudou S, Kokubun M, Okubo K, Kitamura K (1995) Factors affecting isoflavone content in soybean seeds: Changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. J Agric Food Chem 43:1184–1192. https://doi.org/10.1021/jf00053a012

Tsuno Y, Fujimatsu T, Endo K, Sugiyama A, Yazaki K (2018) Soyasaponins: a new class of root exudates in soybean (*Glycine max*). Plant Cell Physiol 59:366–375. https://doi.org/10.1093/pcp/pcx192

Tsurumi S, Tsujino Y (1995) Chromosaponin I stimulates the growth of lettuce roots. Physiol Plant 93:785–789. https://doi.org/10.1111/j.1399-3054.1995.tb05132.x

Tsurumi S, Wada S (1995) Chromosaponin I stimulates the elongation of cortical cells in lettuce roots. Plant Cell Physiol 36:925–929. https://doi.org/10.1093/oxfordjournals.pcp.a078840

Vives-Peris V, Gómez-Cadenas A, Pérez-Clemente RM (2018) Salt stress alleviation in citrus plants by plant growth-promoting rhizobacteria *Pseudomonas putida* and *Novosphingobium* sp. Plant Cell Rep 37:1557–1569. https://doi.org/10.1007/s00299-018-2328-z

Vogt T (2010) Phenylpropanoid Biosynthesis. Molecular Plant 3:2–20. https://doi.org/10.1093/mp/ssp106

Wang J, Kuang H, Zhang Z, Yang Y, Yan L, Zhang M, Song S, Guan Y (2019) Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. Crop J 8:432–439. https://doi.org/10.1016/j.cj.2019.08.008

Wegel E, Koumproglou R, Shaw P, Osbourn A (2009) Cell type–specific chromatin decondensation of a metabolic gene cluster in oats. Plant Cell 21:3926–3936. https://doi.org/10.1105/tpc.109.072124

Weiss T, Crisp PA, Rai KM, Song M, Springer NM, Zhang F (2022) Epigenetic features drastically impact CRISPR-Cas9 efficacy in plants. Plant Physiol 190:1153–1164. https://doi.org/10.1093/plphys/kiac285

White LJ, Ge X, Brözel VS, Subramanian S (2017) Root isoflavonoids and hairy root transformation influence key bacterial taxa in the soybean rhizosphere. Environ Microbiol 19:1391–1406. https://doi.org/10.1111/1462-2920.13602

Xu JL, Shin J-S, Park S-K, Kang S, Jeong S-C, Moon J-K, Choi Y (2017) Differences in the metabolic profiles and antioxidant activities of wild and cultivated black soybeans evaluated by correlation analysis. Food Res Int 100:166–174. https://doi.org/10.1016/j.foodres.2017.08.026

Yamada T, Takagi K, Ishimoto M (2012) Recent advances in soybean transformation and their application to molecular breeding and genomic analysis. Breed Sci 61:480–494. https://doi.org/10.1270/jsbbs.61.480

Yano R, Takagi K, Takada Y, Mukaiyama K, Tsukamoto C, Sayama T, Kaga A, Anai T, Sawai S, Ohyama K, Saito K, Ishimoto M (2017) Metabolic switching of astringent and beneficial triterpenoid saponins in soybean is achieved by a loss-of-function mutation in cytochrome P450 72A69. Plant J 89:527–539. https://doi.org/10.1111/tpj.13403

Yano R, Takagi K, Tochigi S, Fujisawa Y, Nomura Y, Tsuchinaga H, Takahashi Y, Takada Y, Kaga A, Anai T, Tsukamoto C, Seki H, Muranaka T, Ishimoto M (2018) 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 59:792–805. https://doi.org/10.1093/pcp/pcy019

Yates PS, Roberson J, Ramsue LK, Song B-H (2021) Bridging the gaps between plant and human health: a systematic review of soyasaponins. J Agric Food Chem 69:14387–14401. https://doi.org/10.1021/acs.jafc.1c04819

Yoshiki Y, Kahara T, Okubo K, Sakabe T, Yamasaki T (2001) Superoxide- and 1,1-diphenyl-2-picrylhydrazyl radical-scavenging activities of soyasaponin beta g related to gallic acid. Biosci Biotechnol Biochem 65:2162–2165. https://doi.org/10.1271/bbb.65.2162

Yoshiki Y, Okubo K (1995) Active Oxygen Scavenging Activity of DDMP (2,3-Dihydro-2,5-dihydroxy-6-methyl-4*H*-pyran-4-one) Saponin in Soybean Seed. Biosci Biotechnol Biochem 59:1556–1557. https://doi.org/10.1271/bbb.59.1556

Zhang Q, Yu S, Lyu Y, Zeng W, Zhou J (2021) Systematically engineered fatty acid catabolite pathway for the production of (2*S*)-naringenin in *Saccharomyces cerevisiae*. ACS Synth Biol 10:1166–1175. https://doi.org/10.1021/acssynbio.1c00002

Zhou H, Lin J, Johnson A, Morgan RL, Zhong W, Ma W (2011) *Pseudomonas syringae* type III effector HopZ1 targets a host enzyme to suppress isoflavone biosynthesis and promote infection in soybean. Cell Host Microbe 9:177–186. https://doi.org/10.1016/j.chom.2011.02.007

Zilic S, Dodig D, Vancetovic J, Grcic N, Peric V, Titan P, Maksimovic V (2019) Composition of anthocyanins in colored grains and the relationship of their non-acylated and acylated derivatives. Pol J Food Nutr Sci 69:137–146. https://doi.org/10.31883/pjfns/105100

Note

Part of the results presented in this thesis has been published in Plant Cell Reports.

Asa H, Kuwabara C, Matsumoto K, Shigeta R, Yamamoto T, Masuda Y, Yamada T (2025) Simultaneous site-directed mutagenesis for soybean β-amyrin synthase genes via DNA-free CRISPR/Cas9 system using a single gRNA. Plant Cell Rep 44:40. https://doi.org/10.1007/s00299-025-03433-w

Appendix

Supplemental Materials and Methods

Preparations of soybean seed specimens and MALDI-TOF MSI

Mature seeds were placed on wet paper towels and incubated for 9 h at 25°C and 4 h on ice in the dark. Water-absorbed seeds were frozen in a freezer at –30°C. Frozen seeds was immersed into 2% carboxymethyl cellulose (CMC) solution, the volume depending on the size of the seeds, in a container made of aluminum fil, and cooled down to –196°C by dipping the container into liquid nitrogen. A longitudinal thin (40 μm) specimens from a seed was cut on a cryomicrotome (CM 3050 S, Leica Biosystems) at a chamber temperature of –20°C along with the central axis of the hypocotyl. The specimen was mounted on an indium tin oxide (ITO) coated glass slide (Bruker), and dried quickly using a vacuum desiccator at 21.3 kPa for 30 min. The dried specimens were stored at –80°C until analysis using MALDI-TOF MSI.

The soybean seed specimens were sprayed with 1,5-Diaminonaphthalene (DAN) solution (10 mg DAN in 1 mL of 70% (v/v) acetonitrile) as a matrix, at an amount of 250 μL/cm^2 on the ITO glass slide using an air brush (MX 2370, Anest Iwata Corp.). MALDI-TOF MSI analysis was performed by ultrafleXtreme (Bruker) equipped with a Nd:YAG laser (355 nm) producing 0.5 s pulses with the repetition rate set to 1,000 Hz, with an accelerating voltage of 20 kV and the reflector negative-ion mode with delayed extraction at 280 ns. The laser spot size was set to “medium”, with an approximate diameter of 50 μm. The interval of the probing spots on the specimen was set to 150 and 50 μm for the analysis of whole seeds and hypocotyls, respectively. All mass spectra were calibrated with cesium iodide previously loaded at three points on the slide (outside of the specimen), and the *m/z* shifts of the three points were confirmed to be within $\pm$ 0.1. The analysis was performed without any biological replicates. The fleximaging 5.0 (Bruker) was employed to process the raw data and construct the selected ion maps upon acquired mass spectra at every MSI grid array coordinate.

Acknowledgement

本研究の遂行にあたり、ご指導をいただきました主査の北海道大学大学院農学研究院の山田哲也准教授、副査として本論文をご校閲いただきました金澤章教授、久保友彦教授、Maria Stefanie Dwiyanti 助教にこの場をお借りして心より御礼申し上げます。山田哲也准教授には、ご多忙な中でも常に研究内容をはじめ研究に取り組む姿勢から今後の研究を取り巻く情勢まであらゆる事柄について議論及び相談のために時間を設け、懇切丁寧にご指導をいただきました。また、研究に限らず課外活動も含めたあらゆる場面で私の意思を尊重し、自由な学生生活を送るための多大なご支援とご配慮をいただきました。本研究に係る充実した 5 年間の研究生活を授けていただきましたことを、謹んで深謝申し上げます。金澤章教授ならびに Maria Stefanie Dwiyanti 助教には、研究発表やセミナーをはじめとして、随所でご助言を賜りましたことを、厚く御礼申し上げます。

また、大阪大学大学院工学研究科の關光教授と同修士課程の森田遙絵さんには、GC-MS によるステロール分析においてご協力を賜りました。北海道大学大学院農学研究院の鈴木卓特任教授と同農学院博士課程の藤木卓巳さんには、FBBB 法によるポリフェノール分析及び MALDI-TOF MSI による DDMP ソヤサポニンのイメージングにおいてご協力を賜りました。特に、藤木卓巳さんには FBBB 法の技術指導、MALDI-TOF MS による分析、HPLC 分析に関する技術的な相談に加え、共同での研究資金の獲得など研究活動の多くの場面でご協力いただきました。この場をお借りして心より御礼申し上げます。

また、植物遺伝資源学研究室研究員の宇梶慎子博士、遠藤亜矢さん、徐美蘭博士には、実験技術の基礎から研究に対する姿勢、キャリア設計まで多くの面でご助言をいただきました。研究補助員の鈴木元美さん、野口早百合さん、山本貴子さんには、試薬の調製や備品の管理など研究活動に欠かせない多くの業務において多大なご支援をいただきました。事務員の佐々木眞美さん、中野育枝さん、半澤文さんには、備品の発注や出張の手配等の事務的な業務においてご支援をいただきました。

最後に、研究活動だけでなく日々の学生生活を共に過ごした植物遺伝資源学、細胞工学、植物病原学の諸研究室の学生、農学院博士課程の同期の皆様にも感謝申し上げます。

令和 8 年　1 月

麻　裕毅