Elucidation of steroidal glycoalkaloid biosynthesis in Solanum species.

Akiyama, Ryota

(Degree)
博士（農学）

(Date of Degree)
2020-03-25

(Date of Publication)
2022-03-25

(Resource Type)
doctoral thesis

(Report Number)
甲第7797号

(URL)
https://hdl.handle.net/20.500.14094/D1007797

Doctoral Dissertation

Elucidation of steroidal glycoalkaloid biosynthesis in *Solanum* species

ナス属植物におけるステロイドグリコアルカロイド

生合成の解明

January, 2020

Ryota AKIYAMA

Graduate School of Agricultural Science

Kobe University

Contents

General Introduction

Steroidal glycoalkaloids (SGAs) and are nitrogen-containing specialized metabolites which mainly occur in Solanaceae and Liliaceae plant families[1,2]. This class of compounds are famous as toxic and anti-nutritional substances in major Solanaceae vegetable crops such as Potato (*Solanum tuberosum*), Tomato (*Solanum lycopersicum*), and Eggplant (*Solanum melongena*)[3]. SGAs are associated with plant resistance to pathogens and predators due to their toxicity to a wide range of organisms including bacteria, fungi, insect and animal[4–6]. The SGA α-solanine and α-chaconine are the principal toxic substances in potato. These SGAs cause gastrointestinal and neurological disorders. The mechanisms of toxicity include the disruption of membranes and the inhibition of acetylcholine esterase activity[7]. Therefore, since these potato SGAs are toxic and cause a bitter taste, reducing SGA level in potato tubers is an important focus of potato breeding[4]. Similarly, α-tomatine exists in all tomato plant tissues, but it is richly accumulated in the leaves and immature fruit, with its content decreasing during fruit ripening[5].

SGAs consist of two structural components; the aglycone unit composed of C_{27} steroid containing and oligosaccharide attached to the hydroxy group at C-3[6] (Fig. I). The aglycones of SGAs are broadly classified into five different types depending on their structures; solanidanes, spirosolanes, epiminocholestanes, α-epiminocyclohemiketals, and 3-aminospirostanes[8] (Fig. II). Most of the SGAs produced in *Solanum* plants composed of spirosolane-type or solanidane-type of aglycone[6]. The structural difference between spirosolane and solanidane is occurred in the structure of E,F-rings (Fig. II). Minor structural variations of these two ring types such as Δ5 saturated/unsaturated or isomerization at C-22, in combination with various sugar moieties, generate enormous

structural diversity SGAs and their chemical structures determine their specific biological activities[6,9–13].

Potato typically contains solanidane-type SGAs. α-Solanine and α-chaconine, which share the same solanidane-type of aglycone solanidine (Fig. I), comprise upward of 90% of the total SGA contents in domesticated potato. Additionally, more than 50 different SGAs, including spirosolane-type SGAs, have been identified in a variety of commercial cultivars and wild potato species[14,15]. In tomato, α-tomatine and dehydrotomatine, which are composed of spirosolane-type aglycone, are major SGAs predominantly found in leaves and immature fruits of tomato[5,9,10] (Fig. I). Although more than 100 SGAs have been described in various tissues and developmental stages[16–18], accumulation of solanidane-type compounds have not been reported to the best of our knowledge. Eggplant produces two types of spirosolane-type SGA; α-solasonine and α-solamargine, as major SGAs[19] (Fig. I).

SGAs are biosynthesized from common precursor, cholesterol[20]. The SGA biosynthesis can be divided into two main parts; SA aglycones formation part and glycosylation part[16,21] (Fig. III). During aglycones formation part, cholesterol is subsequently modified through multiple reaction steps including hydroxylation, transamination, E- and F-ring closure[2,4,22,23]. Recently, three cytochrome P450 monooxygenases, a 2-oxoglutarate dependent dioxygenase (DOX), and an aminotransferase were reported to be required for SA aglycones formation part during SGA biosynthesis in potato and tomato[14,21,24]. In addition, several UDP-glycosyltransferases (UGTs) involved in the glycosylation steps of SGA biosynthesis have been identified in potato and tomato[16,25–28]. These gene discovery open the way for controlling toxic SGA contents in important food crops of the Solanaceae, through

genetic engineering or more classical breeding[20,21,24,29]. However, genes and pathways generating the structural diversity of aglycone structures, such as E and F-ring formation mechanism and the reduction of the C-5,6 double bond reactions, are completely unknown.

This study was designed to modify the SGA profile by genetic manipulation of known SGA biosynthesis genes and to identify novel SGA biosynthesis genes. In chapter I, CRIPSR/Cas9-mediated gene knockout of *St16DOX*, which is known SGA biosynthetic gene, was performed in hairy root transformation. *St16DOX*-disrupted potato hairy root showed no accumulation of SGAs, and this result have provided clues for breeding of SGA-free potato. Furthermore, combination of CRISPR/Cas9-mediated genome editing technique and hairy root transformation enabled rapid functional analyses of candidate genes. Chapter 2 describes identification of steroid 5α-reductase required for α-tomatine biosynthesis in tomato. Biochemical analysis of the recombinant proteins and CRISPR/Cas9 mediated genome editing revealed that SlS5αR2 is responsible for the C5α reduction in α-tomatine. In chapter 3, we elucidate the slanidane-skeleton formation mechanism in potato. Feeding experiment using *St16DOX*-knockout hairy roots revealed that solanidane-type SAG is biosynthesized from spirosolane-type SGA in potato. Additionally, 2-oxoglutalate dependent dioxygenase (DOX), named StDOX130, was identified as spirosolane E-ring cleavage enzyme involved in the first step of bioconversion from spirosolane-type compound to solanidane-type compound. Based on these findings, chemical structural evolution of SGAs in *Solanum* species is discussed in concluding discussion.

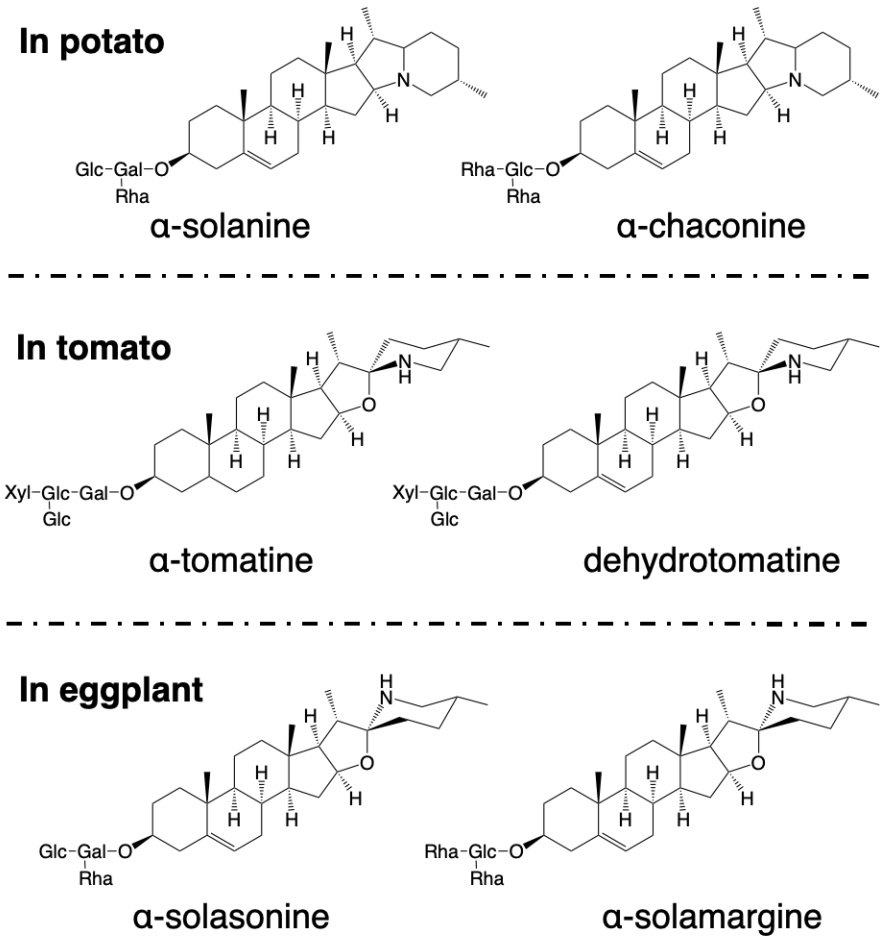


Fig. I. The structures of steroidal glycoalkaloid (SGAs) in *Solanum* plants

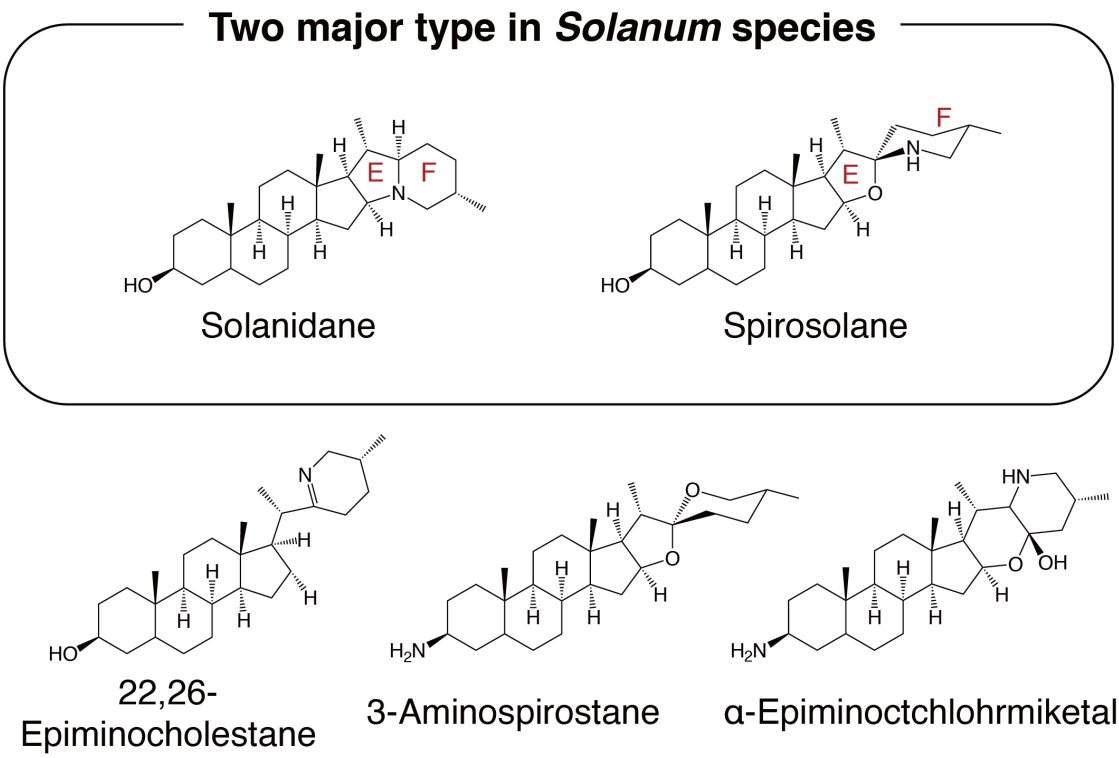


Fig. II. Structures of the aglycone units of steroidal glycoalkaloids (SGAs).

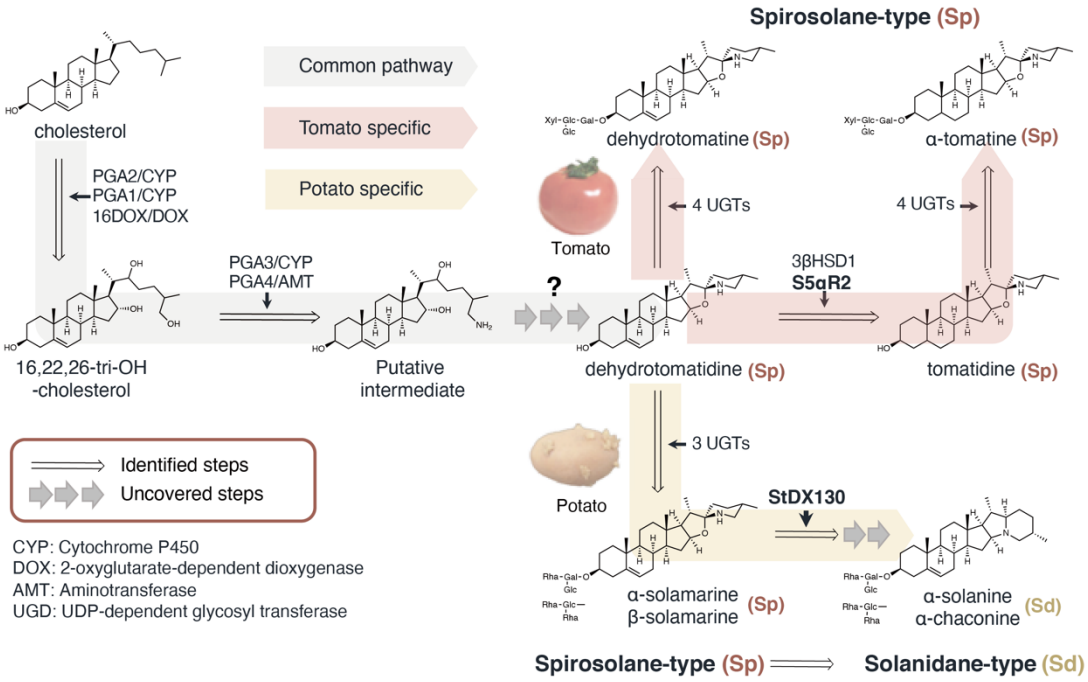


Fig. III. The putative biosynthetic pathway of SGAs in potato and tomato. White arrows indicate identified reaction stages. Gray arrows indicated unidentified reaction stages.

Chapter1

Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the *St16DOX* gene

Introduction

Potato (*Solanum tuberosum*) is one of the worldwide major food crops, and however, potato produces toxic compounds of steroidal glycoalkaloids (SGAs), α-solanine and α-chaconine particularly in the flowers and the tuber sprouts at high levels[22,30]. These SGAs confer a bitter taste, and exhibit toxicity against human[4]. Edible tuber containing over 200mg/ kg fresh weight of SGAs are deemed unsafe for human consumption (7). Therefore, reducing the SGA contents in the tubers is an important focus of potato breeding[4].

Concern for food safety prompted research about SGA biosynthesis. SGAs are biosynthesized from cholesterol[20], which is subsequently modified by the multiple reaction steps during SGA biosynthesis: hydroxylation at C-16, C-22, and C-26, transamination at C-26, E− and F-rings closure, and glycosylation at the C-3 hydroxy group[2,4,22,23] (Fig. 1-1). To date, several SGA biosynthetic genes have been isolated and characterized as described in general introduction. Recently, A 2-oxoglutrate-dependent dioxygenase gene, named *St16DOX* was identified as SGA biosynthetic gene. St16DOX catalyzes 16α-hydroxylation of (22S)-22,26-dihydroxycholesterol in earlier step of conversion of cholesterol to SGAs[24] (Fig. 1-1). *St16DOX*-silencing by RNA interference in transgenic potato plants led to significantly reduced endogenous SGA levels and furthermore, did not affect potato tuber yield[24], indicating that *St16DOX* preferable target for controlling SGA contents in edible tuber of potato.

The emergence of genome editing technologies using custom engineered nucleases enabling direct and precise genome modification have allowed rapid generation of desirable loss-of-function mutant via site-specific gene deletion or insertion. The most widespread genome editing technologies in current use are Transcriptional Activator-Like Effector Nucleases (TALEN) and Clustered Regulatory Interspaced Short Palindromic Repeat (CRISPR) and CRISPR-Associated protein 9 (Cas9) system (CRISPR/Cas9). TALEN and CRISPR/Cas9 system have been applied to induce site-directed mutagenesis in potato[30]. Sawai et al.[20] utilized TALEN to disrupt *sterol side chain reductase 2* (*SSR2*) which is involved in cholesterol and related SGAs. The transgenic potato with disrupted *SSR2* gene showed significantly reduced SGA levels without affecting plant growth. In addition, transient expression of TALEN in protoplast was performed for the disruption of the *ACETOLACTATE SYNTHASE* (*ALS*) and *VACUOLAR INVERTASE* (*VInv*) genes[31,32]. Knockout of potato *ALS* was also performed using CRISPR/Cas9 mediated by *A. tumefaciens*[33]. In another CRISPR/Cas9 study, the GRANULE-BOUND *STARCH SYNTHASE* (*GBSS*) gene was targeted by protoplast transfection system[34].

Here, we applied the CRISPR/Cas9 system to the production of SGA-free potato via disruption of the *St16DOX*. In this study, we selected nine candidate guide RNA (gRNA) sequences targeted for *St16DOX* using *in silico* analysis, and these gRNAs were inserted into a new CRISPR/Cas9 vinery vector designated as pMgP237-2A-GFP that permit expression of multiplex gRNAs based on the pre-tRNA processing system in plant. We also used a hairy root transformation system for a rapid evaluation of the efficacy of the constructed pMgP237-2A-GFP vectors. Consequently, generation of SGA-free transgenic potato hairy roots was achieved.

Materials and methods

Chemicals

Authentic samples of α-solanine and α-chaconine were purchased from Sigma-Aldrich. Authentic compound of (22*S*,25*S*)-22,26-dihydroxycholesterol was synthesized as previously described[24].

Sequencing analysis of the St16DOX gene from S. tuberosum cv. Mayqueen

Genome DNA was extracted from the leaves of the in vitro-grown shoots of *S. tuberosum* cv. Mayqueen using the DNeasy Plant Mini Kit (QIAGEN) according to the manufacturer's instructions. Contaminant total RNAs in the extracted genome DNA were removed by RNase (TOYOBO). A partial fragment of the 5′-terminal side of Mayqueen *St16DOX* genome DNA was PCR amplified using primer set 1 and 2 (Table 1-1). The two primers were designed from potato unigene sequence PGSC0003DMG400011751 obtained by BLASTN search of the *St16DOX* nucleotide sequence against the Spud DB potato genomics resource (http://potato.plantbiology.msu.edu/). The PCR fragment was applied to agarose gel electrophoresis and then purified from the gel using the Wizard® SV Gel and PCR Clean-Up System (Promega, Japan) according to the manufacturer's instructions, and cloned into pCR™4Blunt-TOPO® (Invitrogen). Randomly selected eight clones were used for Sanger sequencing analysis, and the sequences were determined using ABI 3130 genetic analyzer (Applied Biosystems) and analyzed using a biological sequence alignment editor, BioEdit (http://www.mbio.ncsu.edu) (Fig. 1-2).

Construction of CRISPR/Cas9 vectors

In this study, we used the CRISPR/Cas9 binary vector pMgP237-2A-GFP to express multiplex gRNAs based on the pre-tRNA processing system in plant[35]. pMgP237-2A-GFP was constructed from pEgP237-2A-GFP[36]; briefly, the sequence for 2x*Bsa*I sites-gRNA scaffold under Arabidopsis *U6 snRNA-26* promoter of pEgP237-2A-GFP was replaced to tRNA scaffold-2x*Bsa*I sites-gRNA scaffold (Fig. 1-3C). To design gRNA target sequences in Mayqueen *St16DOX* genome DNA, we conducted *in silico* analysis using a web tool Design sgRNAs for CRISPRko (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) and Cas-OT software[37]. Then, we selected nine candidate target sequences designated T1-T9 (Fig. 1-2 and 1-3A). To enhance the transcription from the U6 promotor, one G was added to the 5′-end of T1, T4, T5, T6, and T8 (Table 1-2). Five DNA fragments that were composed of gRNA scaffold and tRNA scaffold between two target sequences: T1/T2, T3/T4, T5/T6, and three target sequences: T7/T8/T9 (Fig. 1-3B and C), were generated by PCR using pMD-gtRNA containing gRNA and tRNA scaffolds as a template and primer sets containing restriction enzyme *Bsa*I sites: 3 and 4 for the T1/T2 fragment, 5 and 6 for the T3/T4 fragment, 7 and 8 for the T5/T6 fragment, 9 and 10 for the T7/T8 fragment, and 11 and 12 for the T8/T9 fragment (Fig. 1-3B). Next, each PCR fragment shown in Fig. 1C was ligated into the *Bsa*I sites of the CRISPR/Cas9 vector pMgP237-2A-GFP using Golden Gate Cloning methods[38] to generate four individual CRISPR/Cas9 vinery vectors named T1/T2-pMgP237, T3/T4-pMgP237, T5/T6-pMgP237, and T7/T8/T9-pMgP237 (Fig. 1-3C). Each vector was electroporated into *Agrobacterium rhizogenes* strain ATCC15834 for generation of Mayqueen hairy roots.

Plant growth conditions and generation of the transformed hairy roots

The Mayqueen in vitro-grown shoots were grown under conditions of 20 °C with 16 h light/8 h dark on Murashige and Skoog (MS) medium solidified with 0.3% (w/v) Gelrite® (Wako) containing 2% (w/v) sucrose. The in vitro-grown shoots were infected with *A. rhizogenes* strain ATCC15834 harboring each constructed CRISPR/Cas9 vector as previously described [39], and cultured for three weeks under at 20 °C in the dark on Gamborg B5 (B5) medium solidified with 0.3% (w/v) Gelrite (Wako) containing 2% (w/v) sucrose. Hairy roots emerging from shoots were cultured as previously described [40] with minor modifications. They were cultured for one week on MS medium solidified with 0.3% (w/v) Gelrite (Wako) containing 2% (w/v) sucrose and 250 μg ml^{-1} cefotaxime for removing *A. rhizogenes*, and then cultured twice each week on B5 medium solidified with Gelrite containing sucrose, and cefotaxime. The established independent lines were suspended for three weeks with 50 ml of liquid B5 medium with 2% (w/v) sucrose and 250 μg ml^{-1} cefotaxime in 200 ml glass flasks. Transformants were selected by genomic PCR targeting a partial region of T-DNA fragment integrated into the genome using primer set 13 and 14 (Table 1-1).

SGA analysis

Steroidal compounds accumulated in the harvested hairy roots were extracted from 100 mg fresh weight samples with methanol, and the extracts were analyzed using liquid chromatography-mass spectrometry (LC-MS) as previously described[40]. Selected ion recording (SIR) mode with *m/z* 868.9 and *m/z* 852.8 was used to detect α-solanine and α-chaconine, respectively, and the two SGAs were quantified using each calibration curve as previously described[21]. For the aglycone analysis, the acid hydrolysates of the

methanol extracts were prepared, trimethylsilylated, and analyzed by gas chromatography-mass spectrometry (GC-MS) as previously described[24].

Genotyping of the transgenic hairy roots

Genomic DNA was extracted from each established transgenic hairy root as described above. Region surrounding *St16DOX* sequences targeted by each CRISPR/Cas9 vector were PCR-amplified using primer sets: 15 and 16, 17 and 18, 19 and 20, and 21 and 16, respectively (Table 1-1). The PCR amplicons were separated on 2.5% agarose gel. The PCR reaction mixture of two independent lines T1/T2_#6 and T7/T8/T9_#4 was cleaned using the Wizard® SV Gel and PCR Clean-Up System (Promega, Japan) according to the manufacturer's instructions. Then, the PCR amplicons were cloned into pCR™4Blunt-TOPO® and the sequences of randomly selected 16 clones were analyzed described above, compared with those of wild-type potato.

Results

Construction of CRISPR/Cas9 vectors for St16DOX-knockout

In this study, CRISPR/Cas9 vector, designated as pMgP237-2A-GFP, that enable multiple gRNAs expression by endogenous pre-tRNA processing mechanism were constructed (Fig. 1-3C). To design precise gRNA that target *St16DOX* of cv. Mayqueen, we first determined the genomic sequence of the Mayqueen *St16DOX* gene in the region from exon1 to exon4 (Fig. 1-2). Next, we selected nine candidate target sequences (T1-T9) of the gRNAs (Table 1-2; Fig. 1-2) based on *in silico* analysis using Design sgRNAs for CRISPRko and Cas-OT software[37]. Cleavage sites by T7 and T8 were present in exon1 of the *St16DOX* gene, those by T1, T2, and T9 were in exon2, and those by T3, T4,

T5, and T6 were in exon3 (Fig. 1-3A), since Cas9 cleaves specifically between the 3rd and 4th nucleotides from the PAM sequence (NGG)[41]. Targets T1, T3-T6, and T8 were designed within exon1, exon2, or exon3, respectively (Fig. 1-3A and Fig. 1-2). Although T2, T7, and T9 were designed boundary of exon2-intron2 or 5′UTR-exon1, putative cleavage sites of the three targets were positioned in the exon of *St16DOX* (Fig. 1-2). Then, we constructed four vectors carrying target sequence sets as followings: T1/T2-pMgP237, T3/T4-pMgP237, T5/T6-pMgP237, and T7/T8/T9-pMgP237 (Fig. 1-3C).

Endogenous SGA analysis of the transgenic hairy roots

Each of the four constructed CRISPR/Cas9 vectors was individually transformed into in vitro-grown shoots of cv. Mayqueen by *A. rhizigenes*-mediated hairy root transformation, and transgenic hairy roots were selected by genomic PCR targeting a partial region of the T-DNA region integrated into the genome (Fig. 1-4). Then, six transgenic lines for T1/T2-pMgP237, nine lines for T3/T4-pMgP237, five lines for T5/T6-pMgP237, and five lines for T7/T8/T9-pMgP237 were generated. Because the analysis of SGA levels in the transgenic hairy roots is easier than the analysis of individual mutations at each target site, their endogenous SGA levels were preliminarily analyzed by LC-MS. Among all the transgenic hairy roots established, two independent lines, T1/T2_#6 and T7/T8/T9_#4, showed no detectable accumulation of SGAs (α-solanine and α-chaconine). The other lines contained significant amounts of SGAs although some lines had significantly lower SGA levels than control (Fig. 1-5).

PCR-based analysis of mutation by agarose gel electrophoresis

To detect mutation in the CRISPR/Cas9 transgenic hairy roots, we conducted agarose gel electrophoresis. While the control lines transformed with the empty vector gave a single band on the gel, extra bands were observed in the following transgenic lines: T1/T2_#1, T1/T2_#6, T3/T4_#2, T3/T4_#5, T3/T4_#6, and T7/T8/T9_#4 (Fig. 1-6), suggesting the occurrence of the large segment deletion in the *St16DOX* gene.

Quantification of SGA in the positive lines

T1/T2_#1, T3/T4_#2, and T3/T4_#5, showed significant reduction of SGAs but incomplete abolition of SGAs (Fig. 1-7A), suggesting their hetero-allelic or mosaic mutation. In contrast, SGAs were not detected in T1/T2_#1 and T7/T8/T9_#4 (Fig. 1-7A), suggesting their tetra-allelic mutation on the *St16DOX* gene. The two SGA-free lines gave two new peaks with retention times of 25.3 min and 26.1 min (Fig. 1-7B), and deduced parent ions of *m/z* 1068 and *m/z* 1052, respectively (Fig. 1-7C). Both the two peaks had a mass fragment ion of *m/z* 383 corresponding to that of dihydroxycholesterol[24]. Next, to investigate the aglycone structure of the steroidal compounds accumulated in SGA-free potato hairy roots, the trimethylsilyl derivatives of the hydrolysates of the extracts from these lines were analyzed using GC-MS. These lines had a distinct peak with a retention time of 22.8 min and a mass fragment of *m/z* 171 corresponding the trimethylsilyl derivative of 22,26-dihydroxycholesterol (Fig. 1-7D). These results indicate that transgenic hairy roots showing a SGA-free phenotype accumulate the glycosides of 22,26-dihydroxycholesterol. Mass fragments of the two glycosides (Rt 25.3 min and Rt 26.1 min) shown in Fig. 1-7C indicated that they were 22,26-dihydroxycholesterols with four monosaccharides at least, suggesting that an additional

monosaccharide is bound at the C-22 or C-26 hydroxy groups besides attaching solatriose or chacotriose at the C-3 hydroxy group.

Genotyping of the transgenic hairy roots

To confirm the *St16DOX* mutation of the two SGA-free lines T1/T2_#6 and T7/T8/T9_#4, we analyzed the sequence of PCR products of randomly selected clones by the Sanger method. All the PCR fragments showed some deletions in the *St16DOX* gene (Fig. 1-8). Among 19 PCR fragments from the T1/T2_#6 line (Fig. 1-8A), one fragment contained deletion mutation in both the T1 and T2 targeting sites, indicating that both T1-and T2-gRNAs were successfully expressed from the T1/T2-pMgP237 vector and caused the mutation of both the target sites. The other 18 fragments contained deletion mutations in only the T2 cleavage site, suggesting that T2-gRNA causes the targeted mutation of *St16DOX* more efficiently than T1-gRNA. Consequently, nine sequences showed the translational frameshift mutation in exon2, while the others exhibited 18-bp deletion at the 3′-end of exon2 with deletion of 10 bp or 54 bp at the 5′-end of intron2 (Fig. 1-8). Among 13 PCR fragments from the T7/T8/T9_#4 line (Fig. 1-8B), one fragment contained 1-bp insertion in the T7 target site and 1-bp deletion in the T9 target site. Four fragments contained 20-bp deletion in T7 and 1-bp insertion in T9, and the other eight fragments showed deletion of more than 500-bp region including the T7/T8/T9 target sites. The large segment deletions were probably caused by cleavage at both the T7 and T9 target sites, resulting in chromosomal fragment deletion that has been observed in the multiplex gene editing of rice, maize, and kiwifruit[35,42,43]. These results indicated that the gRNAs for the T7/T8/T9 target sites were successfully expressed and caused the mutation at the target sites. In the T7/T8/T9_#4 line, we could not detect the

mutation at the T8 target site because of the long-range deletion including the T8 target site. Taken together, these results indicated that the mutagenesis by the pMgP237 vectors caused the complete abolition of the functional 16DOX protein.

Discussion

Reducing the SGA content in potato is required for producers and consumers, since SGAs are harmful compounds against human. Silencing SGA biosynthetic genes such as *PGA1*, *PGA2*, *PGA3*, and *SSR2* by RNA-interference (RNAi) prevent accumulation of SGAs in potato tubers. However, not negligible amount of SGA was detected in all the silenced lines, because the gene expression by RNAi could not be blocked completely. Similarly, *SSR2*-disrupted potato plants constructed by TALEN contained a small amount of SGA[20]. The residual SGA might result from the presence of *sterol* side chain reductase 1 (*SSR1*) that shares 81% nucleotide identity with *SSR2*. Thus, the complete abolition of SGAs in potato has not been achieved. Recently, we reported that St16DOX catalyzes steroid 16α-hydroxylation and that *St16DOX*-silenced potato plants contain significantly lower levels of SGAs[24]. St16DOX showed the highest identity (61.6%) with PGSC0003DMP400034797/Sotub06g034370 in the potato genome, while the Sotub06g034370 protein did not catalyze steroid 16α-hydroxylation[24], suggesting that *St16DOX* is the single copy gene necessary for steroid 16α-hydroxylation in SGA biosynthesis. In accordance, the *St16DOX*-mutated hairy root lines in this study contained no detectable SGAs (Fig. 1-7A and B). In agreement with the metabolic profile, we detected the mutations of the *St16DOX* genome DNA sequence in all the sequenced PCR-fragments from the two lines T1/T2_#6 and T7/T8/T9_#4 (Fig. 1-6). All the observed mutagenesis caused St16DOX inactivation due to the inadequate translation of

the St16DOX protein. Thus, *St16DOX* is a suitable target gene for genome editing to generate SGA-free potato plants.

Previously, the CRISPR/Cas9 system has been applied in many crops[44]. Genome manipulation in polyploid heterozygous crops such as tetraploid potato and hexaploid bread is more challenging than in diploid crops because it includes the task of simultaneously targeting multiple alleles. To date, multiplexed CRISPR/Cas9 methods have been developed, for example the co-expression of multiple gRNAs by connecting multiple single gRNA expressing cassettes tandemly in one plasmid construct[45], and by direct co-microinjection of in vitro synthesized gRNAs and Cas9 mRNA in the embryo[46]. More recently, Xie et al.[35] developed the CRISPR/Cas9 multiplex gene editing system in which multiple gRNAs are produced as a single transcript followed by division into individual gRNAs by pre-tRNA specific ribonuclease processing mechanism. In this study, a new CRISPR/Cas9 vector pMgP237-2A-GFP, which is designed to allow expression of multiplex gRNAs by the pre-tRNA processing system, was used to facilitate mutation on all four target gene alleles in tetraploid potato. We examined to evaluate four kinds of the pMgP237 vectors with nine gRNAs that were targeted to the different sites in *St16DOX* gene. The two vectors, T1/T2-pMgP237 and T7/T8/T9-pMgP237, were found to contribute to the SGA abolition in two lines, T1/T2_#6 and T7/T8/T9_#4 (Fig. 1-5 and 1-7A). Our sequence analysis indicates that target sequences T2, T7, and T9 were highly effective gRNA, resulting in the abolition of *St16DOX* intact sequences in tetraploid potato (Fig. 1-8). Since the mutagenesis efficiency depends on gRNA target sequences, multiplex gRNA expression system can raise the possibility that the highly efficient gRNAs are adopted for mutagenesis of the target gene. Multiple mutagenesis at the different sites and chromosomal fragment deletion in the *St16DOX* gene were

observed in the two lines (Fig. 1-8), demonstrating that the pMgP237 vector successfully induced expression of multiple gRNAs. Thus, the complete disruption of *St16DOX* and the SGA abolition were achieved in tetraploid potato by using the pMgP237 vector. Therefore, this pMgP237 system can be applied to mutagenesis of multiple target genes and large chromosomal fragment deletion in various plant species.

One of key factors determining mutagenesis efficiency by a CRISPR/Cas9 system is the selection of gRNA sequences in a target gene, and a short-term method such as the hairy roots culture system is helpful to evaluate the efficacy of candidate gRNAs. Hairy roots have several attractive features such as hormone-independent growth, high genetic stability, mass-production of specialized metabolites and relatively fast growth rates. Ron et al.[47] previously demonstrated site-directed mutagenesis in tomato using CRISPR/Cas9 system with the hairy root transformation. In this study, we adopted the hairy root culture system of potato by *A. rhizogenes*-mediated transformation as delivery DNA method. The transgenic potato hairy roots were established by easy operation for two months after *A. rhizogenes* infection, and then we obtained several gRNAs that showed relatively high mutagenesis efficiency. Porter and Flores[48] reported that more than 450 plant species of many different genera and families are known to be susceptible to the infection by *A. rhizogenes*. Therefore, site-directed mutagenesis using the hairy root culture system can be applied in various plant species.

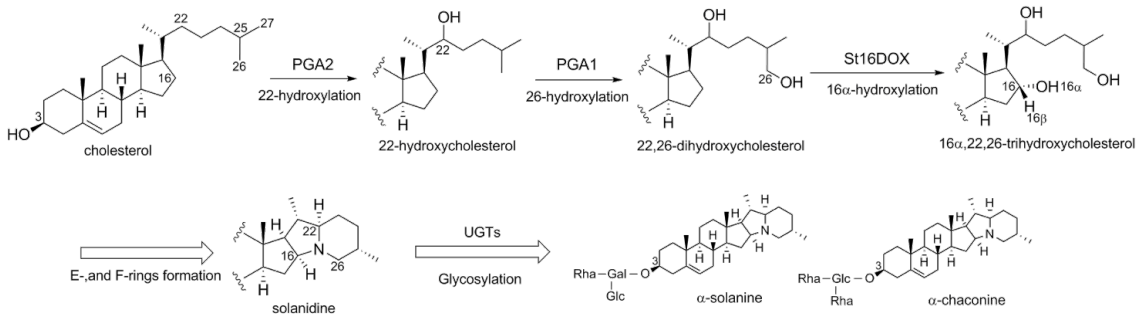
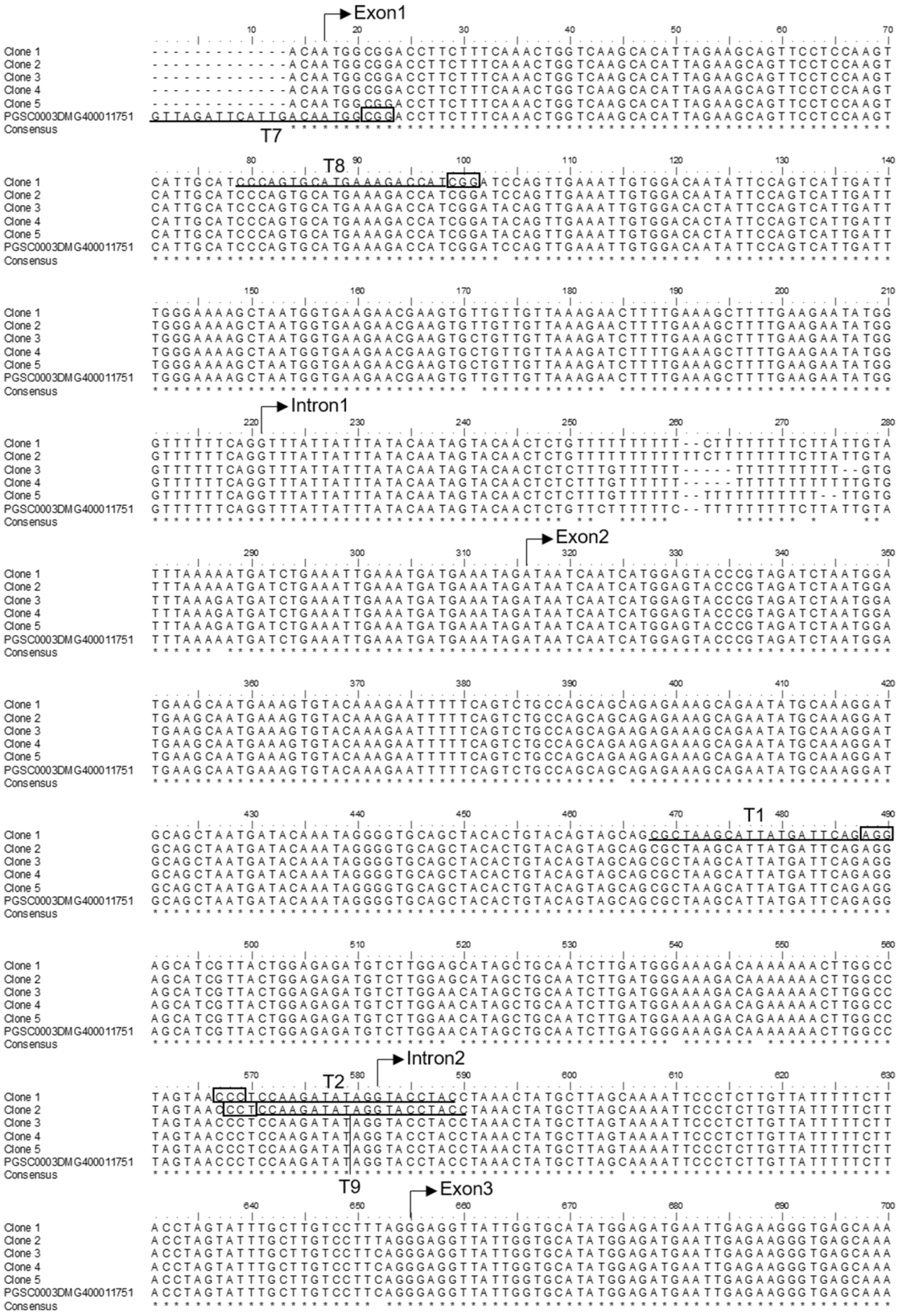


Fig. 1-1. Putative biosynthetic pathway for SGAs in potato. The solid arrows indicate the reaction steps reported by Umemoto et al.[21] and Nakayasu et al.[24]. White arrows represent multiple reaction stages.


Exon1
T7
T8
Intron1
Exon2
T1
Intron2
T2
T9
Exon3
Clone 1
Clone 2
Clone 3
Clone 4
Clone 5
PGSC0003DMG400011751
Consensus

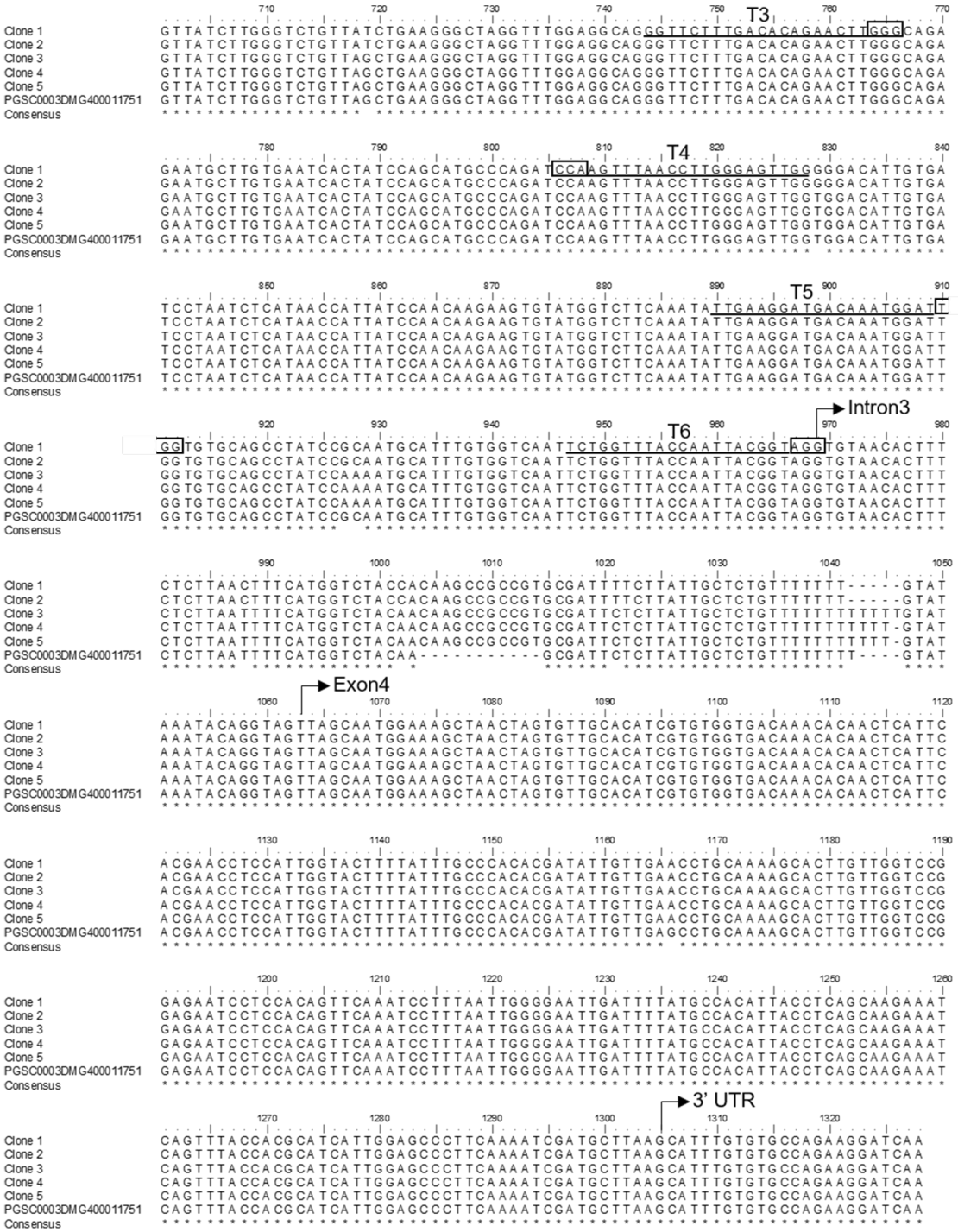


Fig. 1-2. Partial St16DOX genome DNA sequences in potato cv. Mayqueen determined by sequencing analysis. Underlines indicate gRNA target sequences. Boxes indicate PAM sequences (NGG). Arrows indicate start sites of exon, intron, and UTR.

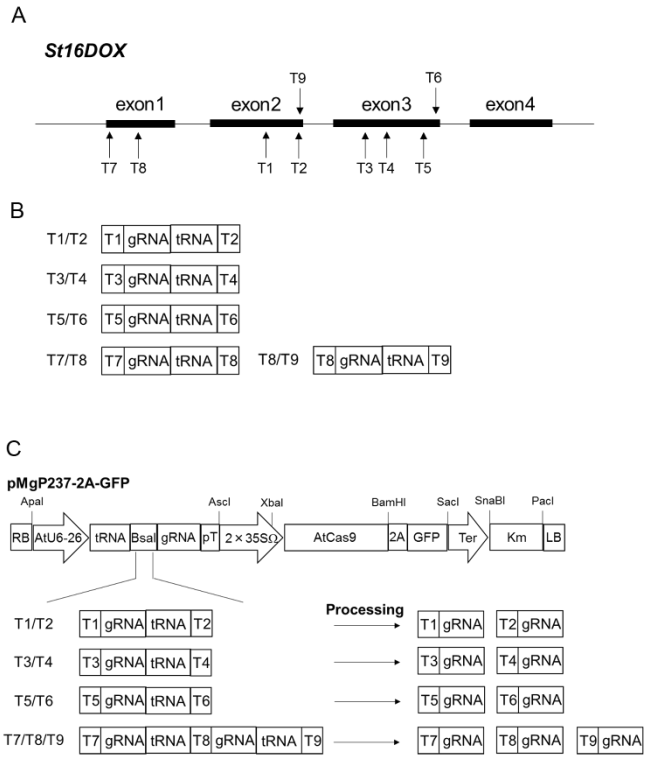


Fig. 1-3. Design of CRISPR/Cas9 construct for site-directed mutagenesis in the *St16DOX* gene in potato cv. Mayqueen. **(A)** The sites of nine candidate gRNA target sequences named T1-T9 in *St16DOX*. **(B)** Schematic representation of DNA fragments composed of gRNA scaffold and tRNA scaffold between two targets generated by PCR. **(C)** Schematic representation of T-DNA regions of the CRISPR/Cas9 binary vector designed pMgP237-2A-GFP constructed in this study. AtU6-26, *Arabidopsis thaliana* U6 snRNA-26 (U6-26) promoter; 2×35SΩ, 2 × *CaMV35S* promoter with the omega enhancer sequence; AtCas9, Arabidopsis-codon optimized *Streptococcus pyogenes* Cas9; 2 A, 2 A self-cleavage peptide from *Thosea asigna*; Km, the kanamycin resistant marker expression cassette; RB, right border of T-DNA; LB, left border of T-DNA. Each PCR fragment shown in (B) was ligated into pMgP237-2A-GFP.

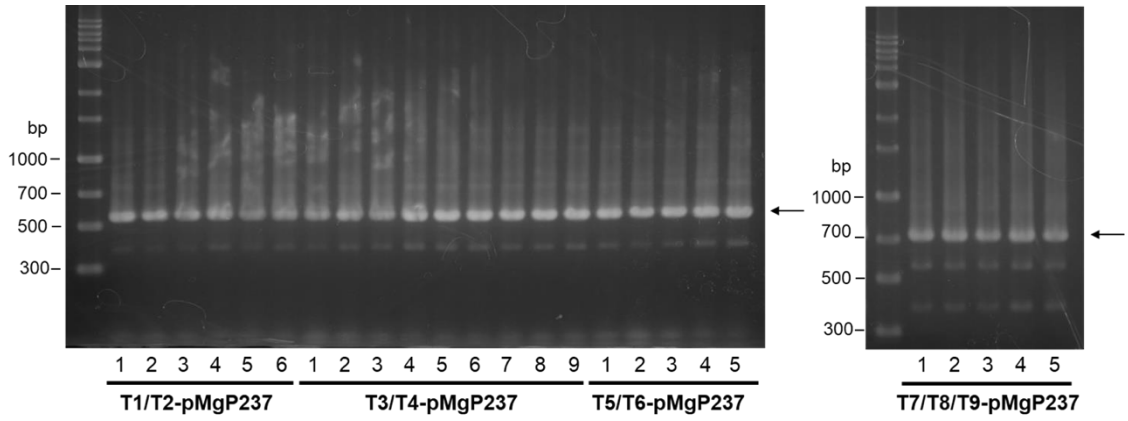


Fig. 1-4. Agarose gel electrophoresis of genomic PCR amplicons of a partial region of the T DNA region integrated into the genome in the generated hairy roots. Arrows indicate the PCR amplicons.

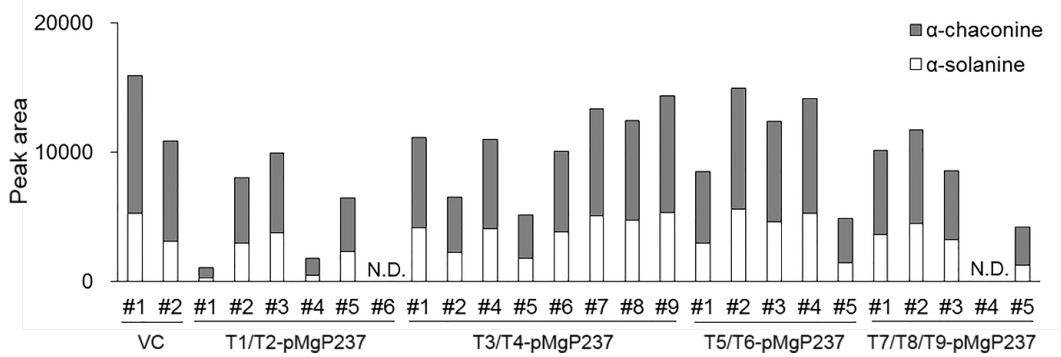


Fig. 1-5. Peak area calculated by LC-MS analysis of SGAs in the transgenic hairy roots. VC, vector control. N.D., Not detected.

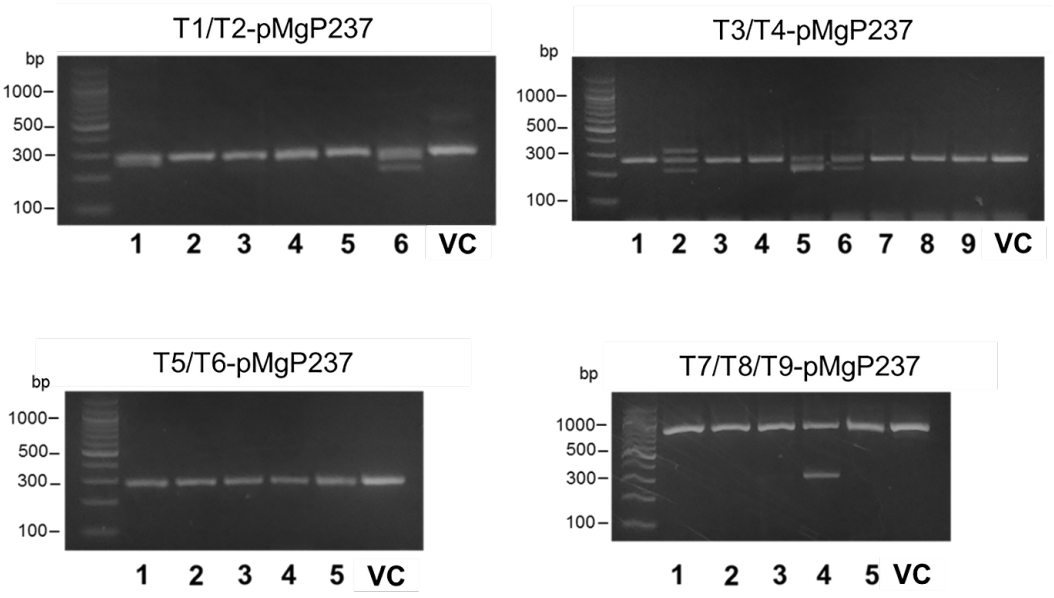


Fig. 1-6. Agarose gel electrophoresis of PCR amplicons of partial *St16DOX* genome DNA in the transgenic hairy roots. VC, vector control.

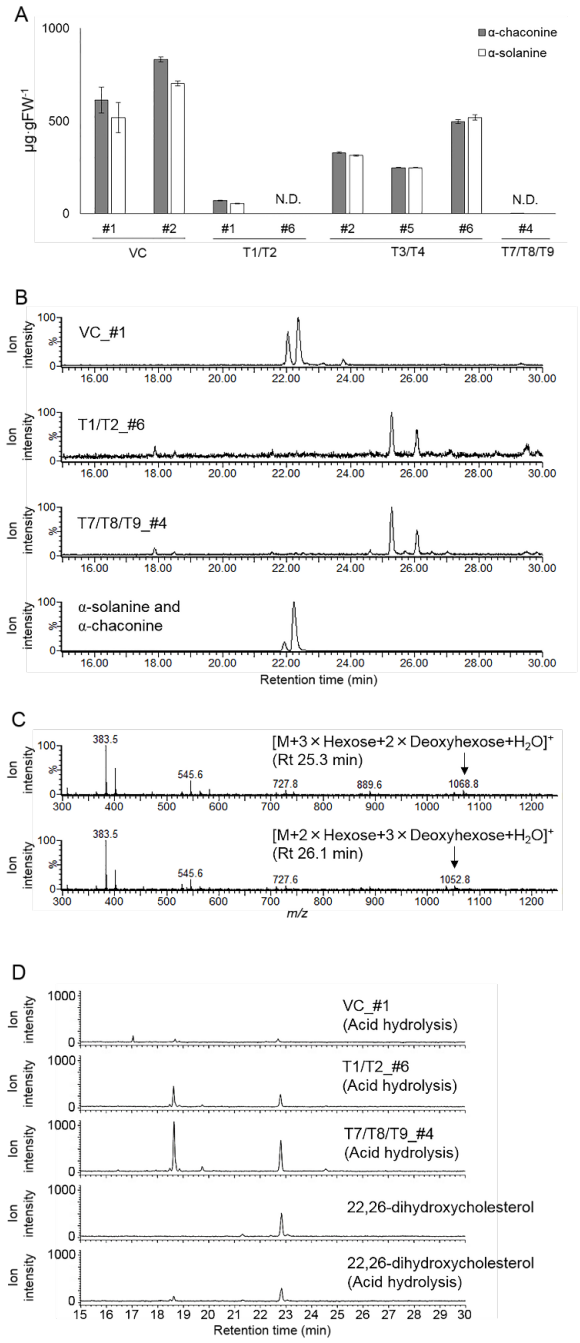


Fig. 1-7. Metabolic analysis of steroidal compounds in *St16DOX*-mutated hairy roots. **(A)** SGA contents in *St16DOX*-mutated hairy roots. **(B)** Total ion chromatogram obtained by LC-MS analysis of the accumulated compounds and the authentic compound. **(C)** Mass spectra of the peaks shown in (B) at retention times of 25.3 min and 26.1 min of T7/T8/T9_#4. **(D)** The extracted ion chromatogram targeting ion *m/z* 171 obtained by GC-MS analysis of the accumulated compounds and the authentic compound. Error bars indicate SD (n = 3). FW, Fresh weight; VC, vector control; N.D., Not detected.

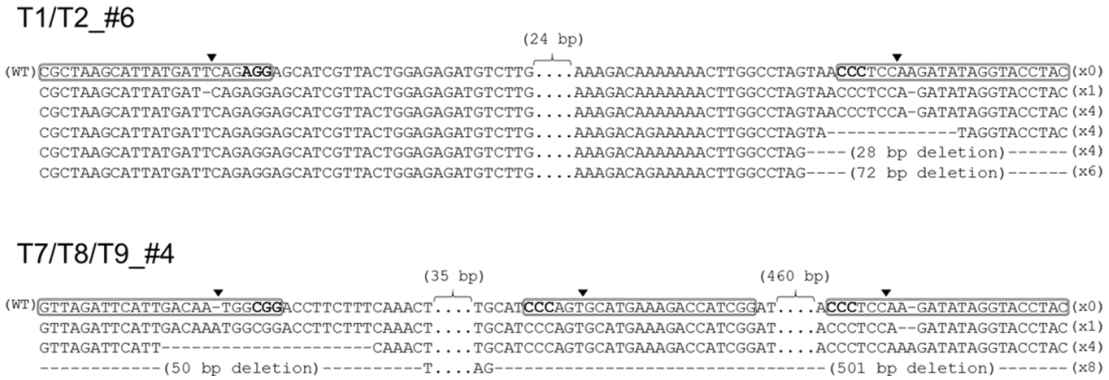


Fig. 1-8. Genotyping of sequences surrounding the *St16DOX* gene in the SGA-free transgenic hairy roots: T1/T2_#6 and T7/T8/T9_#4, compared to the wild-type (WT) sequence shown on top. Boxes indicate gRNA target sequences. Bold letters indicate PAM sequences. Black triangles indicate putative cleavage sites of the targets. The number of deleted and inserted nucleotides is shown on the right.

Table 1-1. The oligonucleotides that were used in the current study.

Primer No.	Sequences (5' to 3')
1	ACAATGGCGGACCTTCTTTCAAAC
2	TTGATCCTTCTGGCACACAAATGC
3	TTGGGTCTCGTGCAGCGCTAAGCATTATGATTCAGGTTTTAGAGCTAGAAATAGCA
4	TTGGGTCTCCAAACTCCAAGATATAGGTACCTACTGCACCAGCCGGGAATCGAA
5	TTGGGTCTCGTGCAGGTTCTTTGACACAGAACTTGTTTTAGAGCTAGAAATAGCA
6	TTGGGTCTCCAAACAGTTTAACCTTGGGAGTTGGCTGCACCAGCCGGGAATCGAA
7	TTGGGTCTCGTGCAGTTGAAGGATGACAAATGGATGTTTTAGAGCTAGAAATAGCA
8	TTGGGTCTCCAAACACCGTAATTGGTAAACCAGACTGCACCAGCCGGGAATCGAA
9	TTGGGTCTCGTGCAGTTAGATTCATTGACAATGGGTTTTAGAGCTAGAAATAGCA
10	TTGGGTCTCCTCATGCACTGGGCTGCACCAGCCGGGAATCGAA
11	TTGGGTCTCGATGAAAGACCATGTTTTAGAGCTAGAAATAGCA
12	TTGGGTCTCCAAACCCAAGATATAGGTACCTACCTGCACCAGCCGGGAATCGAA
13	GGCCCCTGGGAATCTGAAAG
14	GGAAGAAGAAATCGATCTGGAAAATTTTGC
15	AGCAGAATATGCAAAGGATGCAGC
16	CTCCATATGCACCAATAACCTCCCT
17	AGGGAGGTTATTGGTGCATATGGAG
18	GCACACCAATCCATTTGTCATCCTTC
19	AGATCCAAGTTTAACCTTGGGAGTTGG
20	ACTAGTTAGCTTTCCATTGCTAACTACCTG
21	CCCTCCATAGCTCAAATGAGATATCAGAC

Table 1-2. Nine candidate gRNA target sequences. Underlines indicate PAM sequences. Bold letters indicate G added to the 5'-end of targets.

Name	gRNA target sequences (5' to 3')
T1	**G**CGCTAAGCATTATGATTCAGAGG
T2	GTAGGTACCTATATCTTGGAGGG
T3	GGTTCTTTGACACAGAACTTGGG
T4	**G**CCAACTCCCAAGGTTAAACTTGG
T5	**G**TTGAAGGATGACAAATGGATTGG
T6	**G**TCTGGTTTACCAATTACGGTAGG
T7	GTTAGATTCATTGACAATGGCGG
T8	**G**CCCAGTGCATGAAAGACCATCGG
T9	GGTAGGTACCTATATCTTGGAGG

Chapter 2

Characterization of steroid 5α-reductase involved in α-tomatine biosynthesis in tomatoes

Introduction

As described in general introduction, tomato contains α-tomatine and dehydrotomatine as major SGAs[5] in its vegetative tissues. and dehydrotomatine are glycoside of tomatidine and dehydrotomatidine, respectively. α-Tomatine possesses a single bond between C5 and C6 (dihydro type), while dehydrotomatine has a $\Delta^{5,6}$ double bond (dehydro type), and they are biosynthesized via four steps of glycosylation from their respective aglycones, tomatidine and dehydrotomatidine (Figure2-1A). Tomatidine is biosynthesized from dehydrotomatidine by the four-step reaction process catalyzed by 3bhydroxysteroid dehydrogenase (3bHSD), 3-ketosteroid isomerase (3KSI), steroid 5a-reductase (S5aR), and 3-ketosteroid reductase (3KSR) as shown in (Figure2-1A). We recently identified that the short-chain alcohol dehydrogenase/reductase gene named *Sl3βHSD1* is involved in the conversion of dehydrotomatidine to tomatidine[10]. Sl3βHSD1 is a multifunctional enzyme involved in three of three of four reactions converting dehydrotomatidine to tomatidine[10], and however, it has not yet been identified as the gene involved in a 5α-reduction.

The 5α-reduction step is thought to be catalyzed by a steroid 5α-reductase (S5αR), which is involved in the NADPH-dependent reduction of the $\Delta^{4,5}$ double bond in steroids and steroid hormones in various organism. In plants, Arabidopsis *DET2* (*AtDET2*) has been identified as a biosynthetic gene of plant hormone brassinosteroids (Fig. 2-1B), and it encodes an S5αR protein that shares significant sequence identity with mammalian S5αRs[49]. The Arabidopsis *det2* mutant shows a small dark-green dwarf

phenotype, due to of brassinosteroid deficiency caused by lack of the conversion reaction of campest-4-en-3-one to campest-3-one[49]. Tomato *LeDET2*, which is an ortholog of AtDET2 has been characterized as a functional steroid 5α-reductase[50]. Moreover, the presence of two isozymes was suggested by a assay using crude enzyme prepared with tomato plants[50]. We recently isolated and characterized steroid 5α-reductase named SlS5αR1, and co-incubation of SlS5αR1 with Sl3βHSD1 and dehydrotomatidine *in vitro* demonstrated that the four reaction steps converting dehydrotomatidine to tomatidine were reconstituted in the in vitro assay[10]. However, whether *SlS5αR1* contributes to α-tomatine biosynthesis in vivo has remained unknown.

In the present study, we isolated a second *DET2* homolog from tomato, designated as *SlS5αR2*. The *SlS5αR2* gene was co expressed with known SGA biosynthetic genes in tomato. In vitro assay of recombinant SlS5αR2 using progesterone as a substrate revealed that *SlS5αR2* encodes a functional steroid 5α-reductase. Additionally, simultaneous incubation of SlS5αR2 with Sl3βHSD1 resulted in the successive conversion of dehydrotomatidine to tomatidine *in vitro*, as observed in the case of recombinant SlS5αR1. Furthermore, CRISPR/Cas9-mediated knockout of either *SlS5αR2* or *SlS5αR1* in tomato hairy roots was conducted. While disruption of SlS5αR1 did not affect the endogenous SGA levels, SlS5αR2-knockout tomato hairy roots showed drastic reduction in the α-tomatine level and significant accumulation of dehydrotomatine. Taken together, we conclude that *SlS5αR2*, but not *SlS5αR1*, is essential for the 5α-reduction step of α-tomatine biosynthesis in tomatoes.

Materials and methods

Plant materials

The tomato plant used in this study was *S. lycopersicum* cv Micro-Tom (TOMJPF00001), obtained from the NBRP (MEXT, Japan). The seeds were germinated and cultivated under a 16/8-h light/dark photoperiod in plant thermostatic chambers at 25ºC until the plants produced ripening red fruits.

Chemicals

Progesterone were purchased from Tokyo Chemical Industry Co., Ltd. (Tokyo, Japan), and 5α-pregnane-3,20-dione was purchased from Santa Cruz Biotechnology (Satnta Cruz, CA, USA). Dehydrotomatidine was isolated in our lab from tomatidine purchased from Chromadex (Irvine, CA, USA) using high-performance liquid chromatography (HPLC). α-Tomatine was purchased from Sigma-Aldrich.

RNA extraction and reverse transcription

Total RNAs were extracted using the RNAeasy Plant Mini Kit (QIAGEN, Hilden, Germany) and the RNase-Free DNase Set (QIAGEN) according to the manufacturer's instructions. The extracted total RNAs were used to synthesize the first strand of cDNA using the Transcriptor First Strand cDNA Synthesis Kit (TOYOBO, Osaka, Japan) for real-time quantitative polymerase chain reaction (PCR).

Real-time quantitative PCR analysis

Real-time quantitative PCR (qPCR) was performed with the LightCycler®Nano (Roche, Basel, Switzerland) using GeneAce SYBR® qPCR Mix α (Nippon Gene), with the

following primers 1 and 2 for *SlS5αR1*, 3 and 4 for *SlS5αR2*, 5 and 6 for *ubiquitin* (Table 2-1), using as template cDNAs from the various tomato tissues; flower, green fruit (G fruit), yellow fruit (Y fruit), orange fruit (O fruit), and red fruit (R fruit) as templates. Cycling was conducted at 95℃ for 10 min, 45 cycles at 95℃ for 30 s, and 60℃ for 1 min for amplification, followed by holding at 95℃ for 30 s and ramping up from 60℃ to 95℃ at 0.1℃ s^{-1} to perform a melting curve analysis. Each assay was repeated three times. The expression levels were normalized against the values obtained for the *ubiquitin* gene, which was used as an internal reference in tomato. Data acquisition and analysis were carried out using LightCycler®Nano software (Roche).

Cloning of SlS5αR1, SlS5αR1 and Sl3βHSD1

The DNA fragments containing the ORF of *Solyc09g013070* (*SlS5αR1*) and *Solyc10g086500* (*SlS5αR2*) were isolated using the cDNA template from the tomato leaves with primer sets 7 and 8 for *SlS5αR1*, and 9 and 10 for *SlS5αR2* (Table 2-1). The PCR product was purified using Wizard® SV Gel and the PCR Clean-Up System (Promega, WI, USA) and cloned into the pMD19 T-vector (TaKaRa, Shiga, Japan). The sequence of the cloned DNA was confirmed using the ABI 3130 genetic analyzer (Applied Biosystems, CA, USA) and analyzed using BioEdit, a biological sequence alignment editor (http://www.mbio.ncsu.edu). Similarly, the open reading frame (ORF) of *Sl3βHSD1* was isolated using a clone of the LEFL1039AB03 (identical to *Solyc01g073640* in the tomato ITAG 2.4) in the Kazusa Full-length Tomato cDNA Database (http://www.pgb.kazusa.or.jp/kaftom/clone.html), purchased from Kazusa DNA Research Institute. The ORF of *Sl3βHSD1* was amplified using PCR with the clone as a template and primer sets 11 and 12 (Table 2-1).

Heterologous expression of the recombinant SlS5αR1, SlS5αR2 and Sl3βHSD1 proteins

The DNA fragments containing the ORF of the *SlS5αR1* and *SlS5αR2* gene were ligated into a pFastBac1 vector (Invitrogen, Carlsbad, CA), and were then introduced into *Escherichia coli* DH10Bac (Life Technologies) to generate the corresponding recombinant bacmid DNAs. Preparation of recombinant baculovirus DNAs that contained *SlS5αR1* and *SlS5αR2* ORFs and transfection of *Spodoptera frugiperda 9* insect cells were carried out according to the manufacturer's instructions (Life Technologies). Heterologous expression of SlS5αR1 and SlS5αR2 proteins in insect cells was conducted as described by Ohnishi et al[51]. Similarly, the DNA fragments containing the *Sl3βHSD1* gene ORF were ligated into pCold ProS2 vector (TaKaRa), and recombinant Sl3βHSD1 protein was prepared using bacterial expression system in *E.coli* strain BL21 (DE3) as described by Lee et al.[10]

In vitro enzyme activity assay

To prepare tomatid-4-en-3-one, we conducted in vitro enzyme assay of recombinant Sl3βHSD1 protein, and the assay was performed using 100 μL of reaction mixture comprising 100 mM bis-tris-HCl (pH 7.2), 2.5 mM NAD^+ as a coenzyme, 50 μM dehydrotomatidine as a substrate, and the purified Sl3βHSD1. The enzymatically synthesized tomatid-4-en-3-one was further metabolized by recombinant SlS5Rα1 to produce tomatid-3-one, and the reaction was carried out as described below.

Microsomal fractions of the insect cells expressing SlS5αR1 or SlS5αR2 were obtained from 100 mL of cultured cell suspension. Infected cells were washed with

phosphate-buffered saline and suspended in buffer A, consisting of 20 mM potassium phosphate (pH 7.5), 20% (v/v) glycerol, 1 mM ethylenediaminetetraacetic acid, and 1 mM dithiothreitol. The cells were sonicated and cell debris was removed by centrifugation at 10,000g for 15 min, and the resulting pellets were homogenized with buffer A to provide the microsomal fractions. The assay of SlS5αR1 and SlS5αR2 for steroid C5 reduction were performed using 100 μL reaction mixture comprising 100 mM potassium phosphate (pH 7.2), 1 mM NADPH or NADH as a coenzyme, the enzymatically synthesized tomatido-4en-3-one, and the microsomal fraction containing SlS5αR1 or SlS5αR2. The conversion of dehydrotomatidine to tomatidine was reconstituted in an in vitro assay by mixing the purified Sl3βHSD1 protein with SlS5αR1 or SlS5αR2-containing microsomes. The reaction mixture consisted of 100 mM bis-tris-HCl buffer (pH 7.2), purified Sl3βHSD1 protein, microsomes containing SlS5αR1 or SlS5αR2, 2.5 mM NAD^+, 1 mM NADPH, and 25 μM dehydrotomatidine. All the reactions were performed at 30°C for 2 h and were terminated by the addition of 100 μL ethyl acetate. The reaction products were extracted three times with an equal volume of ethyl acetate and the organic phase was collected and evaporated. The residues were analyzed using gas chromatography–mass spectroscopy (GC-MS) and/or liquid chromatography–mass spectroscopy (LC-MS).

GC-MS analysis of the reaction product was conducted as described previously[10]. Briefly, the residue was trimethylsilylated, and then subjected to a GC-MS-QP2010 Ultra (Shimadzu, Kyoto, Japan) with a DB-1MS (30 m × 0.25 mm, 0.25 μm film thickness; J&W Scientific, CA, USA) capillary column. An MS scan mode with a mass range of *m/z* 50-700 were used. Progesterone and 5α-pregnane-3,20-dione were monitored at a selected ion monitoring (SIM) mode chromatogram at *m/z* 229 and 231, respectively.

Tomatid-4-en-3-one and tomatid-3-one were analyzed via SIM mode at *m/z* 125. For LC-MS analysis, the residue was dissolved in 200 μL of methanol. LC-MS analysis of each reaction product was carried out as described above Lee et al[10], using an ACQUITY UPLC H-Class System (Waters, MA, USA) with an Acquity UPLC HSS-T3 column (1.8 μm, 2.1 × 100 mm^2, Waters) at 40°C and an SQ Detector 2 (Waters). Data acquisition and analysis were performed using MassLynx 4.1 software (Waters). Reaction products were analyzed by SIM-mode chromatography at *m/z* 412, 414, and 416.

Construction of CRISPR/Cas9 vectors

The knockout tomato plants for each of *SlS5αR1* and *SlS5αR2* were generated by targeted genome editing with the CRISPR/Cas9 system. We used the CRISPR/Cas9 binary vector pMgP237-2A-GFP to express multiplex gRNAs[52,53]. To design a gRNA target with low off-target effect in *SlS5αR1* and *SlS5αR2*, we conducted *in silico* analyses using the Web tool Design sgRNAs for CRISPRko (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) and Cas-OT software[37]. Then we selected target sequences named SlS5R1_492-511 and SlS5R1_552-571 in the coding region of *SlS5αR1*, and SlS5R2_449-468 and SlS5R2_511-530 in the coding region of *SlS5αR2*, respectively (Fig. 2-2). To enhance the efficiency of gRNA transcription from U6 promoter, one G was added to the 5'- end of SlS5R1_492-511, SlS5R1_552-571 and SlS5R2_511-530. Two DNA fragments composed of the gRNA scaffold and tRNA scaffold between two target sequences, SlS5R1_492-511/ SlS5R1_552-571 and SlS5R2_449-468/SlS5R2_511-530, were generated by PCR using pMD-gtRNA containing gRNA and tRNA scaffolds as a template and primer sets containing restriction enzyme BsaI sites (13 and 14 for SlS5R1_492-511/ SlS5R1_552-

571, 15 and 16 for SlS5R2_449-468/SlS5R2_511-530) (Table 2-1). The units containing two gRNAs–tRNAs were then independently inserted into the BsaI site of pMgP237-2A-GFP using Golden Gate Cloning methods to construct the CRIPR/Cas9 vectors, labeled pMgP237-SlS5αR1ko and pMgP237-SlS5αR2ko, respectively.

Generation of SlS5RαR1- and SlS5RαR2-disrupted transgenic tomato hairy roots

The vectors pMgP237-SlS5αR1ko and pMgP237-SlS5αR2ko were independently introduced into *Agrobacterium rhizogenes* ATC15834 by electroporation. Transgenic tomato hairy roots were generated as previously described by Thagun et al.[39] Genome DNA was extracted from each established hairy root and transformants were selected by genomic PCR to amplify a partial fragment of the T-DNA region integrated into the genome using primers 17 and 18 (Table 2-2). To analyze the mutations or large deletions in the transgenic hairy roots, the region including the target sites of gRNAs was amplified by PCR using primers 19 and 20 for pMgP237-SlS5αR1ko transformants and primers 21 and 22 for pMgP237-SlS5αR2ko transformants (Table 2-2). The PCR fragments obtained from each transformant were then subjected to agarose-gel electrophoresis or microchip electrophoresis using MultiNA (Shimadzu). Furthermore, each PCR fragment from four independent transgenic hairy root lines (pMgP237-SlS5αR1ko_#1 and #6, pMgP237-SlS5αR2ko_#7 and #12) was cleaned using the Wizard® SV Gel and PCR Clean-Up System (Promega, Japan) according to the manufacturer's instructions, and cloned into pCR™4Blunt-TOPO® (Invitrogen). Sanger sequencing of each of the cloned DNAs was performed using a sequencing service (Eurofins Genomics).

Quantification of SGAs in tomato hairy roots

SGAs were extracted from 100 mg fresh weight of harvested transgenic lines, and each sample was prepared and analyzed by LC-MS. The LC-MS analysis was performed using the LC-MS/MS system; water with 0.1% (v/v) formic acid (A) and acetonitrile (B) were used as mobile phases, with an elution gradient of 10%–55% B from 0 to 30 min and 55%–75% B from 30 to 35 min (linear gradient). MS scan mode was used with a mass range of *m/z* 400–1200. The quantities of α-tomatine and dehydrotomatine were calculated from the peak area using an α-tomatine calibration curve.

Phylogenetic analysis

Amino acid sequences of AtDET2 homologs were retrieved from the Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) and Sol Genomics networks (https://solgenomics.net). A phylogenetic tree was constructed using the MEGA7 program and the maximum likelihood method with the following parameters: Poisson correction, pairwise deletion and bootstrap (1,000 replications; random seed).

Results

Selection of candidate S5αR genes involved in α-tomatine biosynthesis

To survey the candidate gene encoding functional S5αR, we performed BLASTX analysis against tomato transcripts database from Sol Genomics (http://solgenomics.nrt.) used AtDET2 as a query. This analysis allowed us to identify two distinct transcripts encoded by *Solyc09g013070* and *Solyc10g086500*. *Solyc09g013070* was identical to *SlS5αR1*, which has been already characterized as functional steroid 5α-reductase[10] in our previous work, and *Solyc10g086500* was identical to *LeDET2*, which was previously

characterized to encode a functional steroid 5α-reductase[50], and here we designated it as *SlS5αR2*. SlS5αR1 and SlS5αR2 show 63% and 57% amino acid identity to AtDET2, respectively, and share 67% amino acid identity to each other. Both *SlS5αR1* and *SlS5αR2* contain six (Arg-150/148, Pro-185/183, Gly-187/185, Asn-197/195, Gly-200/198 and Arg-251/249) of seven conserved amino acid residues that are part of a typical cofactor binding domain in mammalian 5α-reductase[54] (Fig. 2-3). Both SlS5αR1 and SlS5αR2 contain six potential transmembrane-spanning domains based on the TMHMM program (http://www.cbs.dtu.dk/services/TMHMM), as well as AtDET2 predicted by the same program.

Expression analysis of SlS5αR1 and SlS5αR2 in various tissues of tomato

We analyzed expression of *SlS5αR1* and *SlS5αR2* in the RNA-seq data of the Tomato Functional Genomics Database (http://ted.bti.cornell.edu). SlS5αR2 showed similar expression pattern to previously identified SGA biosynthetic genes in terms of higher expression in immature fruits and decreasing expression levels during fruits ripening, whereas the transcript levels of *SlS5αR1* remained high even in red ripe fruits (Table 2-2). To confirm we analyzed the expression of *SlS5αR1* and *SlS5αR2* by real-time PCR using the reverse-transcription products of total RNAs extracted from various tissues of *S. lycopersicum* cv. Micro-Tom. *SlS5αR2* was highly expressed in flowers, where a large amount of α-tomatine is accumulated [55], while the expression level of *SlS5αR2* was low in red mature fruits (Fig. 2-4), in which α-tomatine is further metabolized to esculeoside A and the α-tomatine level is very low [11,12,17]. On the other hand, *SlS5αR1* was abundantly expressed in red mature fruits (Fig.2-4). These results are in good agreement with the observation in RNA-seq data analysis.

In vitro functional analysis of the C5α-reduction activity of SlS5αR2 on tomatid-4-en-3-one to tomatid-3-one

In our recent studies, the recombinant SlS5αR1 protein was expressed with a baculovirus-mediated expression system in *Spodoptera frugiperda* 9 insect cells and confirmed to catalyze C5α-reduction of tomatid-4-en-3-one to yield tomatid-3-one in vitro[10]. Previously, it was reported that recombinant SlS5αR2 expressed in mammalian cells (COS-7) was active on typical substrates of mammalian S5αRs[50] (progesterone, testosterone and androstenedione). In this study, to compare the catalytic activities of SlS5αR1 and SlS5αR2, we prepared recombinant SlS5αR2 with a baculovirus-mediated expression system and C5-reduction activities were analyzed. The functional expression of SlS5αR2 was confirmed by *in vitro* enzyme assays using progesterone as a substrate. GC-MS analysis of the reaction products detected a peak with a retention time of 9.2 min and a major mass fragment ion at *m/z* 123. This product was identical to the authentic compound 5α-pregnane-3,20-dione in terms of both the retention time the mass spectrum (Fig. 2-5). Although SlS5αR2 catalyze the C5 reduction with either NADPH or NADH as a coenzyme, the reduction activity in the presence of NADPH was much higher than that in the presence of NADH. Thus, the functional expression of SlS5αR2 in the baculovirus–insect cell system was confirmed.

Next, we evaluated the C5α-reduction activities on tomatid-4-en-3-one. Because authentic compounds of tomatid-4-en-3-one and tomatid-3-one are not commercially available, we obtained them by enzymatic synthesis using Sl3βHSD1 and SlS5αR1[10]. Tomatid-4-en-3-one was synthesized from dehydrotomatidine by enzymatic conversion with the recombinant Sl3βHSD1 using NAD^+ as a coenzyme, and tomatid-3-one was synthesized from tomatid-4-en-3-one by enzymatic conversion with SlS5αR1 using

NADPH as a coenzyme. Then, the recombinant SlS5αR2 was assayed with the enzymatically synthesized tomatid-4-en-3-one in the presence of NADPH, and the reaction products were analyzed using GC-MS (Fig. 2-6A). SlS5αR2 metabolized tomatid-4-en-3-one to a product with a retention time of 18.9 min, and this product was identical to tomatid-3-one, obtained by incubating tomatid-4-en-3-one with SlS5αR1 (Fig. 2-6A). Thus, SlS5αR2 could catalyze the reduction of tomatid-4-en-3-one at C5α to yield tomatid-3-one.

Next, we conducted a co-incubation assay of SlS5αR2 with Sl3βHSD1 using dehydrotomatidine as a substrate, and the reaction products were analyzed by LC-MS (Figure 5B). When incubating dehydrotomatidine with Sl3βHSD1, tomatid-4-en-3-one was detected at a retention time of 19.6 min and with a parent mass at *m/z* 412.6 (Fig. 2-6B). Simultaneous incubation of dehydrotomatidine with SlS5αR2 and Sl3βHSD1 resulted in the formation of two major compounds with the retention times of 20.4 and 22.4 min, respectively (Figure 5B). The product with a retention time of 22.4 min gave a parent mass at *m/z* 414.6, which corresponded to [tomatid-3-one + H^+]$^+$ (Fig. 2-6B). The product at a retention time of 20.4 min, which had a parent mass at *m/z* 416.6, was identical to the authentic tomatidine (Fig. 2-6B). These results indicate that, during co-incubation with SlS5αR2 and Sl3βHSD1, dehydrotomatidine was at first converted to tomatid-4-en-3-one by Sl3βHSD1, and SlS5αR2 subsequently catalyzed the reduction of tomatid-4-en-3-one at C5α to form tomatid-3-one, and Sl3βHSD1 finally catalyzed the reduction of tomatid-3-one at C3 to form tomatidine. Similar result was obtained in case of co-incubation of dehydrotomatidine with SlS5αR1 and Sl3βHSD1 (Fig. 2-6B), and this result was consistent with our previous work[10]. Taken together, either SlS5αR2 or

SlS5αR1, in combination with Sl3βHSD1 and cofactors, could reconstitute the conversion of dehydrotomatidine to tomatidine in an *in vitro* assay.

CRISPR/Cas9-mediated genome editing of SlS5αR1 or SlS5αR2 in tomato hairy roots.

To clarify the contribution of *SlS5αR1* and *SlS5αR2* in α-tomatine biosynthesis *in vivo*, *SlS5αR1* and *SlS5αR2* were each independently disrupted using CRISPR/Cas9-mediated genome editing in tomato hairy roots. In this study, we used a CRISPR/Cas9 binary vector designated as pMgP237-2A-GFP that permits expression of multiplex gRNAs[52,53]. gRNAs specific to *SlS5αR1* and to *SlS5αR2* were selected by *in silico* analysis using Design sgRNAs for CRISPRko and Cas-OT software [37]. Then, the selected target sequences were inserted into a pMgP237-2A-GFP vector, and we constructed a targeting vector targeting *SlS5αR1* and one targeting *SlS5αR2*, designated as pMgP237-SlS5αR1ko and pMgP237-SlS5αR2ko, respectively (Fig 2-2).

Among the pMgP237-SlS5αR1ko transgenic hairy roots, a formation of a heteroduplex, which indicates the mutation at the target site, was detected as shifted peaks in #1 and #6 (Fig. 2-7). Among the pMgP237-SlS5αR2ko transgenic hairy roots, deletion of a large segment or a formation of a heteroduplex was detected in #5, #7, #8, #10, #11, and #12 (Fig. 2-7). To confirm the mutations at the target sites, four lines for each of pMgP237-SlS5αR1ko (lines #1 and #6) and pMgP237-SlS5αR2ko (line #7 and #12) were selected, and sequencing analyses were performed. The lines #1 and #6 for pMgP237-SlS5αR1ko had no intact *SlS5αR1* sequences, and all the sequences showed the translational frameshift mutations (Fig. 2-8A). In the lines #7 and #12 for pMgP237-SlS5αR2ko, all the detected sequences showed some deletions in the *SlS5αR2* gene (Fig. 2-8A).

Endogenous SGAs of these selected lines were extracted and analyzed by LC-MS. The level of α-tomatine in the pMgP237-SlS5αR1ko_#1 and #6 was comparable to that in the vector control, and the amounts of dehydrotomatine were not affected (Fig. 2-8B, C, D). On the other hand, pMgP237-SlS5αR2ko_#7 and #12 showed significantly reduced α-tomatine level and a corresponding increased dehydrotomatine compared with that in the vector control (Fig. 2-8B, C, D). These results indicate that *SlS5αR2* is involved in the conversion of dehydrotomatidine to tomatidine in tomato, while *SlS5αR1* is not responsible for biosynthesis of SGAs.

Discussion

In the present study, we found that SlS5αR1 and SlS5αR2 catalyze the conversion of tomatid-4-en-3-one to tomatid-3-one *in vitro* (Fig. 2-6). In the *SlS5αR2*-knockout tomato hairy roots, α-tomatine was dramatically decreased and corresponding amounts of dehydrotomatine were accumulated, while the disruption of *SlS5αR1* did not affect the α-tomatine and dehydrotomatine levels (Fig. 2-8). These results clearly demonstrate that *SlS5αR2*, but not *SlS5αR1*, is involved in the C5α reduction step of the metabolic conversion of dehydrotomatidine to tomatidine in α-tomatine biosynthesis.

Arabidopsis DET2 (*AtDET2*) is the first gene identified in plants coding functional steroid 5α-reductase, and the *Arabidopsis* det2-1 mutant shows a dwarf and de-etiolated phenotype because of a deficiency of bioactive brassinosteroids[49]. In plants, the reduction of the steroid $\Delta^{4,5}$ double bond is an important reaction in the biosynthesis of BRs since all known bioactive BRs lack $\Delta^{4,5}$ double bond[50]. Steroid 5α-reductase is likely to be conserved in plant kingdom, and we found that some plant species have multiple S5αR genes in their genome. We also found that all Solanaceae species analyzed

here have two or more *S5αR* genes in the genome, and phylogenetic analysis demonstrated that the S5αR proteins of Solanaceous plants are clearly divided into two distinguishable clades (Fig. 2-9), suggesting that duplication of the ancestral *S5αR* gene has occurred in ancestral Solanaceous plants. SlS5αR1 shows higher homology to AtDET2, suggesting that SlS5αR1 would keep to function only in the biosynthesis of phytosterols and brassinosteroids. Then, the duplication gives rise to SlS5αR2, which is specialized in α-tomatine biosynthesis in tomato. The duplication and following neofunctionalization have been reported for several enzymes involved in phytosterols, BRs and cholesterol biosynthesis[20,56,57]. For instance, potato and tomato have two DWF1 homologs named SSR1 and SSR2[20]. SSR1 corresponds to DWF1 and has $\Delta^{24(28)}$ reduction activity required for phytosterol biosynthesis, while SSR2 catalyzes $\Delta^{24(25)}$ reduction of cycloartenol and desmosterol for cholesterol biosynthesis. Our in vitro assay revealed that both SlS5Rα1 and SlS5αR2 catalyze the conversion of 5α-reduction of tomatid-4-en-3-one to yield tomatid-3-one (Fig. 6A). Furthermore, in combination with Sl3βHSD1, each of SlS5αR1 and SlS5αR2 could complete the conversion of dehydrotomatidine to tomatidine in the in vitro assays (Fig. 6B). These observations suggest that the different physiological roles for *SlS5αR1* and *SlS5αR2* reflect factors other than catalytic activity.

Recently, Thagun et al.[39] reported that *Jasmonate-Responsive ERF 4* (*JRF4*) comprehensively regulates α-tomatine biosynthetic genes, including those involved in the upstream mevalonate pathway, in tomatoes. Quantitative RT-PCR analysis in various tissues of tomatoes showed different expression patterns between *SlS5αR1* and *SlS5αR2*, and the expression pattern of *SlS5αR2* is in good agreement with those of known SGA biosynthetic genes. Therefore, our results indicate that acquisition of the transcriptional regulation, which cooperates with other SGA biosynthetic genes, contributes to

functional specialization of SlS5αR2 for SGA biosynthesis. However, our quantitative RT-PCR analyses and RNA-seq data of the Tomato Functional Genomics Database (http://ted.bti.cornell.edu) demonstrated that *SlS5αR1* is also expressed in certain amounts in SGA containing tissues, suggesting that there are mechanisms, in addition to transcriptional control, to explain the functional differentiation of SlS5αR1 and SlS5αR2 *in vivo*. For example, post-translational regulation such as distinct intracellular localization and/or protein-protein interaction in the metabolon can be one of the mechanisms, as discussed below.

SlS5αR2-knockout tomato hairy roots contained very little α-tomatine, and the amount of dehydrotomatine increased to a level comparable to the level of α-tomatine accumulated in control tomato hairy roots (Fig. 8). Dehydrotomatine is biosynthesized via glycosylation from dehydrotomatidine, which is not a direct substrate of SlS5αR2. Dehydrotomatidine is metabolized via oxidation and isomerization by Sl3βHSD1 to yield tomatid-4-en-3-one, which is a direct substrate of SlS5Rα2 (Fig. 1). Considering that isomerase reaction of mammalian 3βHSD has been reported to be irreversible[58], high accumulation of dehydrotomatine suggests that Sl3βHSD1 does not function well in the *SlS5αR2*-knockout tomato hairy roots. This may be the reason why dehydrotomatine but not tomatid-4-en-3-one accumulated in the *SlS5αR2*-knockout tomato hairy roots. It has been reported that protein–protein interaction that allows the formation of protein complexes of sequential metabolic enzymes, termed metabolon, enhances direct channeling of substrates between the biosynthetic enzymes, providing increased control over metabolic pathway fluxes[59]. So far, to the best of our knowledge, there has been no report of metabolon formation in SGA biosynthesis, but the physical protein–protein interaction of SlS5αR2 with Sl3βHSD1, or other SGA biosynthetic enzymes, may

influence the functional differentiation of SlS5αR1 and SlS5αR2. In order to address this hypothesis, we need to examine in detail the subcellular localization of SlS5αR2 and other SGA biosynthetic enzymes, as well as protein–protein interactions among SGA biosynthetic enzymes.

In conclusion, we showed here that SlS5αR2 is a key enzyme in the production of tomatidine from dehydrotomatidine in tomatoes, thereby contributing to structural diversity of SGAs in the Solanaceae family. SGA is well known as a toxic and antinutritional substance in a Solanaceae crop, but SGAs are considered to play protective roles against plant pathogens and predators for the plants. Through this study, it may be possible to clarify the significance of Solanaceous plants, producing SGA with various structure

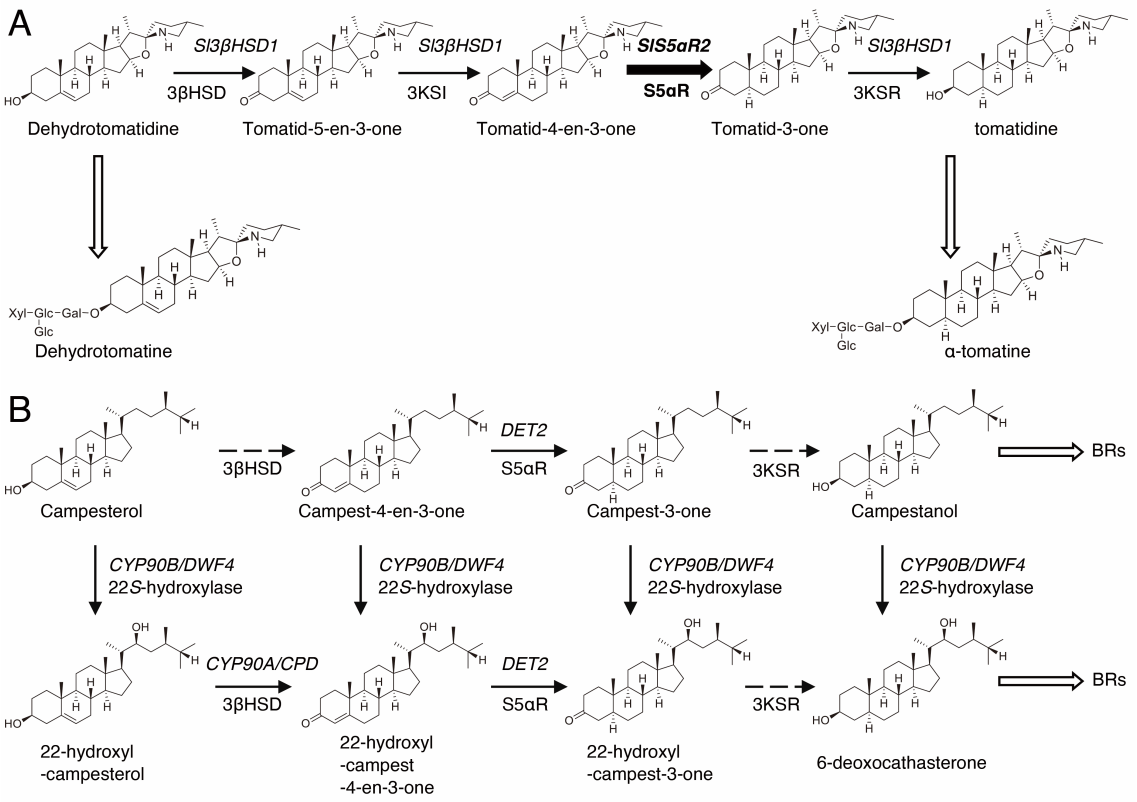


Fig. 2-1. The putative biosynthetic pathways for (A) α-tomatine in tomatoes and (B) brassinosteroids (BRs) in plants. The thick solid arrow indicates the reaction step characterized in this work. Black thin solid arrows indicate the single reaction step characterized previously, and dashed arrows indicate the uncharacterized step. White thick arrows represent multiple reaction steps. The names of enzymes are indicated under the arrows. 3βHSD, 3β-hydroxysteroid dehydrogenase; 3KSI, 3-ketosteroid isomerase; S5αR, steroid 5α-reductase; 3KSR, 3-ketosteroid reductase. The genes previously characterized are shown in the upper part of the arrows. *Sl3βHSD1* involved in the conversion of dehydrotomatidine to tomatidine in tomatoes; *DWF4*, *dwarf4*; *DET2*, *de-etiolated-2*; *CPD*, *constitutive photomorphogenesis dwarf*.

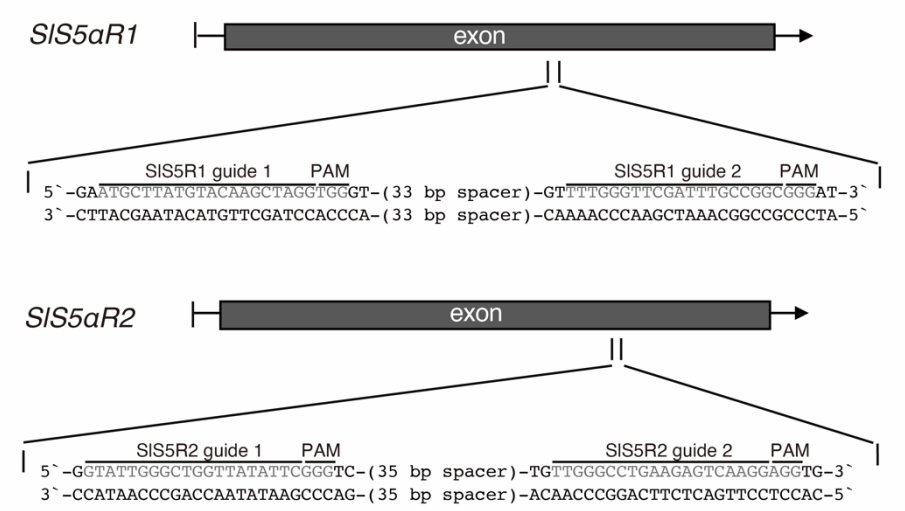


Fig. 2-2. Target sites for *SlS5αR1* and *SlS5αR2*. The 20-bp target sequences and protospacer adjacent motif (PAM) sequences are indicated in gray.

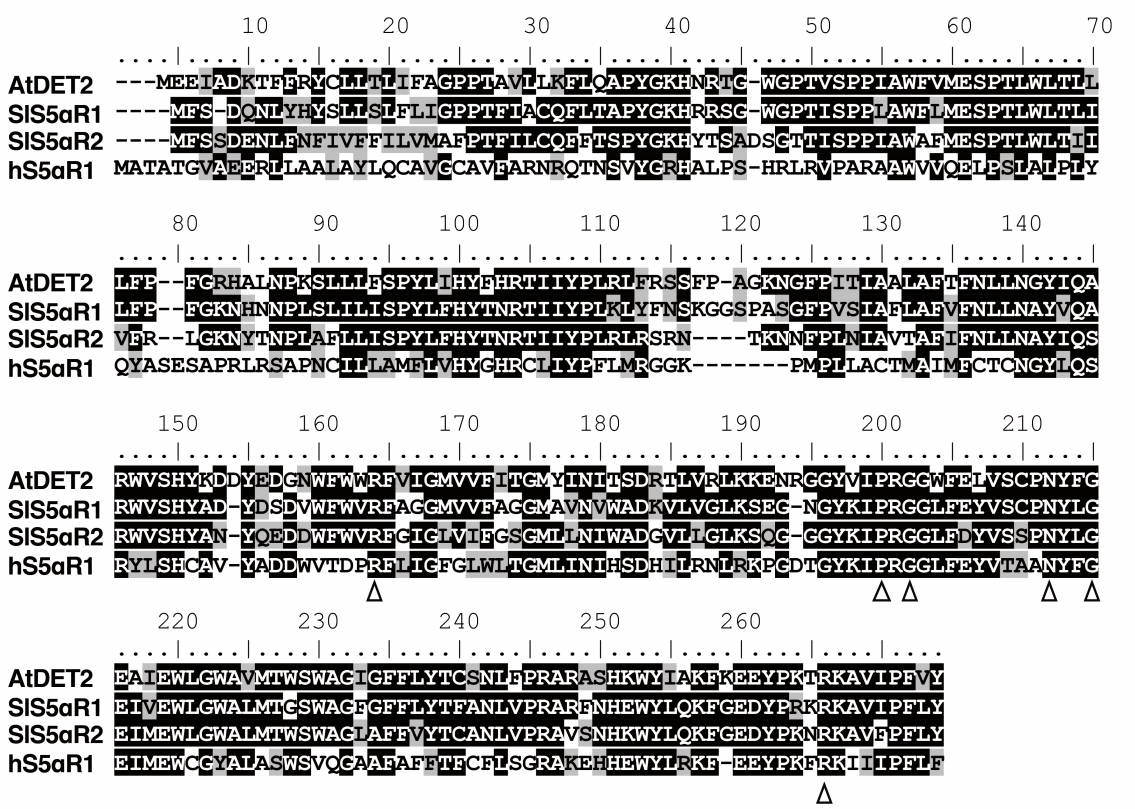


Fig. 2-3. Multiple sequence alignment of SlS5αR1 and SlS5αR2 with AtDET2 from *Arabidopsis thaliana* and human hS5αR1. Black and gray shades indicate identical and similar amino acid residues, respectively. Conserved residues, which are required for steroid 5α-reduction activity, are marked by white triangles.

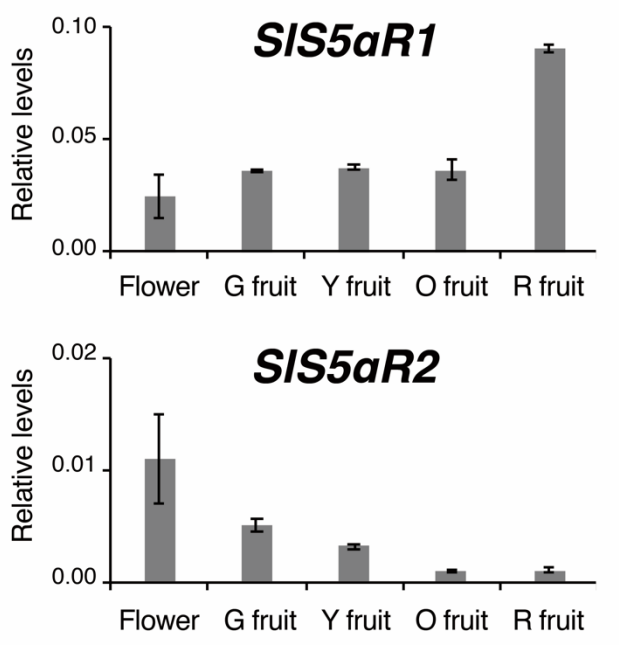


Fig. 2-4. Quantitative RT-PCR analysis of *SlS5αR1* and *SlS5αR2* expression in various tissues of tomatoes. G, Y, O, and R mean green, yellow, orange, and red, respectively. Transcript levels of *SlS5αR1* and *SlS5αR2* are shown relative to that of the *ubiquitin* gene as an internal reference. Error bars indicate standard error from the mean (n = 3).

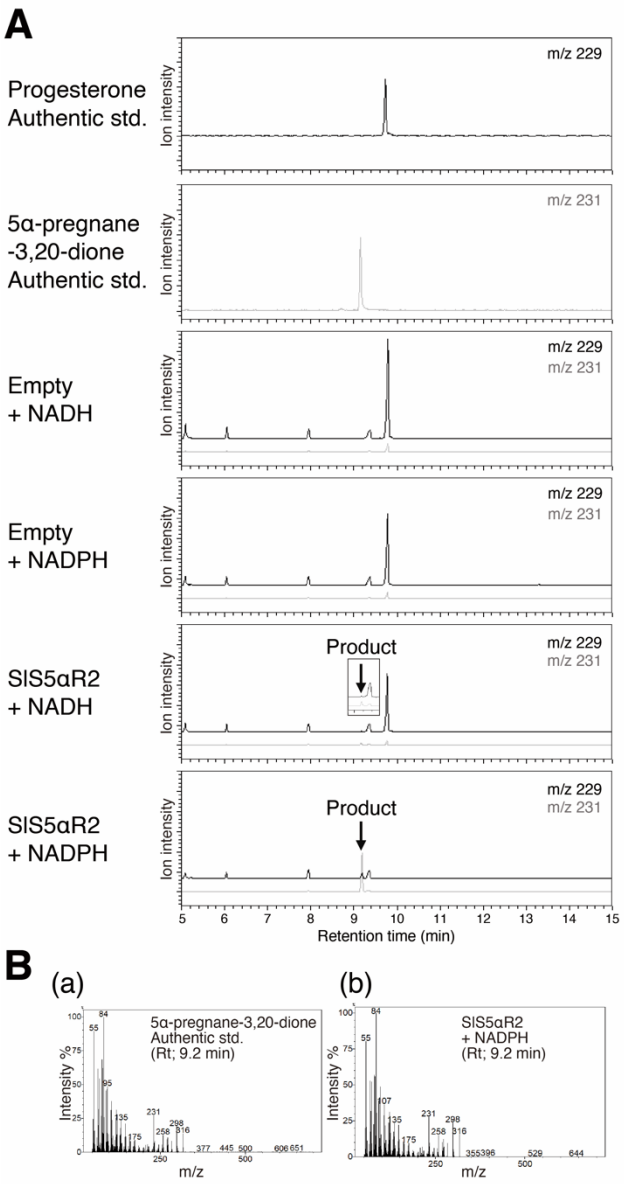


Fig. 2-5. *In vitro* assay of the C5 reduction activity of SlS5αR1 with progesterone as the substrate. (A) Progesterone was monitored by selected ion monitoring (SIM) at *m/z* 229 and is shown in black, and the product 5α-pregnane-3,20-dione was monitored at *m/z* 231 and is shown in gray. (B) Mass spectrum of the (a) authentic 5α-pregnane-3,20-dione at a retention time of 9.2 min and (b) product peak of progesterone incubated with SlS5αR2 and NADPH.

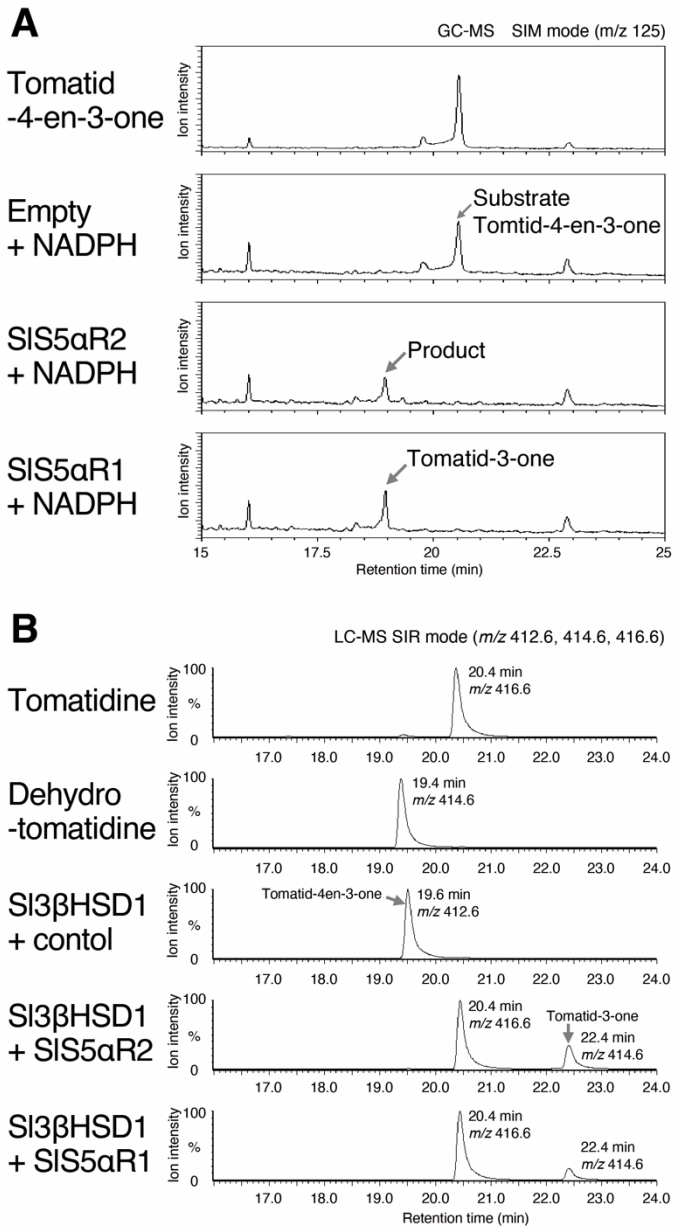


Fig. 2-6. *In vitro* assays of SlS5αR1 and SlS5αR2 with tomatid-4-en-3-one. (A) GC-MS analysis of the in vitro assays of SlS5αR1/2 and NADPH incubated with tomatide-4-en-3-one as a substrate. Substrate and products were monitored by SIM mode at *m/z* 125. (B) LC-MS analysis of the in vitro assays of SlS5αR1/2 and Sl3βHSD1 incubated with dehydrotomatidine as a substrate. Substrate and products were monitored by SIM mode at *m/z* 412, 414, and 416.

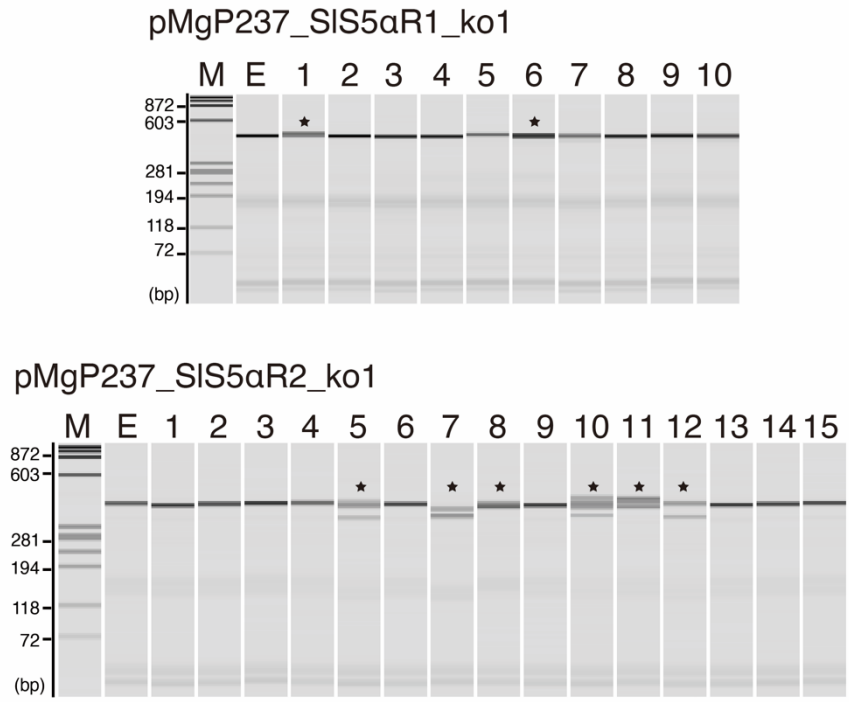


Fig. 2-7. Heteroduplex mobility assay with the MultiNA electrophoresis system. Multiple heteroduplex peaks were detected in PCR amplicons from the CRISPR/Cas9 transgenic tomato hairy roots. Black stars indicate the potential mutated lines. E: hairy roots transformed by empty pMgP237 vector. M: internal marker.

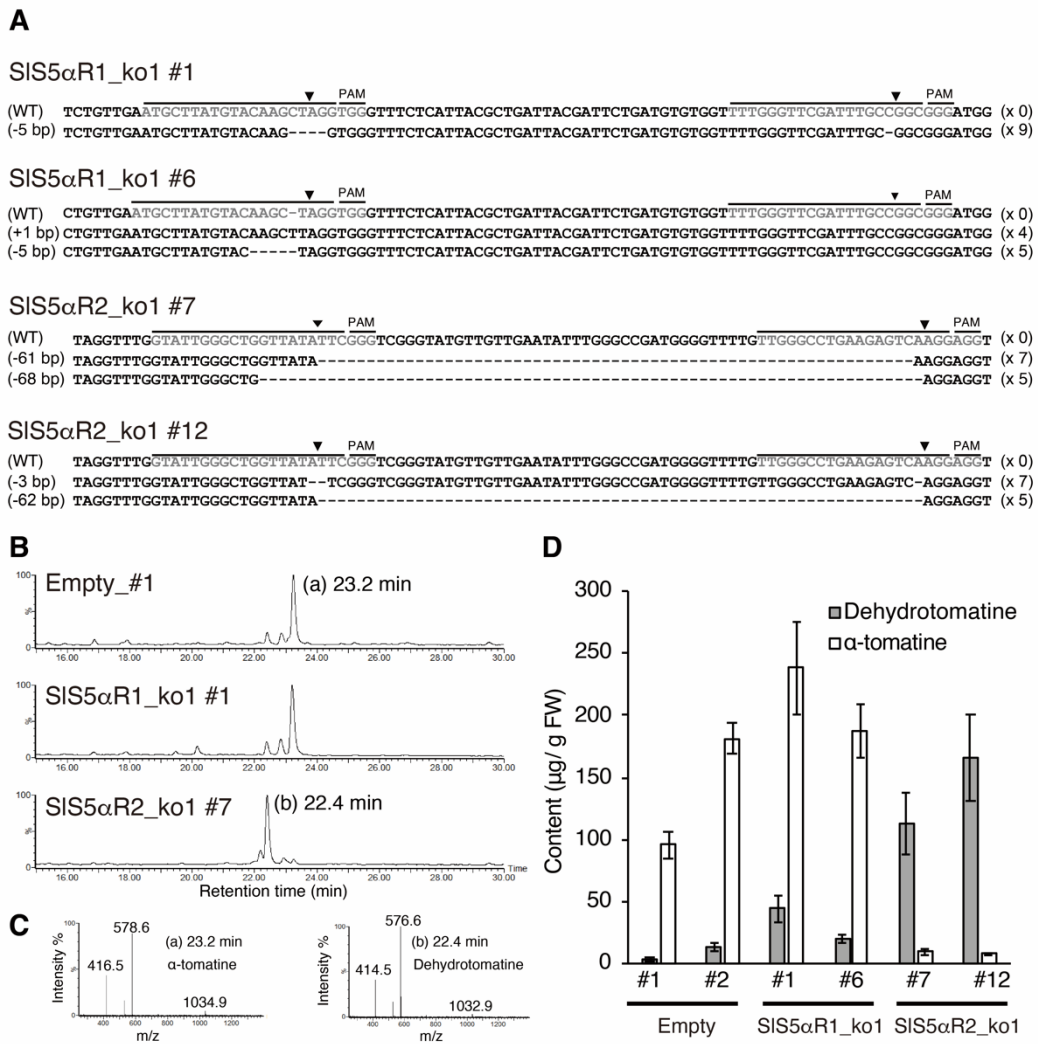


Fig. 2-8. Characterization of *SlS5αR1* or *SlS5αR2*-disrupted tomato hairy roots. (A) DNA sequences of gRNA target sites. Black and gray lines indicate gRNA target sites and PAM sequences, respectively. Black triangles indicate cleavage positions of target sites. Deletions or insertions of nucleotides are presented on the left sides of the sequences, and the numbers of detected sequences are shown on the left sides. (B) LC-MS analysis of SGAs in SlS5αR1_ko1_#1 and SlS5αR2 _ko2_#7. (C) Mass spectrum of the α-tomatine (peak at 23.2 min) and dehydrotomatine (peak at 22.4 min). (D) Quantification of α-tomatine and dehydrotomatine contents using LC-MS. Error bars indicate standard error (n = 3).

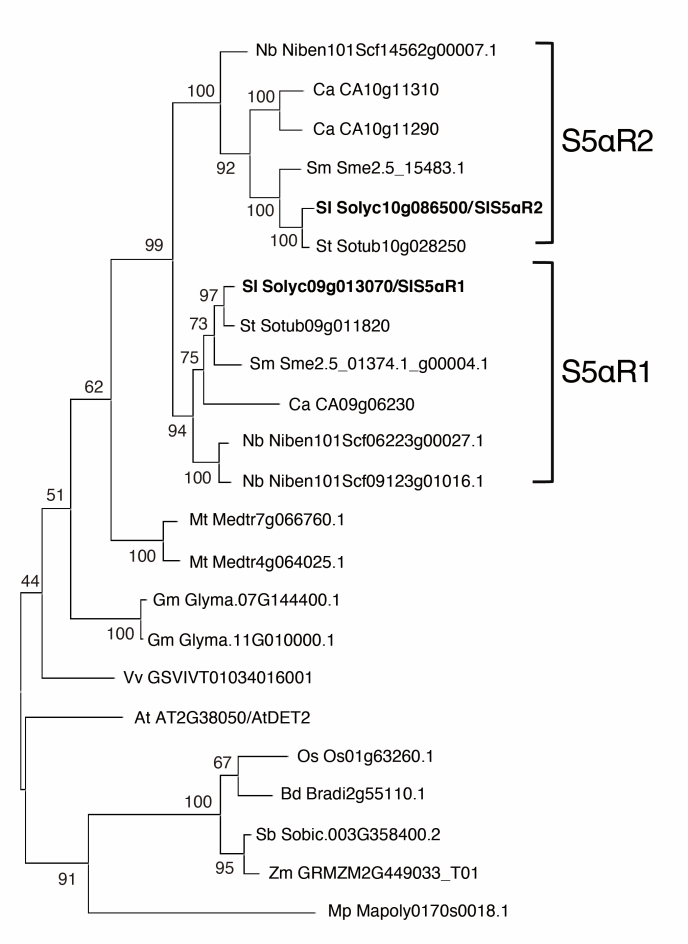


Fig. 2-9. Phylogenetic analysis of plant AtDET2 homologs. Species were represented tomato (Sl), potato (St), eggplant (Sm), *Capsicum annuum* (Ca), *Nicotiana benthamiana* (Nb), *Brachypodium distachyon* (Bd), soybean (Gm), *Medicago truncatula* (Mt), rice (Os), *Sorghum bicolor* (Sb), *Vitis vinifera* (Vv), and maize (Zm). A phylogenetic tree was generated using the maximum likelihood method in MEGA7. Bootstrap values based on 1,000 replicates are shown at the branching points.

Table 2-1. Oligonucleotides used in the present study.

Primer No.	Descripton	Sequences
1	SlS5αR1qPCR Fw	5`-TGTGGTTTTGGGTTCGATTTGCCG-3`
2	SlS5αR1qPCR Rv	5`-TCACTCTTCAGGCCCACTAACACC-3`
3	SlS5αR2 qPCR Fw	5`-TTCACTTCTCCTTACGGCAAACACTACA-3`
4	SlS5αR2 qPCR Rv	5-`CGAAATACGATAATTGTGAGCCAAAGCGT-3`
5	ubiquitin qPCR Fw	5`-CACCAAGCCAAAGAAGATCAAGC-3`
6	ubiquitin qPCR Rv	5`-TCAGCATTAGGGCATCCCTTACG-3`
7	SlS5αR1 ORF BamHI Fw	5`-GGATCCATGTTTTCAGATCAAAACTTGTATCAC-3`
8	SlS5αR1 ORF SalI Rv	5`-GTCGACTCAATACAAAAATGGAATAACAGCTTTT-3`
9	SlS5αR2 ORF BamHI Fw	5`-GGATCCATGTTCTCTTCAGATGAAAATCTCTTC-3`
10	SlS5αR2 ORF SalI Rv	5`-GTCGACTCAATACAAAAAGGGGAAAACAGC-3`
11	Sl3βHSD1 ORF BamHI-NdeI Fw	5`-GAATTCCATATGGCAAATAAGCTCAGGTTGGAGGGC-3`
12	Sl3βHSD2 ORF SalI Rv	5`-GTCGACTTATTGTAGCTTCAAAATAGAACTTAGTCCACC-3`.
13	F2_tgRNA_SlS5R1_492-511	5`-TTGGGTCTCGTGCAGATGCTTATGTACAAGCTAGGGTTTTAGAGCTAGAAATAGCA-3`
14	R2_tgRNA_SlS5R1_552-571	5`-TTGGGTCTCCAAACGCCGGCAAATCGAACCCAAACTGCACCAGCCGGGAATCGAA-3`
15	F2_tgRNA_SlS5R1_449-468	5`-TTGGGTCTCGTGCAGTATTGGGCTGGTTATATTCGTTTTAGAGCTAGAAATAGCA-3‘
16	R2_tgRNA_SlS5R1_511-530	5`-TTGGGTCTCCAAACCCTTGACTCTTCAGGCCCAACTGCACCAGCCGGGAATCGAA-3`
17	AtU6-pMgP237 Fw	5`-GGCCCCTGGGAATCTGAAAG-3`
18	AtU6-pMgP237 Rv	5`-GGAAGAAGAAATCGATCTGGAAAATTTTGC-3`
19	SlS5αR1_g43 Fw	5`-CTCTTCCTCATCGGACCACC-3`
20	SlS5αR1_g501 Rv	5`-CCAAACATTCACCGCCATCC-3`
21	SlS5αR2_g296 Fw	5`-TTCGCCTCCGTAGCAGAAAC-3`
22	SlS5αR2_g678 Rv	5`-TCGAAACAGCTCGAGGAACC-3`

Table 2-2. Candidate S5αRs from tomatoes were selected based on the BLASTP search in the Sol Genomics Network (http://solgenomics.net). RPKM values were obtained from the Tomato Functional Genomics Database (http://ted.bti.cornell.edu).

Clone ID	Bud	Flower	Leaf	Root	1cm_Fruit	2cm_Fruit	3cm_Fruit	Mature green	Breaker
SlS5αR1	34.5	21.4	16.4	15.0	41.9	27.5	38.0	55.3	43.5
SlS5αR2	74.8	12.8	30.2	22.9	105.3	32.3	7.4	3.8	1.2

Chapter 3

Chemical evolution of potato toxin α-solanine from tomato glycoalkaloids

Introduction

Potato (*Solanum tuberosum*), members of the Solanaceae plant family, is one of the worldwide major food crops. Despite its status as global important food source, potato produces toxic compounds of steroidal glycoalkaloids (SGAs), α-solanine and α-chaconine, that confer a bitter taste and cause gastrointestinal and neurological disorders to humans. Therefore, controlling the SGA level is an important target for potato breeding.

SGAs are class of specialized metabolite mainly found in Solanaceae plants, and are famous as antinutritional substance in major *Solanum* vegetable crops such as potato (*S. tuberosum*) tomato (*Solanum lycopersicum*), and eggplant (*Solanum melongena*) [3]. These compounds are associated with plant resistance to pathogens and predators based on their toxicity to a wide range of organisms including bacteria, fungi, insect and animal [4–6]. SGAs consist of two structural components; the aglycone unit composed of C27 steroid containing a nitrogen moiety and oligosaccharide attached to the hydroxy group at C-3 [6]. Various combinations of the aglycone and sugar moieties generate enormous structural diversity of SGAs, and their chemical structures determine their biological activities. Most of the SGAs produced in *Solanum* plants are composed of the two types of aglycones, spirosolane-type or solanidane-type (Supplementary Fig. II), in which the structural difference is at the E,F-ring structure.

Potatoes typically contain α-solanine and α-chaconine, which share the same solanidane-type aglycone, solanidine, and these SGAs comprise upward of 90% of the total SGA contents in domesticated potatoes [14,15]. Additionally, more than 50 different

SGAs, including spirosolane-type SGAs, have been identified in a variety of wild potato species[15]. In contrast, tomatoes contain α-tomatine and dehydrotomatine, which are composed of spirosolane-type aglycones, tomatidine and tomatidenol (dehydrotomatidine), respectively. Although more than 100 SGAs have been described in various tissues and developmental stages[17,18], accumulation of solanidane-type compounds have not been reported in tomatoes.

SGAs are biosynthesized from a common precursor cholesterol[20]. SGA biosynthesis can be divided into two main parts; aglycone formation and glycosylation. During aglycone formation, cholesterol is subsequently modified through multiple reaction steps including hydroxylation, transamination, E- and F-ring closure[2,4,22]. Three cytochrome P450 monooxygenases (CYPs) named as PGA2/CYP72A188, PGA1/CYP72A208/GAME8, PGA3/CYP88B/GAME4 have been found to be involved in hydroxylation of cholesterol at C22 and C26 and oxygenation at C26, respectively[14,21]. A 2-oxoglutarate dependent dioxygenase (DOX), and an aminotransferase were reported to be required for the C16α hydroxylation and C26 amination during SGA biosynthesis in potato and tomato[24]. These enzymes and functions are common to potato and tomato, suggesting that they are involved in biosynthetic steps common to solanidane-SGAs and spirosolane-SGAs (Fig. 3-1). In addition, several UDP-glycosyltransferases (UGTs) involved in the glycosylation steps of SGA biosynthesis have been identified in potato and tomato[16,25–28]. However, the pathways and genes involved in E,F-ring closure leading to solanidane-skeleton in potato and spirosolane-skeleton in tomato, respectively, were completely unknown.

In our effort to elucidate the E,F-ring formation mechanism in SGAs biosynthesis, we focused on solanidane-type SGAs: α-solanine and α-chaconine,

biosynthesis in potato. In the present study, we elucidated that solanidane-type SGA is biosynthesized from spirosolane-type SGA via E,F-ring rearrangement (Fig. 3-1). We discovered *StDOX130*, a member of the 2-oxoglutrate dependent dioxygenase (DOX) superfamily, is required for solanidane-type SGAs biosynthesis. StDOX130 in vitro exhibited novel activity converting spiorosolane-skeleton to solanidane-skeleton (Fig. 3-1). Furthermore, examination of genome structure revealed that production of solanidane-type SGAs has been driven by tandem gene duplication. These results offer insights into the chemical evolution of SGAs in *Solanum* species.

Material and Methods

Chemicals

Authentic samples of α-solanine, α-chaconine, and solanidine were purchased from Sigma-Aldrich, and dehydrotomatine and tomatidenol were isolated in our laboratory from α-tomatine (Tokyo Chemical Industry Co., Ltd.) and tomatidine (Chromadex) using HPLC, respectively. α-solamarine and β-solamarine were purified from diploid potato clone 97H32-6 in our laboratory and identified with LC-MS and NMR analyses. The leaves (350 g) of *S. tuberosum* 97H32-6 were extracted with 100% MeOH (1 L) for 24 h at room temperature. Then, the mixture was filtered *in vacuo* through No. 2 filter paper (Advantec, Tokyo, Japan), and the residue was repeatedly extracted with 100% MeOH (1 L). The solutions were combined and concentrated in a vacuum. The extract (20.0 g) was suspended in water (300 mL) and partitioned with *n*-hexane (300 mL × 3) and EtOAc (300 mL × 3). The H_2O layer was concentrated with a rotary evaporator at 30 °C under reduced pressure. The residue (8.3 g) was subjected to open column chromatography on YMC-GEL ODS-A (YMC Co., Ltd.) and eluted with H_2O/MeOH (100:0, 50:50, 25:75,

0:100; v/v, step-wide system, 800 mL) as elution solvents. The fraction (330.0 mg) eluted with 100% MeOH was subjected to a Sephadex LH-20 column (GE Healthcare) and then eluted with MeOH. The eluted fraction containing steroidal glycoalkaloid was separated with a preparative-scale HPLC (Shimadzu, Kyoto, Japan), operated by the following system: ODS column (5C_{18}-PAQ, 20 × 250 mm, Nacalai) protected by a guard column (5C_{18}-PAQ, 10 × 10 mm, Nacalai) of an isocratic system of H_2O (0.1% trifluoroacetic acid)/MeCN (70:30, v/v) at 40ºC. Flow rate was 5.0 mL min^{-1} and eluents were monitored at 203 nm. Two fractions, E1 (Rt 21.2 min, 52 mg) and E2 (Rt 22.8 min, 101 mg) were obtained. Each fraction was dissolved in pyridine-d_5 (C_5D_5N, Aldrich). NMR spectra were recorded on a Bruker AVANCEIII 600 spectrometer (Bruker BioSpin Corp.,) at 600 MHz for ^1H and 151 MHz for ^{13}C. The NMR chemical shifts were referenced to tetramethylsilane. The assignments of the ^1H and ^{13}C signals were based on ^1H–^1H correlation spectroscopy (COSY), heteronuclear single quantum coherence (HSQC), heteronuclear multiple bond correlation (HMBC), and nuclear Overhauser enhancement and exchange spectroscopy (NOESY) spectra.

Real-Time Quantitative RT-PCR Analysis

Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) and the RNase-Free DNase Set (Qiagen). Total RNAs of potato (*Solanum tuberosum* cv Sassy) were prepared from the leaves, stems, flowers, roots, stolons, tuber peels, and tuber sprouts. The extracted total RNAs of potato were used to synthesize the first-strand cDNAs using the ReverTra Ace ® qPCR RT Master Mix with gDNA Remover. Quantitative RT-PCR was performed with a LightCyclerNano (Roche) using FastStart Essential DNA Green Master (Roche) with the following primer sets: 1 and 2 for *StDOX130*, 3 and 4 for *StDOX070*,

and 5 and 6 for EF1α[60] (Table 3-1). Cycling was carried out at 95°C for 10 min, 45 cycles at 95°C for 10 s, 60°C for 10 s, and 72°C for 15 s for amplification, followed by holding at 95°C for 30 s and ramping up from 60°C to 95°C at 0.1°C s^{-1} to perform a melting curve analysis. Three biological replicates were analyzed in duplicate. The gene expression levels were normalized against the values obtained for the EF1α gene, which were used as internal references in potato. Data acquisition and analysis were performed using LightCyclerNano software (Roche).

Cloning of StDOX130 and StDOX070 cDNAs

The potato cDNA template was prepared from mRNA isolated from sprouts of *S. tuberosum* cv Sassy as described above. The cDNA fragments that contained the open reading frame of *StDOX130* and *StDOX070* gene were amplified by RT-PCR using cDNA from tuber sprout as a template with primers 7 and 8 for *StDOX130* and primers 9 and 10 for *StDOX070* (Table3-1), which were designed from the potato unigene sequences; Sotub01g007130 and Sotub01g007070 in the Potato ITAG protein database, respectively. The PCR products were cloned into the pCR®4 Blunt-TOPO® vector (Thermo Fisher Scientific).

Generation of StDOX130i transgenic potato plants.

The *StDOX130*-RNAi binary vector carried the inverted repeat of partial *StDOX130* fragment interposing the third intron of the Arabidopsis (*Arabidopsis thaliana*) *At4g14210* gene under the control of the *cauliflower mosaic virus* 35S (CaMV35s) promoter in the T-DNA region. The construction scheme is shown in Fig. 3-2. Briefly, pRI201-AS (TaKaRa) was digested with KpnI and SmaI to disrupt multi

cloning site 2. The protruding end of the digested site was blunted with T4 DNA polymerase (TaKaRa) and then the blunted ends was ligated. The resulting plasmid was named pRI201-⊿MCS2 in this work. The third intron of Arabidopsis gene At4g14210 was PCR amplified from pKT258[24] with primers 11 and 12, which contained restriction sites (Table 3-1). The amplified fragments were cloned into XbaI and NdeI sites of pRI201-⊿MCS2, and resulting plasmid was named pRNAi201. A 262 bp fragment of *StDOX130* coding region was PCR amplified using primers 13 and 14 (Table 3-1). The fragment was inserted into XbaI/BamHI and NdeI/SalI sites of pRNAi201 to construct *StDOX130i* vector. Potatoes (*S. tuberosum* cv Sassy) were transformed using *A. tumefaciens* EHA105 cells harboring pRNAi201-*StDOX130* vector. Briefly, internodal stem sections were excised from in vitro grown potato shoot and cut into about 5 mm. Explants were co-cultivated with agrobacterium for 48 h at 20℃ on MS medium containing 3% w/v sucrose, trans-zeatin (2 μg mL^{-1}), indole-3-acetic acid (50 ng mL^{-1}), Kanamycin (50 μg mL^{-1}) and acetosyringone (100 μM) under dark conditions. After co-cultivation period, stem pieces were transferred to shoot induction medium composed of agarose solidified MS medium containing 3% w/v sucrose, trans-zeatin (2 μg mL^{-1}), indole-3-acetic acid (50 ng mL^{-1}), Kanamycin (50 μg mL^{-1}) and Meropenem Hydrate (50 μg mL^{-1}), and cultured at 20 °C in16-h/8-h light/dark cycles. Stem pieces were transferred to fresh shoot induction medium every 2 weeks, until plantlets emerge. Well-developed plantlets were excised and transferred to MS medium containing 3% w/v sucrose, Kanamycin (50 μg mL^{-1}) and Meropenem Hydrate (50 μg mL^{-1}). After 2-3 weeks, well-rooted shoots were selected and then transformants were individually screened by genomic PCR using primer sets 15 and 16 (Table 3-1), which targeted the kanamycin resistance gene in the T-DNA region integrated into the genome.

Characterization of StDOX130i transgenic potato plants.

Total RNA was then prepared from the leaves of four independent lines of in vitro-cultured plants: #6, #13, #17, and #21. Quantitative RT-PCR analysis of StDOX130 and StDOX70 were performed using primers 1/2 and 3/4, respectively (Table 3-1) as described in "Real-Time Quantitative RT-PCR Analysis." The SGAs accumulated in *StDOX130i* transgenic plants were extracted as described previously as follows: Briefly, 50 mg tissue from fresh plant materials were frozen and homogenized with a mixer mill at 4°C and the homogenate were extracted with 300 μL MeOH. After centrifugation, the supernatant was collected. The extraction was repeated three times and the extract, dried *in vacuo*, was re-dissolved in 400 μL MeOH. 5 μL of solution was diluted with 295 μL MeOH and filtered through a 0.22-μm polytetrafluoroethylene membrane filter. An aliquot (2 μL) was nalyzed by UPLC-MS. LC-MS analysis was performed using a system consisting of an ACQUITY UPLC H-Class System (Waters) and an SQ Detector 2 (Waters), and data acquisition and analysis were performed using MassLynx 4.1 software (Waters). Each sample was injected into an ACQUITY UPLC HSS T3 chromatographic column (100 × 2.1 mm, 1.7 μm; Waters), in which the column temperature was set at 40°C and the flow rate was set at 0.2 ml min^{-1}. The mobile phases were water with 0.1% (v/v) formic acid (A), acetonitrile (B), using a gradient condition as follows: solvent B ramped linearly from 10% to 32.5% over 30 min; solvent B increased linearly to 100% over 1 min, held at solvent B 100% for 4 min; solvent B then returned immediately to 10% followed by a 5 min re-equilibration period. The mass spectra were obtained in positive electrospray ionization, with a capillary voltage of 3 kV and a sample cone voltage of 60 V. Mass spectrometry scan mode with a mass range of *m/z* 350 to 1250 was used. Quantifications of α-solanine, α-chaconine, α-solamarine, and β-solamarine were

calculated from the ratio of peak area at m/z 868, 852, 884, and 868 from positive ion scans using a calibration curve of authentic samples, respectively.

Feeding experiments

St16DOX-disrupted potato hairy roots were established in our laboratory[52] and subcultured every 1 month on orbital shaker (100 rpm) in Gamborg B5 (B5) liquid medium containing 2% (w/v) sucrose and 250μg/ml cefotaxime at 20℃ under 24-h dark condition and used for feeding. α-solamarine, β-solamarine, α-tomatine and tomatidenol (MeOH solution, final concentration 10μM) were individually added to the liquid medium of three-weeks-old the hairy roots. For the inhibition experiment, prohexadione (acetone solution) or uniconazole-P (acetone solution) was concomitantly added to the aquaculture with α-solamarine. After three-days cultivation, steroidal glycoalkaloids contained in the harvested hairy roots were extracted from 100 mg fresh samples with 300μl methanol for three times. The extracting solution was evaporated and the residue was dissolved 200 μl methanol. After centrifugation, 10 μl of supernatant was diluted with 290 μl methanol and aliquot (2 μl) of filtrate was analyzed by LC-MS as described above in "Characterization of StDOX130i transgenic potato plants" with minor modification. Briefly, the mobile phases were water with 0.1% (v/v) formic acid (A), acetonitrile (B), using a gradient condition as follows: solvent B ramped linearly from 10% to 42.5% over 15 min; solvent B increased linearly to 100% over 4 min, held at solvent B 100% for 5 min; solvent B then returned immediately to 10% followed by a 5 min re-equilibration period. In the case of α-tomatine feeding experiment the mobile phases were water with 0.1% (v/v) formic acid (A), acetonitrile (B), and methanol (C), using a liner gradient elution as follows: from solvent A 90% /solvent B 5% /solvent C

5% to solvent A 45% /solvent B 27.5% /solvent C 27.5% over 30 min, from solvent A 45% /solvent B 27.5% /solvent C 27.5% to solvent A 25% /solvent B 37.5% /solvent C 37.5% over 5 min, from solvent A 25% /solvent B 37.5% /solvent C 37.5% to solvent A 0% /solvent B 50% /solvent C 50%, held at solvent A 0% /solvent B 50% /solvent C 50%, solvent B and solvent C then returned immediately to 5% followed by a 5 min re-equilibration period.

Expression of the recombinant DOX proteins in Escherichia coli

The coding sequences of *StDOX130* and *StDOX070* were digested from the pCR®4 Blunt-TOPO® vector (Thermo Fisher Scientific) with NdeI and SlaI. Since *SlDOX6585* could not be amplified from cDNA from any tissue in tomato, coding sequence of *SlDOX6585* with NdeI and Sal1 sites were synthesized (Eurofins Genomics Inc.) The DNA fragments were then ligated into the Nde1-Sla1 sites of pCold ProS2 (TaKaRa). *E.coli* strain BL21 (DE3) (Clontech) transformed with the constructed plasmid was grown at 37℃ in LB medium containing 50μg/ml ampicillin until OD_{600} reached 0.5. Recombinant protein production was induced by adding 0.1mM isopropyl β-D-1-thiogalactopyranoside and the culture were grown for 24h at 15℃. The culture was then centrifuged at 8,000 rpm for 10 min at 4℃, and the cell pellets were resuspended in 5 ml of cold sonication buffer containing 50 mM Bis-Tris-HCl (pH7.2), 150 mM NaCl, 10% (v/v) glycerol, 5mM dithiothreitol. The solution was then sonicated three times for 30 s each on ice using a Bandelin Sonopuls HD 2070 ultrasonic homogenizer type MS73 (Sigma-Aldrich) at a sound intensity of 200 W cm^{-2} and centrifuged at 15,000 rpm for 10 min at 4°C. The recombinant His-tagged protein contained in the soluble fraction was purified using His SpinTrap TALON (GE Healthcare) according to the manufacturer's instructions. After

two column washes, the adsorbed proteins were eluted twice in 200 µl of an elution buffer containing 50 mM sodium phosphate (pH 7.4), 30 mM NaCl, and 150 mM imidazole, and the elution was mixed. The purified recombinant proteins were used for further analyses.

In vitro recombinant enzyme activity assay

The SGAs and their aglycone; α-solamarine, β-solamarine, α-tomatine, α-solanine, α-chaconine, tomatidenol, tomatidine, and solanidine were used as substrates in recombinant protein assay. An in vitro enzyme activity assay was performed using 100 µL of reaction mixture that consisted of 100 mM Bis-Tris-HCl (pH 7.2), 5 mM 2-ketoglutaric acid, 10 mM sodium ascorbate, 0.2 mM $FeSO_4$, 25 µM substrate and 1 µg purified recombinant proteins as an enzyme. The reaction was initiated by the addition of the enzyme and carried out at 30°C for 30 min. The reaction was then stopped by incubation for 2 min at 90℃. After centrifugation, 20 µl of supernatant was diluted with 180 µl methanol and filtered through 0.2-µm nylon membrane filters (Acrodisc; Waters). An aliquot (2 µl) was analyzed using LC-MS as described in "Feeding experiments". We determined the kinetic parameters of recombinant StDOX130 and StDOX070 in triplicate assays. Assay was performed conducted as described above with minor modifications. Briefly, 0.1 µg purified StDOX130 protein or 0.02 µg purified StDOX070 were used. The activity was assayed using α-solamarine or tomatidenol at a concentration ranging from 0.25 µM to 50 µM, and the reaction was carried out at 30°C for 10 min. 20 µl of assay mixture was diluted with 180 µl methanol and filtered through 0.2-µm nylon membrane filters (Acrodisc; Waters). An aliquot (2 µl) was analyzed using LC-MS as described

above. Kinetic parameters were determined by nonlinear regression using the ANEMONA program[61]

Structural determination of the StDOX130 enzymatic reaction product

To determine the structure of the product of the enzymatic reaction catalyzed by recombinant StDOX130, the enzymatic reaction was performed using 50 ml of reaction mixture as described above with a minor modification; the reaction was conducted using the crude enzyme overnight. The reaction mixture was extracted with water saturated butanol after the pH adjusted to 3.0 by the addition of hydrochloric acid. The butanol layer was corrected and StDOX130 product was concentrated *in vacuo* and the dried residue was dissolved in methanol. StDOX130 reaction product dissolved in methanol was subjected to preparative scale HPLC. The sample was injected into an ODS column (COSMOSIL $5C_{18}$-PAQ, 20×250 mm, Nacalai), in which the column temperature was set at 40ºC and the flow rate was set at 4 mL min^{-1}. Separation was performed with isocratic mobile phases of 22% acetonitrile in H_2O containing 0.1% trifluoroacetic acid (v/v) and elute was detected at 203 nm. Then, the separated StDOX130 reaction product was concentrated.

The molecular formula of the DOX 130 enzymatic reaction product was determined as $C_{50}H_{81}NO_{21}$ based on HRMS analysis (*m*/*z* 1032.5380, $[M+H]^+$). Its structure was determined by two-dimensional NMR measurements based on ^1H-^1H correlation (COSY), nuclear Overhauser enhancement and exchange spectroscopy (NOESY), heteronuclear single quantum coherence (HSQC), and heteronuclear multiple bond correlation (HMBC) techniques. NMR measurement in this study was done in CD_3SOCD_3 and CD_3OD. ^{13}C and ^1H assignment is summarized in Table 3-2 and 3-3, and

COSY and HMBC correlation is shown in Fig. 3-3. Sample was dissolved in NMR solvent immediately prior to measurement. In CD_3OD, signals corresponding to H-23, observed at 2.84 ppm (br ddd, J = 21.3, 8.9, 6.8 Hz) and 3.12 ppm (1H, br ddd, J = 21.3, 4.1, 4.1) at the beginning, disappeared during measurement within 21 h. The observation indicated H-D exchange in CD_3OD. Accordingly, C-23 was observed severely broadened signal in CD_3OD by C-D coupling. In both solution, C-16 signals were observed at 110.95 (in CD_3SOCD_3) and 117.85 ppm (in CD_3OD), indicating their acetal-like nature. Since C-16 of α-tomatine (the substrate of the StDOX130 reaction) bears two carbon, one oxygen, and one hydrogen atoms, the chemical shift of C-16 of the DOX130 product indicates that oxidation occurred on C-16. In contrast, chemical shift of C-22, consisted hemiaminal structure in α-tomatine, observed at 192.21 ppm (in CD_3SOCD_3) and 196.80 ppm (in CD_3OD) in DOX130 product, at first suggested that C-22 exists as carbonyl carbon of ketone or aldehyde. However, the HMBC correlation between H-26 and C-22 indicated that C-22 exists as imine carbon instead of carbonyl carbon. Since chemical shift of typical imine carbon observed around 140–170 ppm,[62] the chemical shift of C-22 of the StDOX130 product appeared at considerably lower field. The observation indicates the presence of iminium ion structure in the StDOX130 product, and the reported ^{13}C chemical shift of tri- or tetra-substituted carbon atom of iminium ions (over 180 ppm, Fig. 3-4)[63] is comparable to the C-22 chemical shift of the StDOX130 product. Especially, the chemical shift of the tetra-substituted iminium ion in Fig. 3-4 (195.1 ppm) is in good accordance with those in the StDOX130 product (192.21 or 196.80 ppm). Furthermore, the NOE correlation between H-26_b and H-15β indicates proximity of these hydrogen atoms. Additionally, the above mentioned ^{13}C chemical shift of C-16 indicated not only one oxygen atom but also one nitrogen atom binds to C-16. Accordingly, the structure of

the StDOX130 product was determined as a zwitterion species shown in Fig. 3-5, in which iminium ion structure is constructed between C-16, the nitrogen atom, and C-22 to form. The previously mentioned H-D exchange of H-22 in CD_3OD, which further confirms the iminium ionic form of the StDOX130 product, is illustrated in Fig. 3-6. The cationic nature of the nitrogen atom would increase the acidity of H-22 and accelerate the H-D exchange reaction.

Phylogenetic analysis

Amino acids sequences of StDOX070 (Sotub01g007070), StDOX080 (Sotub01g007080), StDOX090 (Sotub01g007090), StDOX100 (Sotub01g007100), StDOX110 (Sotub01g007100), StDOX120 (Sotub01g007120), StDOX130 (Sotub01g007130), StDOX150 (Sotub01g007150) were obtained from the Spud DB potato genomics resource (http://solanaceae.plantbiology.msu.edu/). Amino acids sequences of Sl23DOX/GAME31 (Solyc02g062460), Sl23DOX/GAME31-like1 (Solyc02g06250), Sl23DOX/GAME31-like2 (Solyc02g062470), Sl23DOX/GAME31-like3 (Solyc02g062490), SlDOX580 (Solyc01g006580), SlDOX585 (Solyc01g006585), and SDOX610 (Solyc01g006610) were downloaded from Sol Genomics Network. Sequence alignments were performed using MUSCLE. The Maximum Likelihood tree was inferred in MEGA10 using 1000 bootstrap replications.

Results

Feeding rescue experiments in α-solanine biosynthesis

Cultivated potatoes (*Solanum tuberosum*) contain only solanidane-type SGAs (α-solanine and α-chaconine), while some wild potato species accumulate spirosolane-

type SGAs, α-solamarine and β-solamarine as well[2,64,65], suggesting that spirosolane-type SGAs might be the biosynthetic intermediates for solanidane-type SGAs. To explore this hypothesis, we conducted feeding experiments using *St16DOX*-knockout potato hairy roots (i.e., 16DOX-koHR), which do not contain detectable SGAs but accumulate the biosynthetic intermediates, the glycosides of 22,26-dihydroxycholesterol[52]. Firstly, the administration of α-solamarine into 16DOX-koHR resulted in the restoration of the accumulation of α-solanine in 16DOX-koHR (Fig. 3-7a and Fig. 3-8). Similar results were obtained from the feeding experiments with β-solamarine, in which the accumulation of α-chaconine was detected (Fig. 3-9). Then, feeding with tomatidenol, which is an aglycone of α-solamarine and β-solamarine, resulted in the accumulation of α-solanine, α-chaconine and uncharacterized minor SGAs (Fig. 3-10). Furthermore, the conversion of α-tomatine to a solanidane-type SGA, demissine, which is known to be accumulated in leaves of wild potato *Solanum acaule*, was confirmed with keeping lycotetraose attached at the hydroxy group at C3 (Fig. 3-11). Furthermore, the addition of the DOX inhibitor prohexadione during the α-solamarine feeding assay caused the suppression of the conversion of α-solamarine to α-solanine in a dose-dependent manner, whereas uniconazole-P, a cytochrome P450 inhibitor, had little effect on the bioconversion (Fig. 3-7b). These data suggest that solanidane-type SGAs (i.e., α-solanine and α-chaconine) are biosynthesized from spirosolane-type compounds in potatoes and also that a DOX family enzyme is involved in the conversion reactions.

Selection of candidate genes

The DOX superfamily is one of the largest enzyme superfamily in plant kingdom and associated with the oxygenation reactions of various specialized metabolites[66]. Recently,

involvement of DOXs in SGA biosynthesis and modification was reported[11,12,14,24]. In our previous study, *St16DOX*, which showed the highest expression in potato tuber sprouts, has been identified to encode steroid 16α-hydroxylase from 256 DOX genes in the potato genome. In this study, we focused on the gene *Sotub01g007130*, which showed the second highest expression in the tuber sprouts among the potato DOX genes (Table 3-4), and the gene was designated as *StDOX130*. Quantitative reverse transcription PCR showed that the expression pattern of *StDOX130* was similar to those of the SGA biosynthetic genes *PGA1*, *PGA2*, *16DOX* [24] (Fig. 3-12). Additionally, we found that *Sotub01g007070*, *Sotub01g007100*, and *Sotub01g007110*, which are clustered in chromosome 1, were highly expressed in potato tuber sprouts (Table 3-4), and we named them as *StDOX070*, *StDOX100*, and *StDOX110*, respectively. Coding sequence of *StDOX130* and *StDOX070* were successfully amplified using the cDNAs synthesized from total RNA extracted from tuber sprouts of *S. tuberosum* cv Sassy. The deduced proteins of StDOX130 and StDOX070 consist of 275 and 311 amino acid residues, respectively, indicating that StDOX130 is 36 residues shorter at the C-terminus than StDOX070 (Fig. 3-13). StDOX130 shares 85.8% amino acid sequence identity to StDOX070 while StDOX130 shows 18.1% identity to St16DOX. Based on the phylogenetic classification of the plant DOX superfamily[66], StDOX130 and StDOX070 were found to belong to clade DOXC20, which includes α-tomatine 23-hydroxylase (Sl23DOX/GAME31) involved in the detoxification of tomato bitter glycoalkaloid[11,12].

Characterization of StDOX130-silenced potato plants.

To investigate the contribution of *StDOX130* to solanidane-type SGA biosynthesis, we generated *StDOX130*-silenced transgenic potato plants (i.e., *StDOX130i* lines) by using

RNA-interference (RNAi) vector containing a stem-loop partial fragment of *StDOX130* cDNA (Fig. 3-2). *StDOX130* expression levels in the leaves of in vitro grown *StDOX130i* plants were significantly reduced, and no off-target silencing effect on *StDOX070* was observed (Fig. 3-14a,b). The leaves of in vitro-grown *StDOX130i* plants showed the remarkable decrease of solanidane-type SGAs (α-solanine and α-chaconine) levels compared with that of not-transformed (NT) potato plants (Fig. 3-14c), suggesting that *StDOX130* is necessary for solanidane-type SGAs biosynthesis in potato. Analysis of endogenous metabolites accumulated in *StDOX130i* plants detected two new peaks, of which the retention time (Rt) and mass spectrum matched with those of the authentic compounds, α-solamarine and β-solamarine (Fig. 3-14d and Fig. 3-15). The amounts of α-solamarine and β-solamarine accumulated in *StDOX130i* plants were comparable to those of α-solanine and α-chaconine in NT potato plants, respectively (Fig. 3-14e). These results indicate that *StDOX130* is involved in the conversion from spirosolane-type SGAs to solanidane-type SGAs in potato.

StDXO130 catalyze solanidane-skeleton formation reaction.

Recombinant His-tagged protein of StDOX130 were prepared with a bacterial expression system in *Escherichia coli* and purified by cobalt-affinity chromatography (Fig. 3-16). Purified StDOX130 was assayed with several spirosolane-type compounds as substrates, and the reaction products were analyzed using LC-MS. StDOX130 metabolized α-solamarine to form an unknown product, of which the retention time and the mass fragment pattern was not identical to α-solanine (Fig. 3-17a,b). The product gave a parental ion mass at m/z 882.6 which is 2 mass smaller than that of the substrate α-solamarine (Fig. 3-17b). Tomatidenol, the aglycone of α-solamarine, was also

metabolized to an unknown product, of which the parent mass ion is 2 mass smaller than that of tomatidenol (Fig. 3-18). The other spirosolane-type compound, β-solamarine, α-tomatine, and tomatidine, were also accepted as substrates by StDOX130, and the reaction products were 2 mass smaller than the respective substrates (Fig. 3-18). On the other hand, StDOX130 was not active toward solanidane-type SGAs (α-solanine and α-chaconine), and the aglycone solanidine (Fig. 3-19). StDOX070 gave the results essentially same as StDOX130 did (Fig. 3-17a, Fig. 3-18 and Fig. 3-19).

To further identify the nature of reactions catalyzed by StDOX130, we performed high resolution mass spectrometry (HRMS) and NMR analysis of the StDOX130 enzymatic product obtained by reacting with α-tomatine, which is commercially available in an enough quantity for structural analysis. The structure of the StDOX130 enzymatic reaction product was determined as shown in (Fig. 3-5), in which solanidane-skeleton scaffold is formed and iminium ion structure is constructed between C-16, the nitrogen atom, and C-22. The reaction mechanism of DOX130 is illustrated in Fig. 3-17c. Thus, we confirm that both StDOX130 and StDOX070 convert spirosolane-skeleton to solanidane-skeleton.

StDXO130 display strong preference for glycosides.

To gain further insight into the substrate specificity of StDOX130 and StDOX070, we analyzed the kinetic parameters for α-solamarine and its aglycone, tomatidenol (Fig. 3-20 and Table 3-5), which are possible endogenous substrates in potatoes. The K_m value of StDOX130 for α-solamarine was 37-fold lower than that for tomatidenol, and the k_{cat} value for α-solamarine was 29-fold higher than that for tomatidenol, resulting that the catalytic efficiency (k_{cat}/K_m) for α-solamarine was 1,123-

fold greater than that for tomatidenol (Table 3-5). These results indicate that StDOX130 shows a very strong preference for the glycoside α-solamarine, suggesting that conversion of spirosolane-skeleton to solanidane-skeleton occurs after the glycosylation step in SGA biosynthesis in potatoes. The finding is consistent with the observation that α-tomatine was converted to its solanidane-type demissine with keeping a lycotetraose bound at the hydroxy group at C-3 (Fig. 3-11). The substrate preference of StDOX070 did not resemble to that of StDOX130 in terms of lower k_{cat} value for α-solamarine than that of for tomatidenol, giving only 4.5-fold higher k_{cat}/K_m than those detected with tomatidenol (Table 3-5). Comparing the two enzymes, the k_{cat}/K_m values of StDOX130 for α-solamarine and tomatidenol were 3.8-hold higher and 65-fold lower than those of StDOX070 (Table 3-5).

Evolution of SGA biosynthetic pathways in Solanum species

BLASTX search against Tomato ITAG protein database (https://solgenomics.net/) exhibited that Solyc01g006585, designated as SlDOX6585, showed the highest amino acid identity to StDOX130. Tomato *SlDOX6585* gene was almost not expressed in all tissues of tomato in RNA-seq data of the Tomato Functional Genomics Database (http://ted.bti.cornell.edu) (Table 3-6). Recombinant assay revealed that SlDOX6585 display solanidane formation activity like StDOX130 and StDOX070 (Fig. 3-21) but no activity with solanidane-type compounds was observed (Fig. 3-22). Very low or no expression of *SlDOX6585* (Table 3-6) explain the lack of solanidane-SGAs and accumulation of spirosolane-SGA (i.e., α-tomatine etc.) in tomato.

We found that *StDOX130* and *StDOX070* gene are located on chromosome 1, as a part of DOX-gene cluster which is spanning ~210 k bp genomic regions interposed by

50 k bp undetermined sequence gaps and contains 8 *DOX* genes (Fig. 3-23a). Among these clustered DOXs, deduced proteins of *StDOX090*, *StDOX100*, *StDOX120*, and *StDOX150* appeared truncated protein lacking canonical motif conserved among the DOX superfamily. Additionally, SlDOX6585 is embedded in gene cluster contains three DOX genes in tomato chromosome 1 (Fig. 3-23a). It is notable that genes flanking the clustered DOXs regions were highly conserved among potato and tomato, suggesting a common origin of these regions. Based on the phylogenetic classification of the plant DOX superfamily[66], all the clustered DOXs in potato and tomato were classified into DOXC20 clade which include Sl23DOX/GAME31 that encode spirosolane-23-hydroxylase[11,12]. In the phylogenetic analysis, StDOX130 and StDOX070 form a clade, designated as solanidane-forming clade, which is clearly separated from the clade including Sl23DOX/GAME31 (Fig. 3-23b). The clustered DOXs of potato were grouped into the SDF clade, excepting StDOX080 and StDOX090 which share a relatively close phylogenetic relationship with Sl23DOX/GAME31 (Fig. 3-23). No solanidane-forming activity of recombinant StDOX080 enzymes (data not shown) is consistent with the phylogenetic analysis. In tomatoes, only SlDOX6585 were found in SDF clade, while the other clustered DOXs were similar with Sl23DOX/GAME31 (Fig. 3-23a,b). These findings indicate that repeated duplication of ancestral SDF gene was occurred at this locus after speciation, and accumulation of multiple mutation gave rise highly active SDF, *StDOX130*, resulting in production of solanidane-type SGA in potato.

Discussion

The data presented here show that solanidane-type SGAs are biosynthesized from spirosolane-type compounds in potato and that StDOX130 is identified as a key enzyme

catalyzing the conversion of spirosolane-skeleton to solanidane-skeleton (Fig. 3-1). Strong preference of StDOX130 to glycosides observed in in vitro assay let us hypothesize that glycosylation occurs before aglycone, solanidine, formation is complete. Previously, SGT1 and SGT2 were reported to catalyze initial galactosylation and glucosylation of solanidine, respectively. These two glycosyltransferases also accept spirosolane-type substrates and the activities were comparable to activity with solanidine[26,27]. Additionally, silencing of *St16DOX* which catalyze 16α-hydroxylation of 22,26-dihydrxycholesterol resulted in accumulation of glycosides of 22,26-dihydrxycholesterol. These previous findings support our hypothesis that glycosylation occurs at spirosolane-skeleton or at more upstream intermediate. Additionally, we propose the following molecular mechanism to complete α-solanine and α-chaconine biosynthesis after ring rearrangement catalyzed by StDOX130. Putative imine reductase reduces the iminium moiety to furnish α-solanine or α-chaconine (Fig. 3-17c).

We demonstrate that duplication of gene belonging to solanidnae-forming clade led to production of solanidane-SGAs in potato. Several studies reported that novel specialized metabolic pathways are generated through gene duplication and subsequent functional divergence of, which includes altered expression patterns and protein functions. For example, three methylthioalkylmalate synthases (MAM1-3), products of serial tandem duplication, have different substrate specificity by which determine the side chain length variation of glucosinolates and different regulation of thse genes expression control the glucosinolates profile in *Arabidopsis thaliana*[67]. In evolutionary scenario of solanidane-SGA production in potato, regulatory neofunctionalization may have been crucial step. A recent report by Cardenas et al[68]. demonstrated that the transcription factor GAME9 comprehensively regulate SGA biosynthetic genes in potato and tomato. In the

same study, significant increase in expression level of *StDOX130* in *GAME9*-overexpressed potato was observed, while expression of SlDOX6585 was not affected by GAME9 in tomato[68]. These data indicate that StDOX130 have evolved to undergo the transcriptional regulation by GAME9 in potato. Change of protein structures may also be impotent factors. SDOX130 protein is characteristically shorter compared to typical DOXs, 36 residues shorter at the C-terminus than StDOX070 and SLDOX6585, respectively (Fig. 3-13). To confirm whether this short C-terminus contributes to the high specificity of StDOX130 to glycosides, functional analysis of the point-mutated enzyme and protein crystal structure analysis will be required.

The selection pressures that has resulted in solanidane-SGAs production also provides an interesting research question, considering that there is large variation in the biological activity of the different SGAs. Several wild potato cultivars contain both spirosolane-type and solanidane-type SGAs, however domesticated potatoes produce almost only solanidane-type SGAs. Therefore, solanidane-SGAs production pathway probably emerged after speciation of tomato and wild potato species, followed by domestication and breeding process seems to have dictated a convergence that resulted in the presence of almost only solanidane-type SGAs in cultivated potatoes. Roddick reported that solanidane-SGAs, compared to spirosolane-type SGAs, had stronger acetylcholinesterase inhibitory activity which is one of a major cause of SGA toxicity and several studies demonstrated the significance of E,F-ring structure to their biological activities. Possibly, the pathway by which metabolize spirosolane-SGAs to solanidane-SGAs initiated by StDOX130 in potato has resulted because of its evolutionary benefit, for example, by making SGAs more effective in protection against their enemies.

In conclusion, we have revealed genetic and enzymatic origin of solanidane-type SGAs in potato. We have demonstrated chemical evolution of SGAs and its evolution`s footprints on genome. Such our findings will provide insight into understanding how plant evolve and developed their specialized metabolic pathways and biological functions of metabolites.

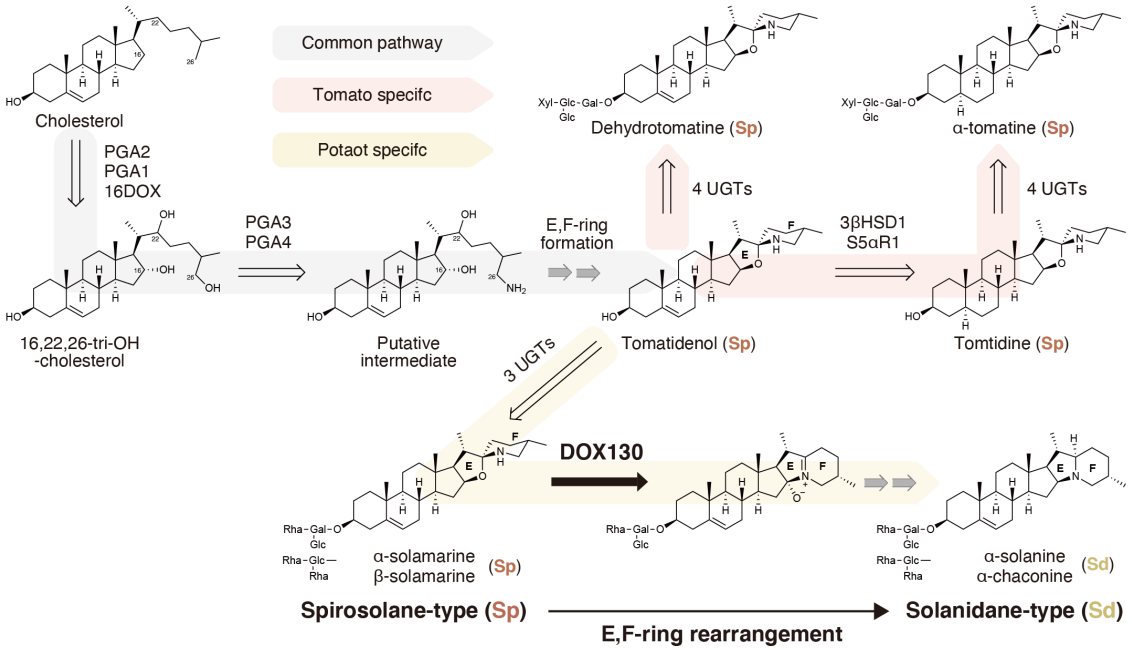
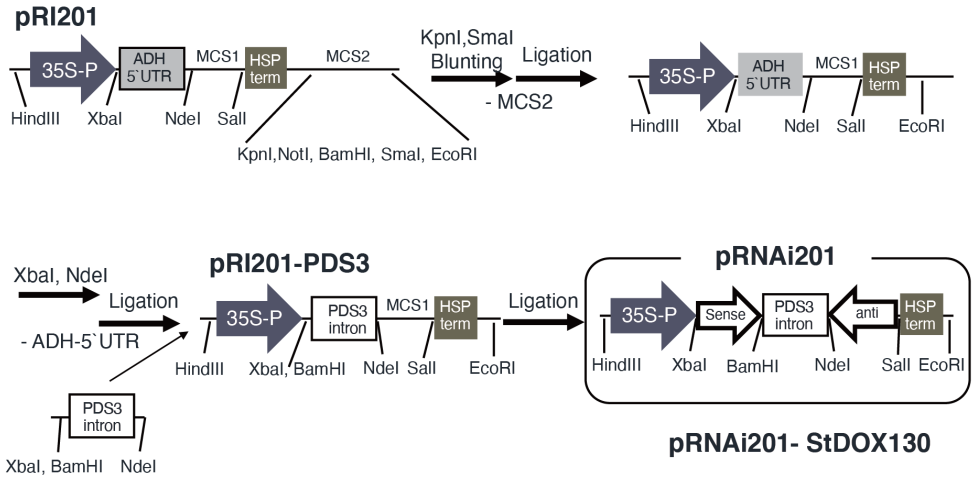


Fig. 3-1 | Biosynthesis of solanidane-type and spirosolane-type SGAs in potato and tomato. Common pathway between potato and tomato are shaded gay. Steps specific to tomato or potato are indicated by red or yellow shade. White and gray arrows represent characterized and uncharacterized reactions, respectively. Reaction step characterized in this work is shown in black arrow. PGA, POTATO GLYCOALKALOID BIOSYNTHESIS; DOX, 2-oxoglutarate dependent dioxygenase; UGT, udp-dependent glycosyltransferases; Glu, Glucose; Gal, Galactose; Rha, Rhamnose; Xyl, Xylose.



Supplementary Figure 3-2 | Construction scheme of *StDOX130*-RNAi vectors.

35-SP, Cauliflower mosaic virus 35S promoter; PDS3 intron, intron of the Arabidopsis phytoene desaturase gene (PDS, At4g14210); ADH 5`UTR, alcohol dehydrogenase gene-derived 5'-untranslated region; HSP term, Arabidopsis heat shock protein gene-derived terminator.

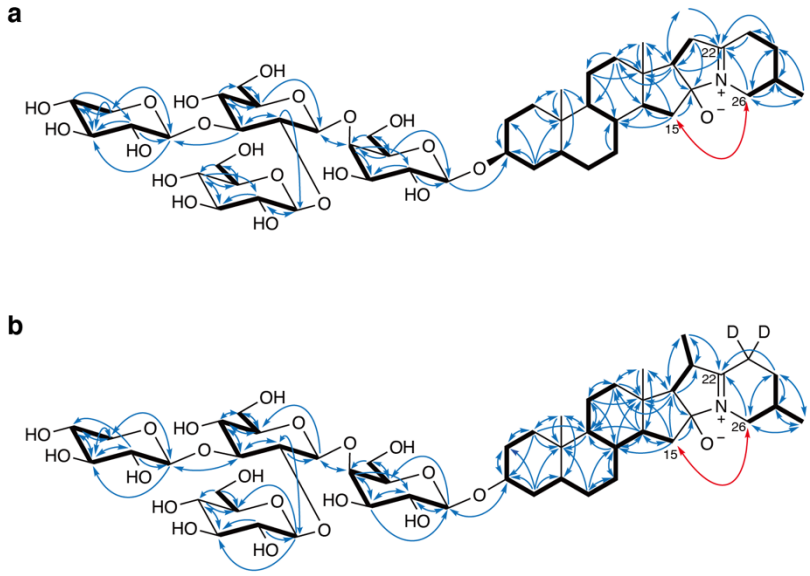


Fig. 3-3 | COSY (bold bond), HMBC (blue arrow), and key NOE (red arrow) correlation of the StDOX130 product in CD_3SOCD_3 (a) and CD_3OD (b). Configuration is omitted for clarity.

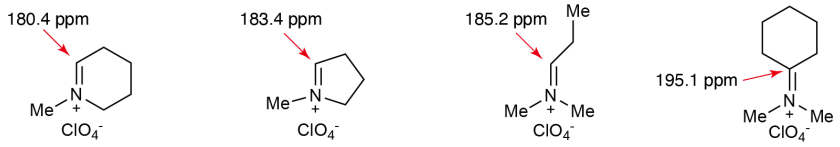


Fig. 3-4 | ^{13}C chemical shift of carbon atom of iminium cations (ppm, CD_3NO_2, TMS).

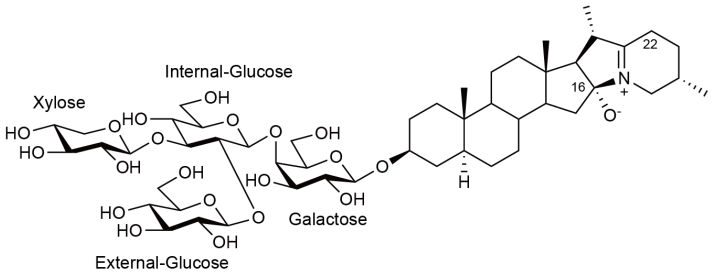


Fig. 3-5 | The structure of StDOX130 enzymatic reaction product confirmed by HRMS and NMR analyses.

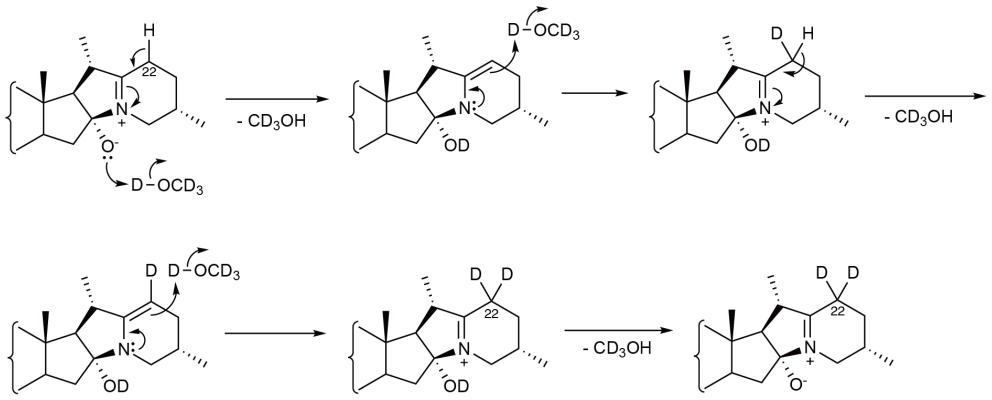


Fig. 3-6| Putative mechanism for H-D exchange of H-22 of the StDOX130 product in CD_3OD.

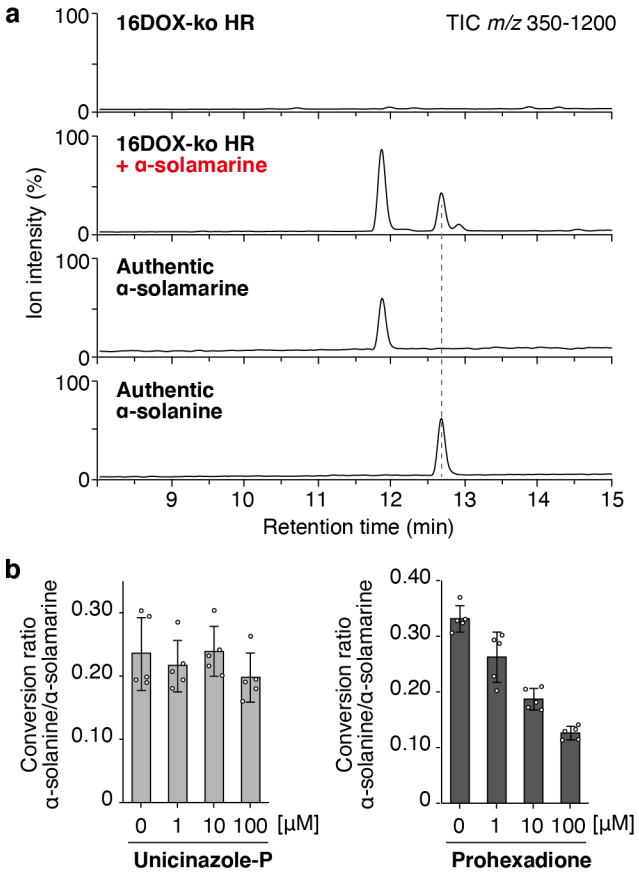


Fig. 3-7 | Bioconversion of solanidane-skeleton to spirosolane-skeleton in potato hairy root. a, Liquid chromatography–tandem mass spectrometry (LC-MS) detection of conversion from administered α-solamarine to α-solanine in *St16DOX*-disrupted potato hairy root (16DOX-ko HR), which contains no detectable SGAs. Mass spectrometry scan mode with a mass range of *m/z* 350 to 1,200 was used. Production of α-solanine was confirmed by comparison with authentic standard of α-solanine in terms of retention time and mass spectrum (Supplementary Fig. 2). **b,** Effect of Uniconazole-P or prohexadione on bioconversion of α-solamarine to α-solanine. After administering α-solamarine for 3 days in the presence of Uniconazole-P or prohexadione, SGAs accumulated in 16DOX-ko HR were extracted and quantified using LC-MS. Bars represent the ratio of α-solamarine to α-solanine accumulated in the hairy root. Error bars represent means ± s.d. (n = 5 independent experiments).

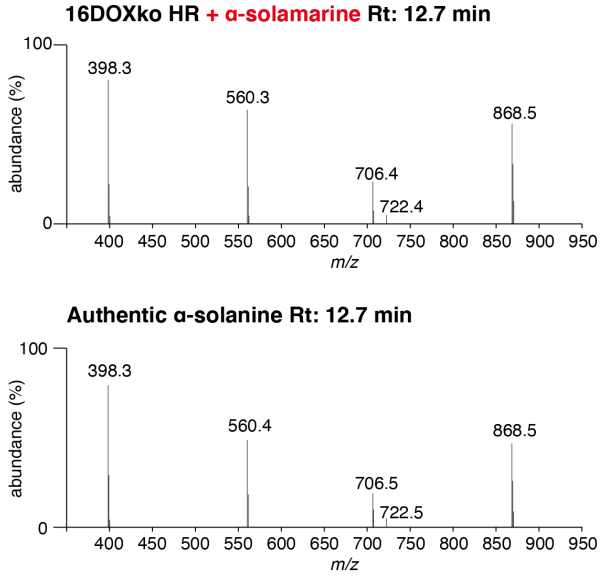


Fig. 3-8 | Comparison of ESI spectra from Fig. 3-7. The spectra verify the products are chemically identical despite their different origins. All spectra were obtained from the center of the peaks. Rt, retention time.

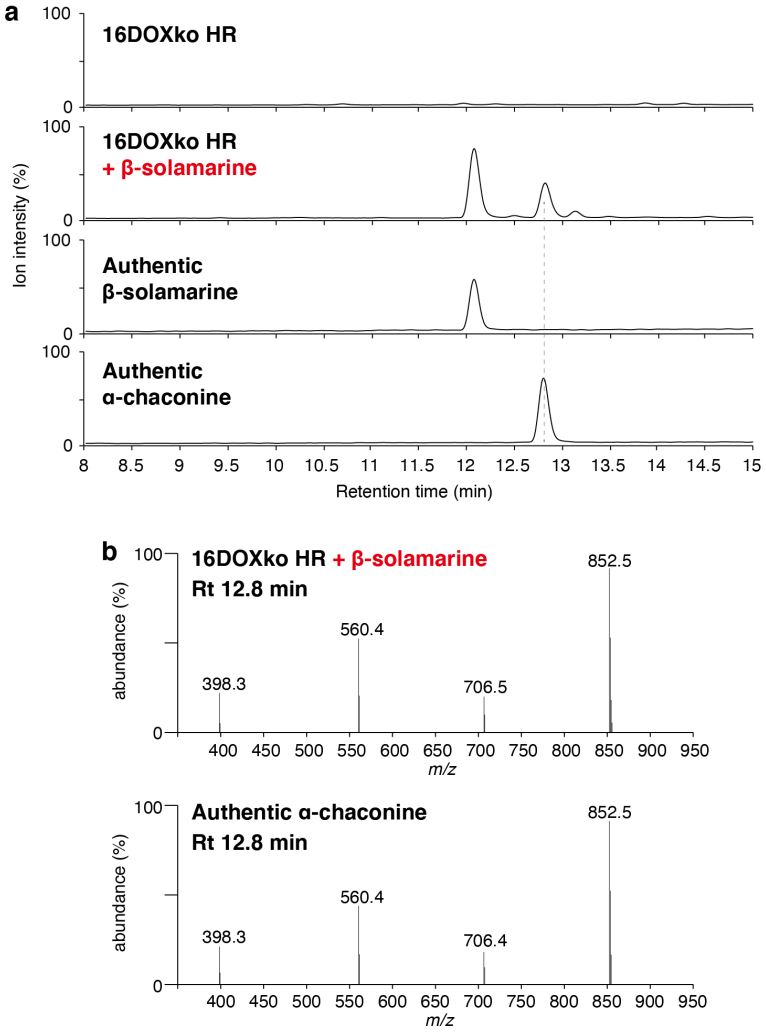


Fig. 3-9 | Bioconversion of β-solamarine to α-chaconine in potato hairy root. a, Liquid chromatography–tandem mass spectrometry (LC-MS) detection of conversion from administered β-solamarine to α-chaconine in *St16DOX*-disrupted potato hairy root (16DOX-ko HR), which contains no detectable SGAs. Mass spectrometry scan mode with a mass range of *m/z* 350 to 1,200 was used. **b,** Comparison of ESI spectra from (**a**). The spectra verify the products are chemically identical despite their different origins. All spectra were obtained from the center of the peaks. Rt, retention time.

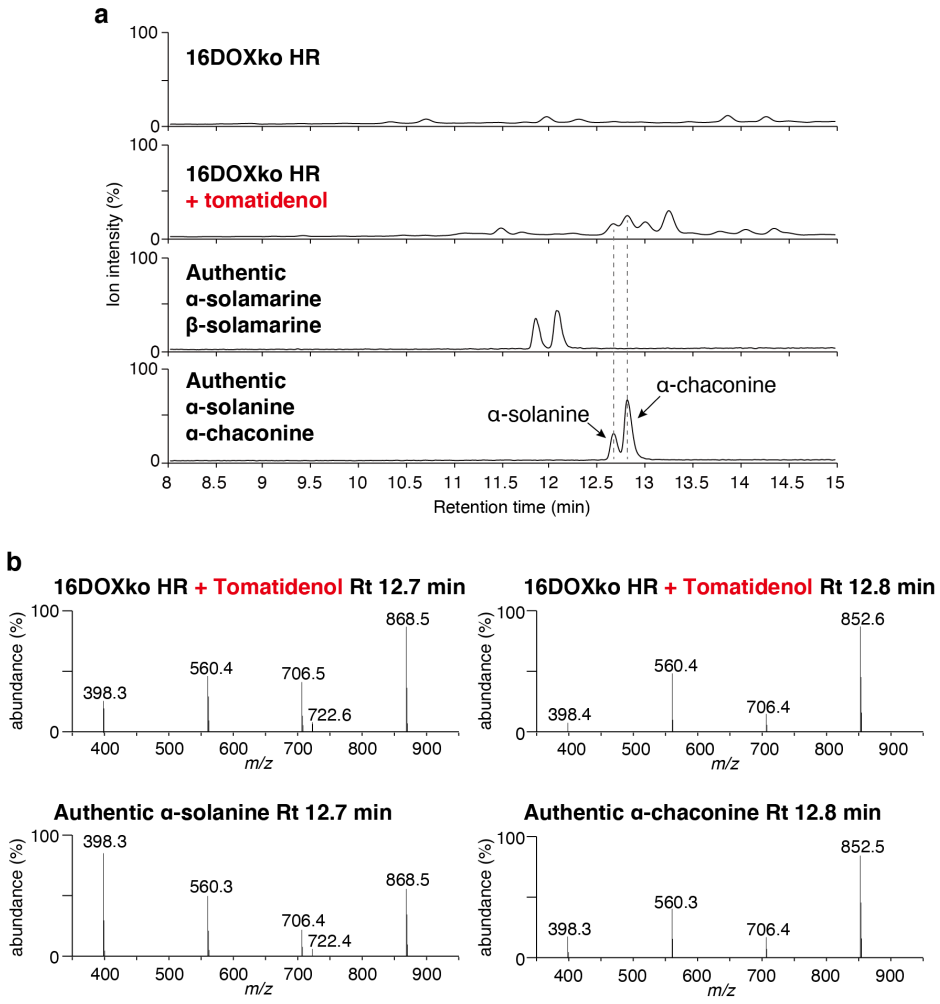


Fig. 3-10 | Bioconversion of tomatidenol in potato hairy root. a, Liquid chromatography–tandem mass spectrometry (LC-MS) detection of conversion from administered tomatidenol to α-solanine and α-chaconine in *St16DOX*-disrupted potato hairy root (16DOX-ko HR. Mass spectrometry scan mode with a mass range of *m/z* 350 to 1,200 was used. **b,** Comparison of ESI spectra from (**a**). The spectra verify the products are chemically identical despite their different origins. All spectra were obtained from the center of the peaks. Rt, retention time.

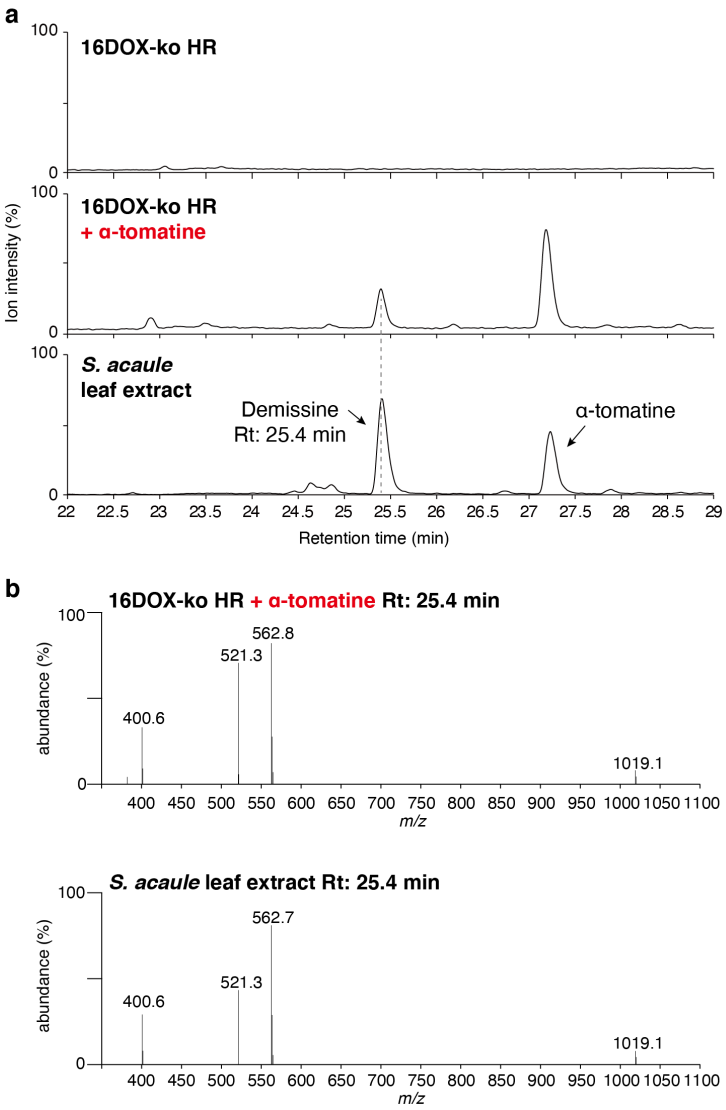


Fig. 3-11 | Bioconversion of α-tomatine to demmisine in potato hairy root. a, Liquid chromatography–tandem mass spectrometry (LC-MS) detection of conversion from administered α-tomatine to demissine in *St16DOX*-disrupted potato hairy root (16DOX-ko HR. Mass spectrometry scan mode with a mass range of *m/z* 350 to 1,200 was used. Production of demissine was confirmed by comparison with demissine extracted from leaves of *S. acaule* in terms of retention time and mass spectrum. **b,** Comparison of ESI spectra from (**a**). The spectra verify the products are chemically identical despite their different origins. All spectra were obtained from the center of the peaks. *S. acaule*, *Solanum acaule*; Rt, retention time.

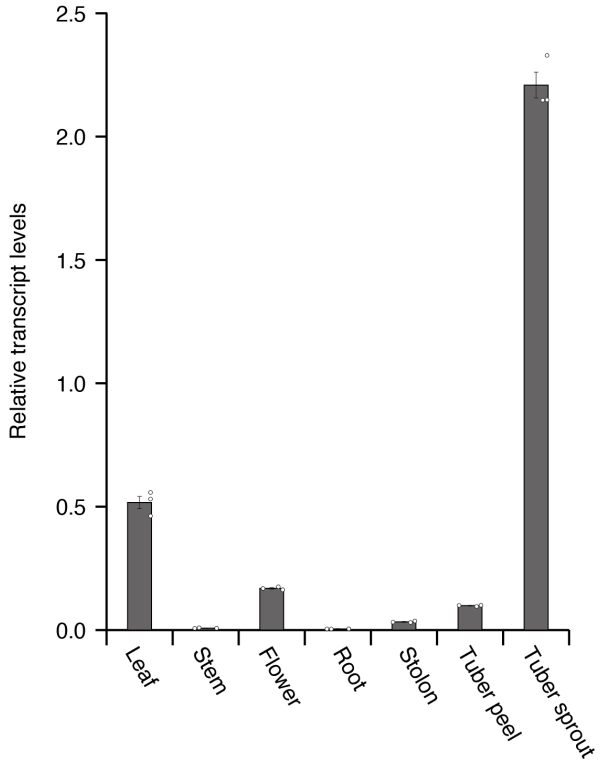


Fig. 3-12 | Quantitative RT-PCR analysis of the expression patterns of SGA biosynthetic genes in various organs of potato plants. Transcript levels of SGA biosynthetic genes are shown relative to that of *EF1α* as an internal reference gene. Error bars indicate s.d. (n = 3 independent experiments).

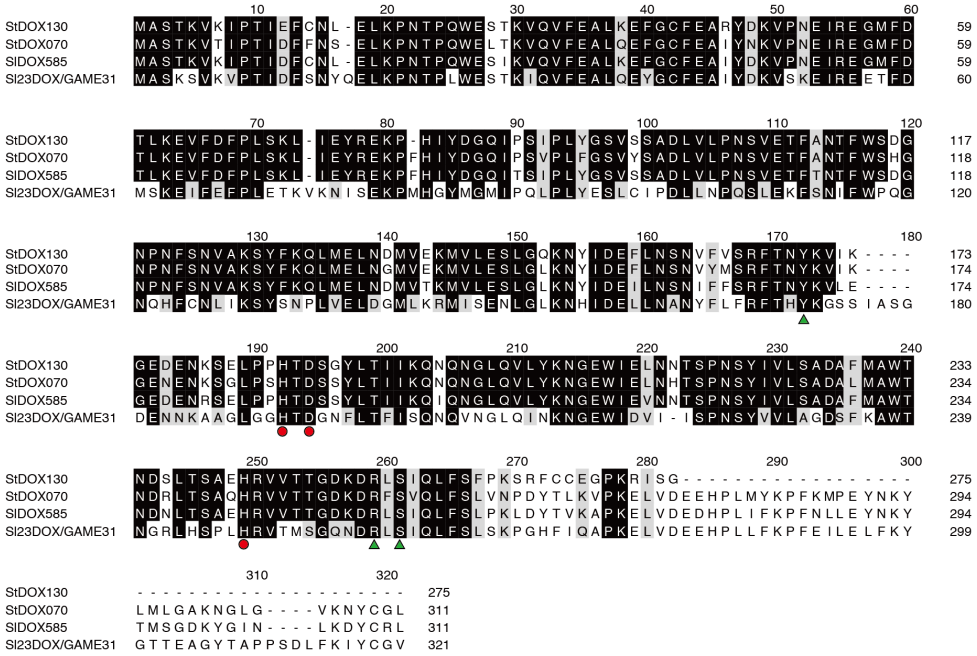


Fig 3-13 | Alignment of the amino acid sequences of StDOX130, StDOX070, SlDOX585 and Sl23DOX. Sequences were aligned using Clustal Omega. Shaded black boxes indicate residues that are identical in at least StDOX130, StDOX070 and SlDOX585. Similar amino acid residues are shaded in gray. Residues corresponding to a canonical HXDXnH catalytic triad (red circles) required fo r coordinating Fe(II) and a YXnRXS motif (green triangles) implicated in 2-oxoglutrate binding are indicated.

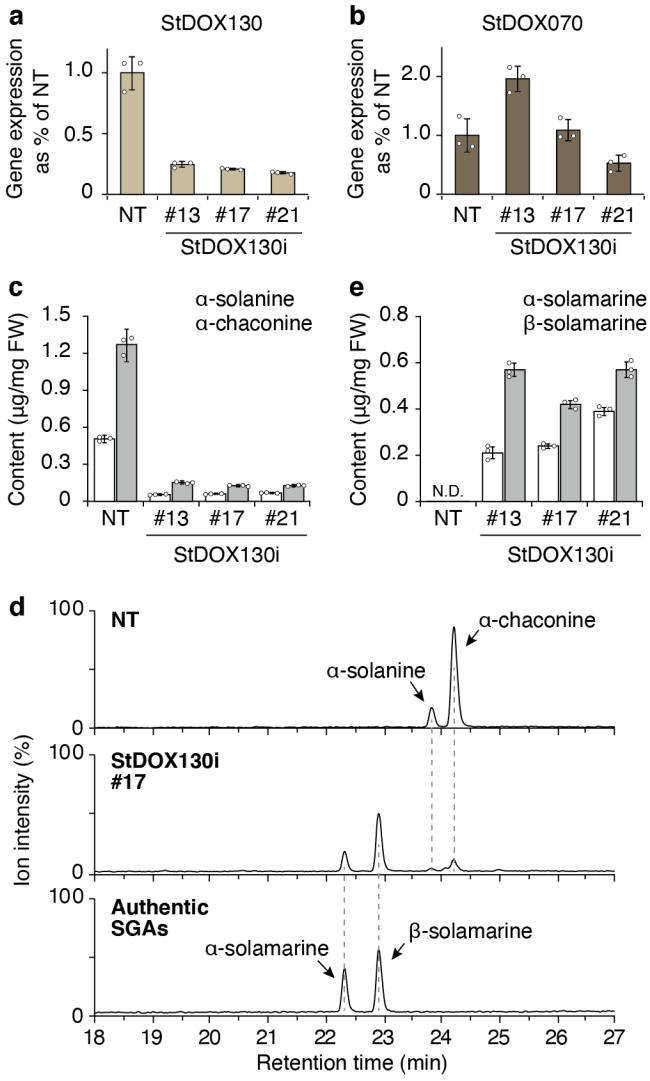


Fig. 3-14 | Silencing of *StDOX130i* potato plants. a,b, Real-time quantitative RT-PCR analysis of *StDOX130* (a) and *StDOX070* (b) gene transcript levels in leaf samples from *in vitro* grown shoot of non-transformed and *StDOX130*-silenced potato plants. **c,** Quantification of solanidane-type SGAs accumulated in non-transformed and *StDOX130*-silenced potato plants using LC-MS. **d,** SGA profiles in of non-transformed and *StDOX130*-silenced potato plants determined by LC-MS. Accumulation of α-solamarine and β-solamarine was confirmed by comparison with authentic standards in terms of retention time and mass spectrum (Fig. 3-15). **e,** Quantification of spirosolane-

type SGAs accumulated in non-transformed and StDOX130-silenced potato plants using LC-MS. All the SGAs analyzed in (**c**), (**d**) and (**e**) were extracted from leaf samples from *in vitro* grown shoot. Error bars indicate s.d. (n = 3 independent experiments). FW, Fresh weight; NT, non-transformed control plants; #13, #17, #21, independent transgenic lines.

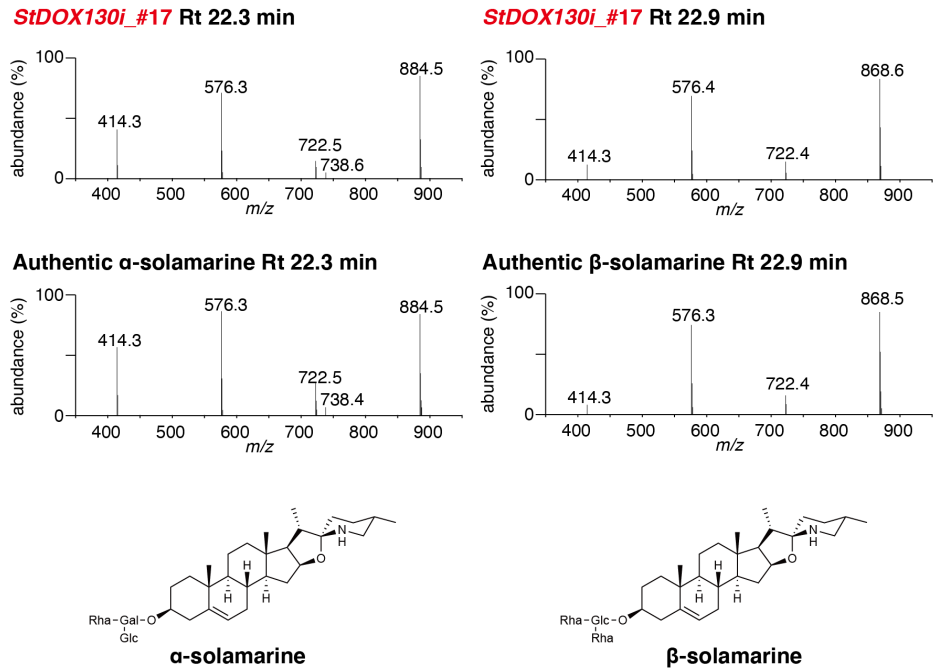


Fig. 3-15 | Comparison of ESI spectra from Fig. 3d. The spectra verify the products are chemically identical despite their different origins. All spectra were obtained from the center of the peaks. Rt, retention time.

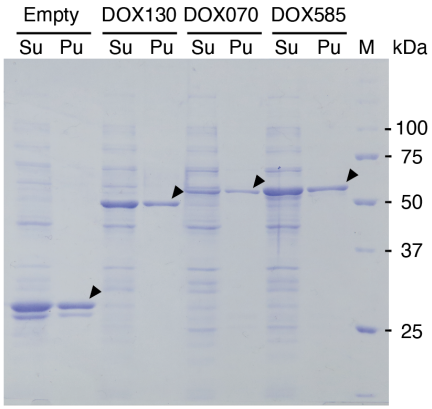


Fig. 3-16 | Heterogous expression of recombinant StDOX130, StDOX070 and SlDOX585 in E. coli BL21 (DE3) and purification using cobalt column. Expression and purification of the recombinant proteins were confirmed on SDS-PAGE analysis. M: Molecular weight protein marker, Su: Supernatant fraction, Pu: Purified protein fraction. Empty: Empty pCold ProS2 vector transformed in to BL21 (DE3) cells was used as negative control. Purified recombinant proteins are marked with black triangle.

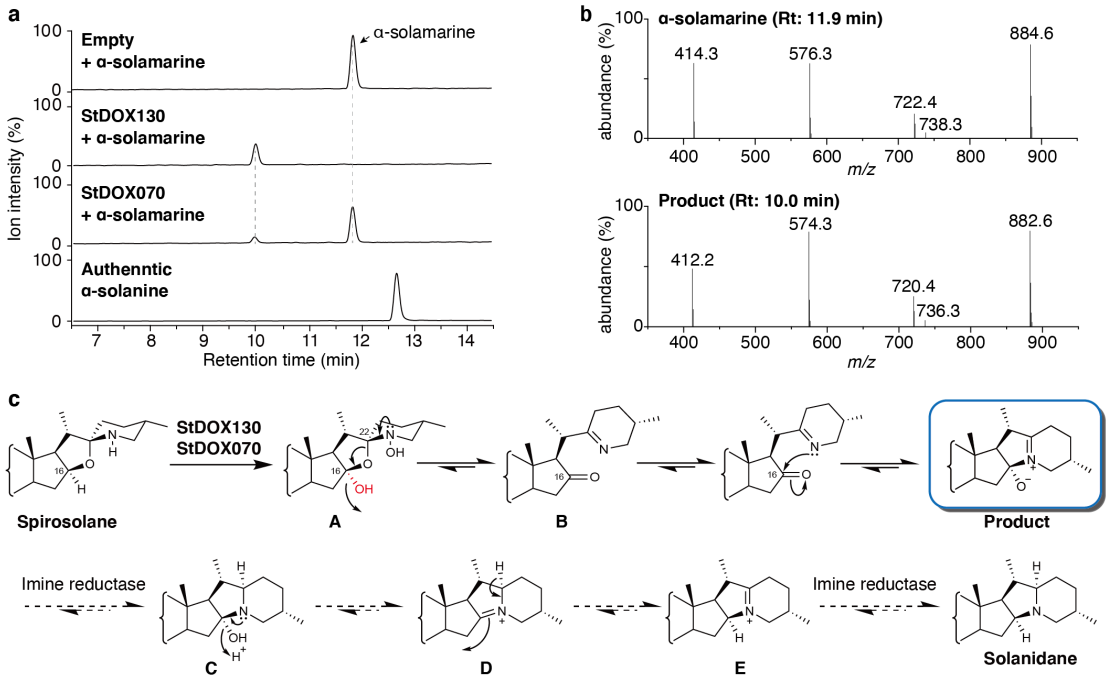


Fig. 3-17 | The activity of StDOX130 and StDOX070. a,b, LC-MS analysis of the reaction products from the recombinant StDOX130 and StDOX070 proteins produced in *E. coli* cells with α-solamarine as a substrate. (**a**) Substrate and product were monitored by mass spectrometry scan mode with a mass range of *m/z* 350 to 950. (**b**) Mass spectrum of the peaks shown in **a** at a retention time of 11.9 min (substrate, α-solamarine) and 10.0 min (reaction product). The negative control reaction was performed using purified protein fraction from *E. coli* cells transformed with an empty pCold ProS2 vector. **c,** Enzymatic reaction of StDOX130 and StDOX070, and putative subsequent reaction steps to complete solanidane-skeleton formation. The structure of the reaction product observed in the enzyme assay was elucidated using NMR analysis (Supplementary Fig. 3-5, Table 3-2 and Table 3-3). First, StDOX130 inserts an oxygen atom on C-16 of α-tomatine to form 16-hemiacetal intermediate **A**. Second, the lone pair on the nitrogen atom of **A** migrates between C-22 and the nitrogen atom. Concomitantly, the carbon-oxygen bond at C-22 cleaves and the newly formed hydroxy group at C-16 leaves as water to affords 16-oxo species **B**. Third, the nitrogen atom of **B** attacks the C-16 carbonyl

carbon to form the zwitterion, the StDOX130 reaction product. Next, putative imine reductase reduces the iminium moiety of the StDOX130 product to give hemiaminal **C**. The lone pair on the nitrogen atom of **C** pushes the hydroxy group at C-16 to leave as water to form iminium species **D**, which would isomerize to **E**. Finally, putative imine reductase again reduces iminium moiety of **E** to complete solanidane-skeleton formation.

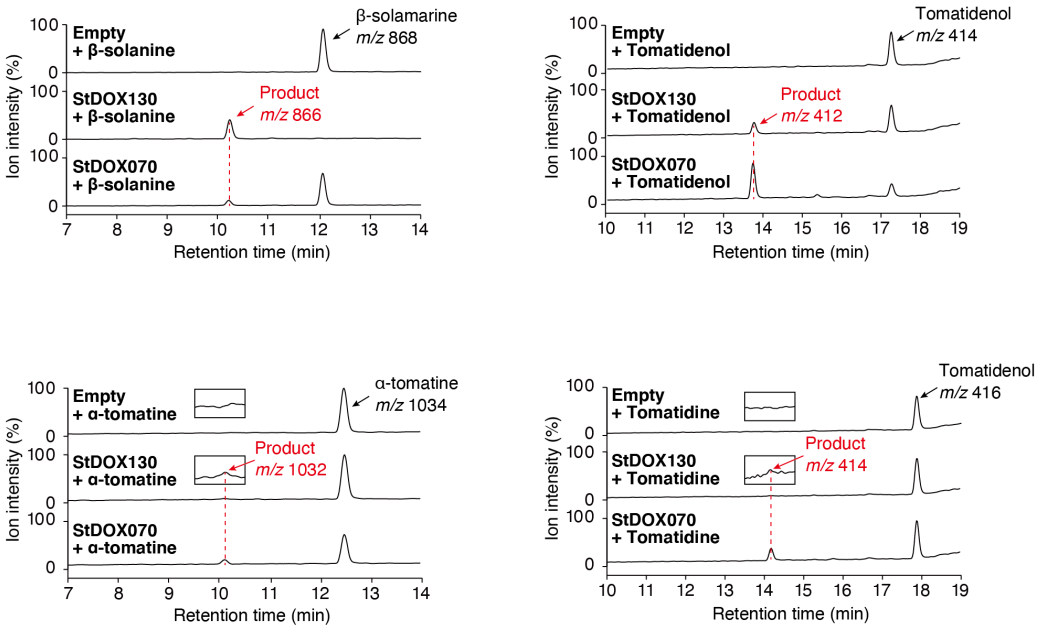


Fig. 3-18 | LC-MS analyses of the reaction products from the recombinant StDOX130 and StDOX070 proteins with various spirosolane-type substrates. Substrate and product were monitored by mass spectrometry scan mode with a mass range of *m/z* 350 to 1,200. The negative control reaction was performed using purified protein fraction from *E. coli* cells transformed with an empty pCold ProS2 vector. Mass to charge (*m/z*) derived from parental ion is shown for substrates and assay products.

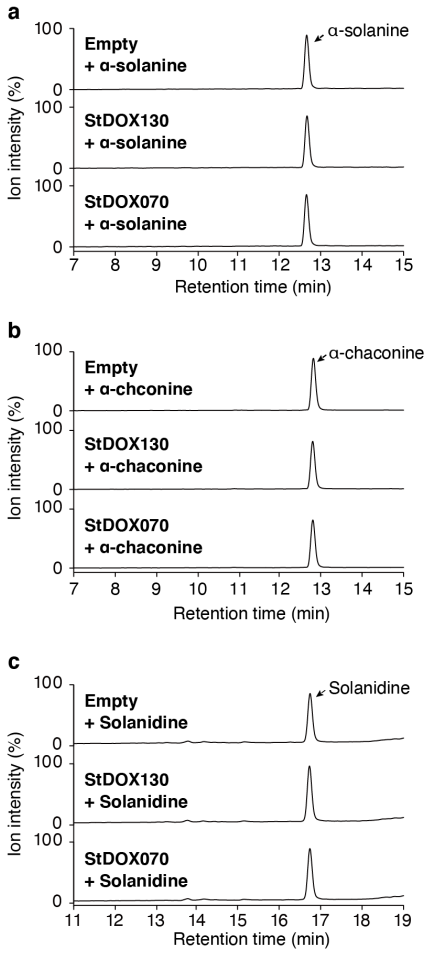


Fig. 3-19 | LC-MS analyses of the reaction products from the recombinant StDOX130 and StDOX070 proteins with solanidane-type substrates. Substrate and product were monitored by mass spectrometry scan mode with a mass range of *m/z* 350 to 950. The negative control reaction was performed using purified protein fraction from *E. coli* cells transformed with an empty pCold ProS2 vector.

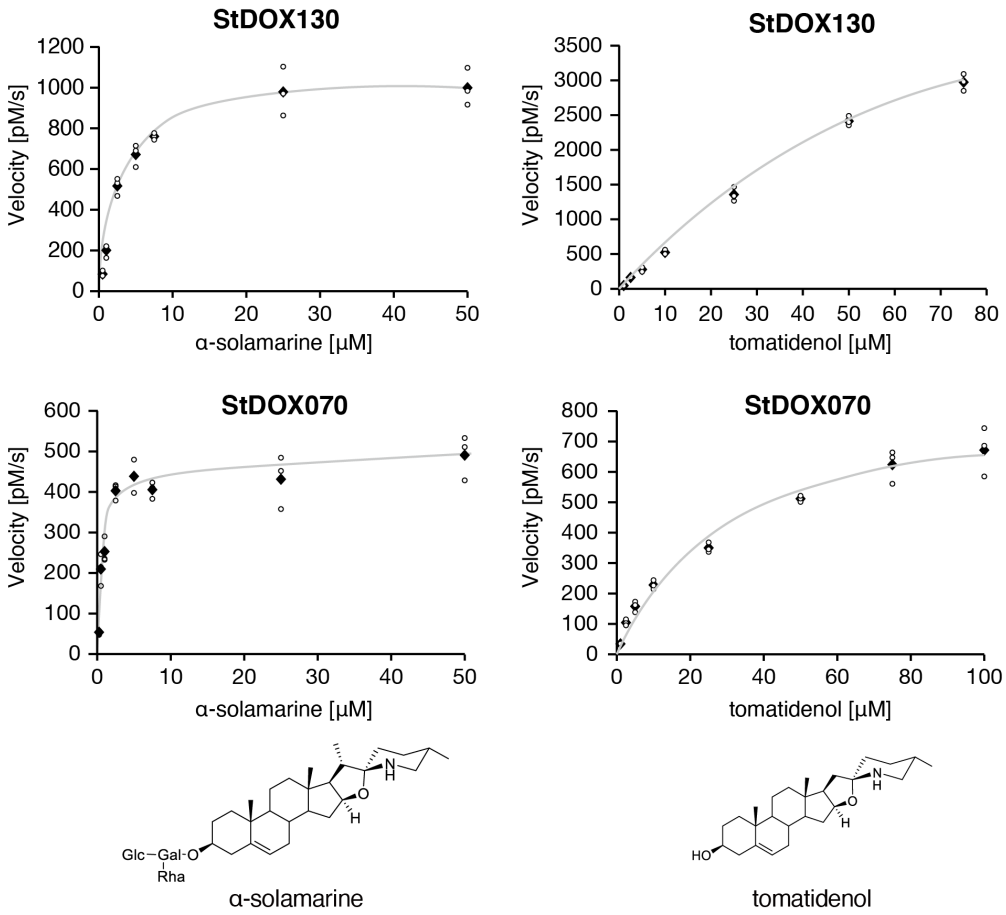


Fig. 3-20 | Enzyme kinetics of purified recombinant StDOX130 and StDOX070 with α-solamarine and tomatidenol. Purified recombinant enzyme of StDOX130 and StDOX070 were assayed using different concentration of α-solamarine (up to 50 μM) at a fixed concentration of 2-oxoglutrate (100 mM), or using different concentration of tomatidenol (up to 75 μM) at a fixed concentration of 2-oxoglutrate (100 mM). A Michalis-Menten curve was fitted to the values obtained and calculated kinetic parameters is shown in Table 3-5.

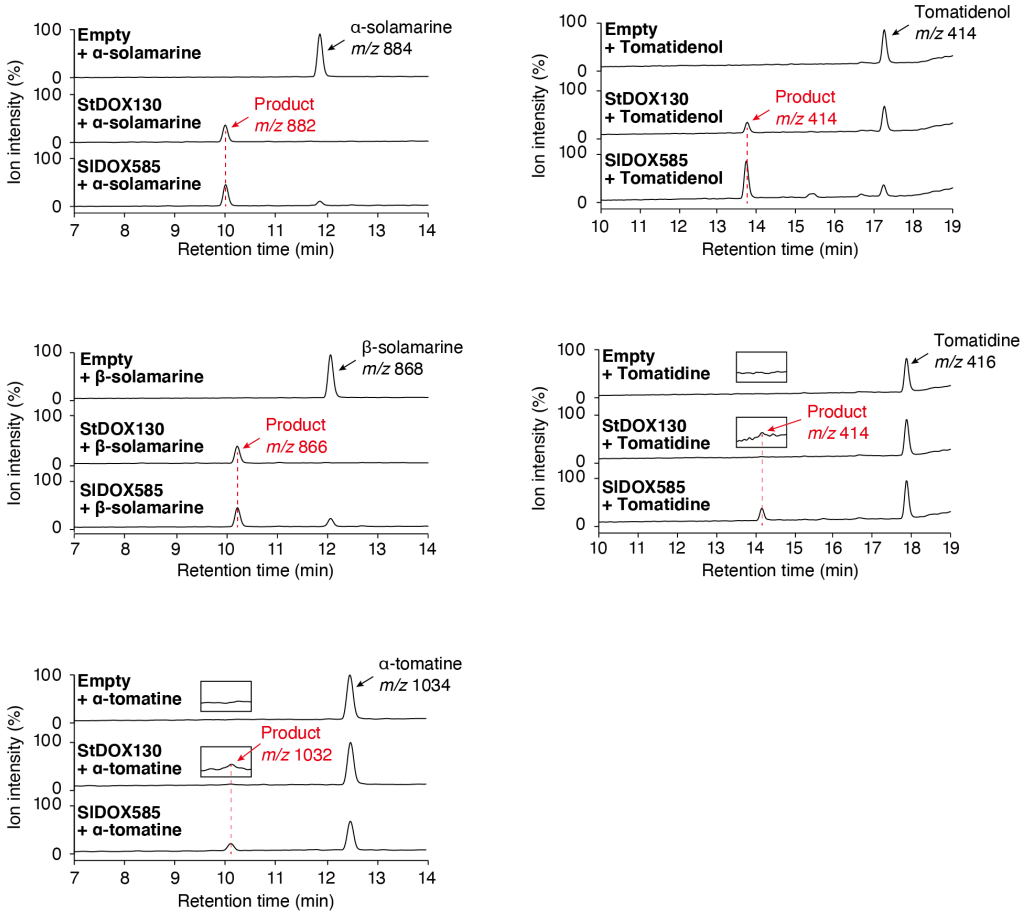


Fig. 3-21 | LC-MS analyses of the reaction products from the recombinant SlDOX585 proteins with various spirosolane-type substrates. Substrate and product were monitored by mass spectrometry scan mode with a mass range of *m/z* 350 to 1,200. The negative control reaction was performed using purified protein fraction from *E. coli* cells transformed with an empty pCold ProS2 vector. Mass to charge (*m/z*) derived from parental ion is shown for substrates and assay products.

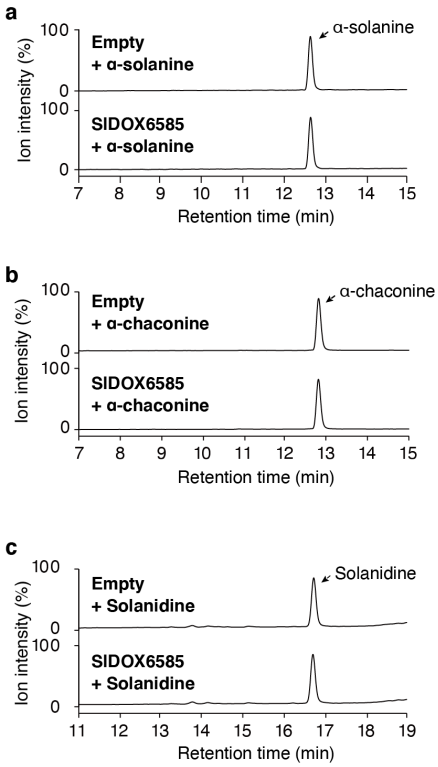


Fig. 3-22 | LC-MS analyses of the reaction products from the recombinant SlDOX585 proteins with solanidane-type substrates.

Substrate and product were monitored by mass spectrometry scan mode with a mass range of *m/z* 350 to 950. The negative control reaction was performed using purified protein fraction from *E. coli* cells transformed with an empty pCold ProS2 vector.

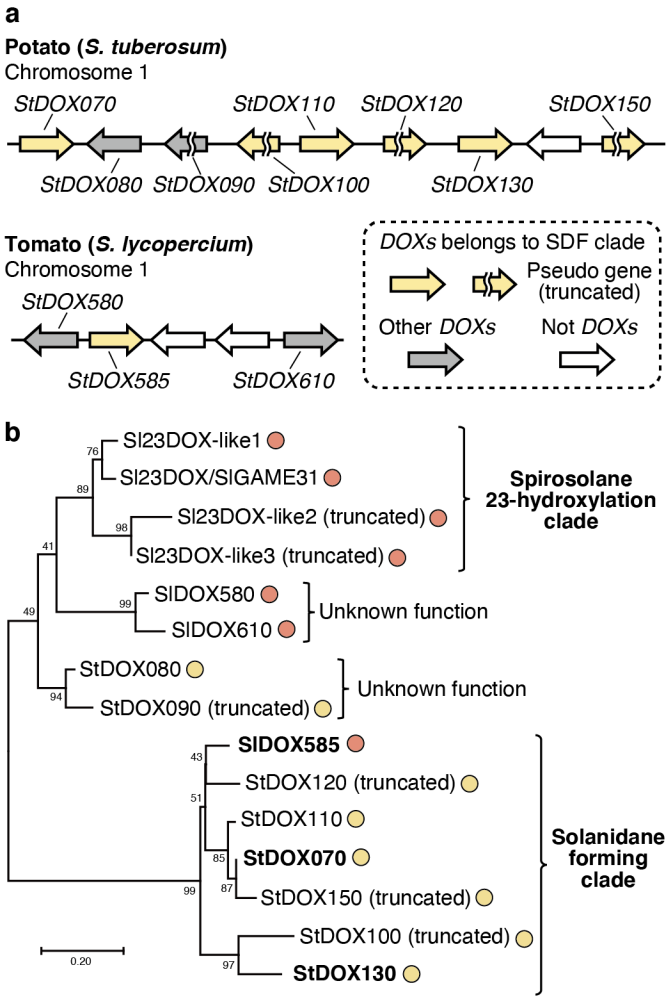


Fig. 3-23 | Evolution of DOX family genes associated with SGA metabolism in potato and tomato. a, Genomic organization of DOX genes in potato and tomato. Yellow arrows indicated DOX genes belonging to the same clade as StDOX130 in phylogenetic analysis. Gray arrows indicate DOX genes belonging to the other clades. White arrows indicated genes not belong to DOX superfamily. **b,** phylogenetic analysis of DOX genes shown in (**a**). Yellow and red dots indicated genes from potato and tomato, respectively. A phylogenetic tree was generated using the maximum likelihood method in MEGA7. Bootstrap values based on 1,000 replicates are shown at the branching points.

Table 3-1: PCR primers used in this work.

No.	primer name	sequence (5` to 3`)
1	StDOX130 qPCR Fw	GAGAGTGGATAGAGCTCAATAATACA
2	StDOX130 qPCR Rv	CTCCTGTTGTTACTACTCTGTGTTC
3	StDOX070 qPCR Fw	GATTACCTTCCCACACAGATAGTTC
4	StDOX070 qPCR Rv	CTATCATTTGTCCATGCCATAAG
5	StEf1α qPCR Fw	ATTGGAAACGGATATGCTCCA
6	StEf1α qPCR Rv	TCCTTACCTGAACGCCTGTCA
7	StDOX130 cloning Fw	CATATGGCATCTACCAAAGTTAAGATTCCC
8	StDOX130 cloning RV	GTCGACTTAAAGACCACAATAATTCTTGAGATTAATTCA
9	StDOX070 cloning Fw	CATATGGCATCTACCAAAGTTACGATTCCC
10	StDOX070 cloning RV	GTCGACTTAAAGACCACAATAATTCTTGACACCC
11	AtPDS third intron Fw	TCTAGATTTGGATCCAAAGGTACTTTGATTGGTCTC
12	AtPDS third intron Rv	GGATCCCAACACTATTTGGGAGGACCAA
13	StDOX130 fragment Fw	GTCGACTCTAGACCAAACACTCCACAATGGGA
14	StDOX130 fragment Rv	CATATGCTCGAGCAACACTATTTGGGAGGACCAA
15	NPT2 Fw	TAAAGCACGAGGAAGCGGT
16	NPT2 Rv	GCACAACAGACAATCGGCT

Table 3-2: ^{13}C and ^{1}H assignment of the aglycon of the DOX130 enzymatic reaction product. [a] TMS (0.00 ppm) was used as an internal standard. [b] Values in parentheses denote multiplicity and coupling constants (in Hz) for each signal. [c] Data after H-D exchange are presented. [d] Severely broadened signal due to C-D coupling. [e] Before H-D exchange, the corresponding signals were observed as following: 2.84 (br ddd, J = 21.3, 8.9, 6.8 Hz), 3.12 (1H, br ddd, J = 21.3, 4.1, 4.1).

solvent		CD_3SOCD_3		CD_3OD[c]	
		δ_C	δ_H	δ_C	δ_H
C/H		201 MHz	800 MHz	151 MHz	600 MHz
1	α	36.29	0.94 (m)	38.12	1.01 (m)
	β		1.64 (m)		1.73 (m)
2	α	28.90	1.74 (m)	30.40	1.88 (m)
	β		1.39 (m)		1.53 (m)
3		76.26	3.53 (dddd, 11.0, 11.0, 4.7, 4.7)	79.27	3.68 (dddd, 11.2, 11.2, 4.8, 4.8)
4	α	33.86	1.62 (m)	35.30	1.71 (m)
	β		1.17 (m)		1.32 (m)
5		43.81	1.07 (m)	45.92	1.13 (m)
6	α	28.07	1.28 (m)	29.72	1.34 (m)
	β		1.21 (m)		1.32 (m)
7	α	31.57	0.96 (m)	33.32	1.04 (m)
	β		1.60 (m)		1.73 (m)
8		33.96	1.43 (m)	35.82	1.60 (m)
9		53.14	0.74 (m)	55.33	0.79 (ddd, 11.4, 10.8, 3.9)
10		35.19	–	36.88	–
11	α	20.01	1.51 (m)	21.81	1.62 (m)
	β		1.23 (m)		1.35 (m)
12	α	37.83	1.24 (m)	40.00	1.37 (m)
	β		1.74 (m)		1.85 (m)
13		42.83	–	45.05	–
14		52.56	1.44 (m)	54.58	1.46 (m)
15	α	36.95	2.00 (dd, 13.1, 4.9)	38.38	2.21 (dd, 12.9, 5.5)
	β		1.83 (m)		1.84 (dd, 13.9, 12.9)
16		110.95	–	117.85	–
17		63.90	1.81 (br s)	58.74	2.16 (d, 2.1)
18		13.16	0.54 (s)	14.93	0.70 (s)
19		11.92	0.77 (s)	12.67	0.86 (s)
20		41.43	3.24 (m)	43.70	3.37 (m)
21		18.61	1.33 (d, 7.6)	17.76	1.45 (d, 7.6)
22		192.21	–	196.80	–
23	a	24.71	2.86 (br ddd, 21.4, 7.5, 7.5)	26.38[d]	–[e]
	b		3.08 (br ddd, 21.4, 5.4, 5.4)		–[e]
24	a	23.61	1.45 (m)	25.35	1.55 (dd, 13.8, 11.1)
	b		1.83 (m)		1.97 (ddd, 13.8, 3.4, 1.2)
25		25.55	2.09 (m)	27.94	2.10 (m)
26	a	47.80	3.28 (dd, 15.0, 8.7)	50.68	3.18 (m)
	b		3.78 (m)		3.81 (m)
27		16.98	0.99 (d, 6.7)	18.03	1.12 (d, 6.7)

Table 3-3: ^{13}C and ^{1}H assignment of the sugar moiety of the DOX130 enzymatic reaction product. TMS (0.00 ppm) was used as an internal standard. Values in parentheses denote multiplicity and coupling constants (in Hz) for each signal. [c] Data after H-D exchange are presented.

solvent		CD_3SOCD_3		CD_3OD[c]	
		δ_C	δ_H	δ_C	δ_H
C/H		201 MHz	800 MHz	151 MHz	600 MHz
Gal	1	100.86	4.21 (d, 7.7)	102.69	4.37 (d, 7.7)
	2	71.39	3.20 (dd, 9.4, 7.7)	73.20	3.61 (dd, 9.8, 7.7)
	3	73.46	3.33 (m)	75.65	3.50 (dd, 9.8, 3.3)
	4	78.91	3.78 (m)	80.21	4.02 (br d, 3.3)
	5	73.74	3.34 (dd, 6.6, 5.7)	75.35	3.50 (m)
	6	59.39	3.39 (dd, 10.6, 5.7)	61.10	3.61 (m)
			3.73 (dd, 10.6, 6.6)		3.91 (m)
Internal-Glu	1	103.23	4.41 (d, 7.7)	104.76	4.58 (d, 7.8)
	2	79.27	3.57 (dd, 8.9, 7.7)	81.09	3.75 (dd, 8.9, 7.8)
	3	84.89	3.60 (dd, 8.9, 8.9)	87.93	3.70 (dd, 8.9, 8.6)
	4	68.71	3.12 (dd, 8.9, 8.9)	70.49	3.27 (m)
	5	75.81	3.23 (m)	77.53	3.33 (m)
	6	61.23	3.37 (dd, 10.7, 7.6)	63.16	3.57 (dd, 11.6, 7.5)
			3.71 (dd, 10.7, 10.7)		3.90 (m)
External-Glu	1	102.32	4.71 (d, 8.1)	104.32	4.91 (d, 8.1)
	2	74.20	2.97 (dd, 8.1, 8.1)	75.91	3.18 (dd, 8.9, 8.1)
	3	75.98	3.15 (m)	78.01	3.35 (m)
	4	69.58	3.14 (m)	71.59	3.34 (m)
	5	76.88	3.14 (m)	78.52	3.32 (m)
	6	60.80	3.56 (m)	62.74	3.81 (dd, 12.3, 5.5)
			3.72 (m)		3.88 (br d, 12.3)
Xyl	1	103.23	4.50 (d, 7.8)	104.99	4.60 (d, 7.7)
	2	73.42	3.04 (dd, 8.9, 7.8)	75.29	3.24 (dd, 9.3, 7.7)
	3	76.46	3.12 (dd, 8.9, 8.9)	78.35	3.30 (m)
	4	69.32	3.32 (m)	71.01	3.51 (m)
	5	65.79	3.10 (dd, 11.0, 8.9)	67.23	3.26 (m)
			3.77 (dd, 11.0, 5.2)		3.91 (m)

Table 3-4: The list of *DOX* genes highly expressed in tuber sprout. The list include top 30 genes with high expression level in tuber sprout. Expression levels are shown in FPKM value

Gene ID	Transcript ID	RH Stamen	RH Tuber Pith	RH Tuber Peel	**RH tuber sprout**	RH Tuber Cortex	RH Flower	RH Leaf	RH Petiole	RH Shoot Apex	RH Stem	RH Stolon	RH Young tuber	RH Mature Tuber	RH Root
Sotub07g016570 (St16DOX)	PGSC0003DMT400030676	1	1	80	**843**	4	339	1954	10	1298	301	2214	2279	109	14
Sotub01g007130 (StDOX130)	PGSC0003DMT400008936	0	0	17	**88**	1	3	22	0	99	3	39	217	7	3
Sotub02g031570	PGSC0003DMT400064315	79	39	58	**51**	36	90	30	42	22	19	23	23	20	67
Sotub01g007110 (StDOX110)	PGSC0003DMT400008930	0	0	7	**49**	0	0	16	0	52	4	27	107	4	2
Sotub01g007100 (StDOX100)	PGSC0003DMT400082027	0	0	7	**49**	0	1	7	0	59	1	17	102	3	2
Sotub01g007070 (StDOX070)	PGSC0003DMT400081912	0	0	6	**48**	0	0	9	0	51	2	16	78	5	1
Sotub09g026180	PGSC0003DMT400044419	1	8	19	**39**	5	15	21	33	31	30	43	57	6	22
Sotub06g018240	PGSC0003DMT400079111	9	63	70	**38**	65	17	17	45	74	32	27	43	79	52
Sotub04g026810	PGSC0003DMT400045704	8	12	14	**36**	16	33	12	19	145	38	73	124	43	17
Sotub03g017050	PGSC0003DMT400001512	1	0	16	**36**	1	4	16	100	3	335	27	10	8	30
Sotub03g029700	PGSC0003DMT400001594	1	5	4	**36**	5	3	7	1	11	4	4	14	8	21
Sotub02g024300	PGSC0003DMT400043087	5	1	0	**35**	2	61	36	51	20	15	27	49	5	90
Sotub03g029710	PGSC0003DMT400001611	0	3	1	**33**	3	1	3	1	7	2	1	11	5	22
Sotub02g031540	PGSC0003DMT400064257	10	23	31	**32**	21	18	13	22	23	13	24	22	22	38
Sotub02g011530	PGSC0003DMT400033881	9	28	27	**31**	23	11	12	16	26	17	26	20	13	32
Sotub01g042800	PGSC0003DMT400083733	0	59	56	**31**	48	8	15	26	7	22	3	3	25	71
Sotub04g007350	PGSC0003DMT400075876	17	18	16	**31**	21	21	17	24	39	28	26	40	24	17
Sotub03g029750 or Sotub03g029760	PGSC0003DMT400084684	0	5	3	**31**	6	5	8	1	12	5	3	16	7	25
Sotub04g007790	PGSC0003DMT400075915	43	10	17	**31**	14	49	56	41	7	16	14	5	11	41
Sotub03g029750	PGSC0003DMT400001604	1	5	3	**30**	5	5	10	1	14	4	3	14	7	23
Sotub02g027300	PGSC0003DMT400009156	6	23	25	**30**	25	8	10	17	14	14	12	15	15	37
Sotub11g028800	PGSC0003DMT400070461	0	1	2	**29**	0	7	29	15	6	38	21	6	0	1
Sotub05g009120	PGSC0003DMT400078389	22	17	14	**29**	18	26	21	32	23	24	24	24	17	29
Sotub02g017740	PGSC0003DMT400007992	1	0	22	**28**	1	12	41	7	5	1	18	6	4	33
Sotub01g022480	PGSC0003DMT400081749	15	33	26	**28**	27	18	11	26	32	33	15	44	20	29
Sotub03g019160	PGSC0003DMT400056112	28	10	14	**27**	13	22	16	25	21	20	53	21	12	34
Sotub02g031170	PGSC0003DMT400064198	1	23	17	**27**	24	3	5	4	37	13	9	17	29	9
Sotub09g021590	PGSC0003DMT400009727	10	5	8	**25**	10	21	19	19	8	16	9	17	10	4
Sotub11g028790	PGSC0003DMT400070465	1	4	4	**23**	2	18	40	33	10	51	27	15	6	4
Sotub02g008310	PGSC0003DMT400055332	29	0	5	**22**	1	248	472	163	2	18	142	30	0	1

Table 3-5: Kinetic parameters of StDOX130 and StDOX070 with α-solamarine and tomatidenol. Kinetic parameters were determined by non-liner regression with ANEMONA[61]

Enzyme	Substrate	K_m (µM)	k_{cat} (s^{-1})	k_{cat}/K_m (μM^{-1} s^{-1})
StDOX130	α-solamarine	4.6 ± 0.19	0.67 ± 0.08	0.15
StDOX070	α-solamarine	1.7 ± 0.24	0.064 ± 0.011	0.038
StDOX130	tomatidenol	170 ± 0.34	0.023 ± 0.013	0.00013
StDOX070	tomatidenol	19.6 ± 0.056	0.17 ± 4.5E-06	0.0085

Table 3-6: RPKM values of *SlDOX6585*, *Sl16DOX* and *Sl23DOX/GAME31*

Gene ID	Unopened flower bud	Fully opened flower	Leaf	Root	1 cm fruit	2 cm fruit	3 cm fruit	Mature green fruit	Breaker fruit	Breaker fruit + 10 days
Solyc01g006585 (SlDOX6585)	0	0	0	0	0	0	0	0	2	0.3
Solyc07g043420 (Sl16DOX)	2474	983	1280	441	5013	1405	329	49	8	0
Solyc02g062460 (Sl23DOX/SlGAME31)	74	27	1	10	14	15	10	117	186	568

Concluding Discussion

Steroidal glycoalkaloids (SGAs) are well known as antinutritional and toxic substance in Solanaceae vegetable crops such as potato, tomato and eggplant. So, controlling and reduction of SGA content is important for breeding of these crops. While toxic to human, SGA are beneficial for plant species that produce them. SGAs play protective role against wide range pathogens and predators, such as bacteria, fungi, viruses, insects and animals. SGAs are known for their enormous structural diversity, and their chemical structure determine their biological activities. This study was designed and implemented to get insight into strategy for breeding of SGA-free potatoes and to elucidate how structural diversity of SGAs is generated in *Solanum* species.

In chapter 1, CRISPR/Cas9-mediated genome editing of St16DOX which is involved in SGA biosynthesis was performed through *A. rhizigenes* mediated hairy root transformation. Complete disruption of *St16DOX* in potato hairy root was achieved and *St16DOX*-knokout hairy roots showed non-detectable accumulation of SGAs, indicating *St16DOX* is good target for breeding of SGA-free potato. Other SGA biosynthesis genes could also be good targets for producing SGA-free potatoes. PGA3/GAME4 was identified as SGA biosynthetic gene and silencing of this gene in potato resulted in remarkable reduction of SGA levels and accumulation of corresponding amounts steroidal saponins; protoneodioscin and 25-*epi*-indioside D[14,29], which are valuable saponins used for semi-synthetic production of pharmaceutical steroidal drugs such as anti-inflammatory, androgenic, estrogenic and contraceptive drugs. Thus, through genome editing of PGA3/GAME4, generation of potatoes that not only do not produce SGA but also produce the materials for the steroidal drugs.

Chapter 2 and chapter3 describe about identification of SGA biosynthetic genes that to structural diversity in *Solanum* plant species. Chapter 2 showed that SlS5αR2 catalyze steroid 5α-reduction and is involved in elimination of double bond from unsaturated SGAs in tomato. The presence or absence of the double bond at the C-5,6 position (dehydro-type and dihydro-type, respectively) is one of the sources for vast structural diversity in SGAs and determines the degree of SGA toxicity. Tomato contains both of dihydro-type SGA (α-tomatine) and dehydro-type SGA (dehydrotomatine). Prior findings suggest that dehydrotomatine is likely a more toxic SGA compared with α-tomatine. So, the reaction, of the C-5,6 double bond removal in SGA aglycones (tomatidine), is a key branch point that not only determines the diversity of SGAs in *Solanum* species but also modulates the toxic effects of this metabolite class to the plant.

Potato produces solanidane-type SGAs while tomato contains spirosolane-type SGAs. In chapter 3, I demonstrated that potato metabolizes spirosolane-type SGAs to produce solanidane-type SGAs (e.g. α-solanine and α-chaconine), and StDOX130 originated through repeated tandem gene duplication is key enzyme for this metabolism. Metabolism of spirosolane-type SGA is also found in tomato[17]. Sl23DOX/GAME31 catalyze the 23-hydroxylation of spirosolane-type SGAs, which is the first step of α-tomatine detoxification metabolic reactions in tomato fruit, and this metabolism seems to be key process in propagation and domestication of tomato[11,12]. Sl23DOX/GAME31 and StDOX130 are classified into DOXC20 subfamily, indicating these geneses arose from common ancestral gene. Interestingly, Sl23DOX/GAME31 is also tandemly located with three closely related DOX genes[11]. Thus, duplication and functional divergence of this class of DOX drive the evolution of structure and biological activity of SGA. Analyses of this class of DOX may lead to elucidation of yet uncovered nature of SGA biosynthesis

and metabolism. Furthermore, our study provide insight into understanding how plant evolve and developed their specialized metabolic pathways and biological functions of metabolites.

Acknowledgement

本研究を進めるにあたり，終始懇切なるご指導とご鞭撻賜り，本論文をまとめるに際して，親身なご助言をいただきました，水谷正治 准教授（神戸大学大学院農学研究科)に心より感謝を申し上げます.

本研究および本論文作成の過程における議論・検討にあたって多大なるご教示ならびにご激励を賜りました杉本幸裕 教授（神戸大学大学院農学研究科)，ならびに本論文の審査において有益な議論と情報交換をしていただきました宇野知秀 教授（神戸大学大学院農学研究科）に深く御礼申し上げます．また，本研究について貴重なご助言をいただきました山内靖雄 准教授（神戸大学大学院農学研究科）に深謝申し上げます.

これまでの研究において，様々なご支援をいただきました斉藤和季 先生（理化学研究所 環境資源科学研究センター)，村中俊哉 教授（大阪大学大学院工学研究科)，梅基直行 博士（理化学研究所 環境資源科学研究センター）に御礼申し上げます．ゲノム編集ベクターのご恵与ならびにゲノム編集技術についてご教示いただきました，刑部敬史 教授（徳島大学生物資源産業学部)，刑部祐里子 准教授（徳島大学生物資源産業学部)に深く感謝いたします．NMR 構造解析等を行っていただきました渡辺文太 助教授（京都大学化学研究所）に御礼申し上げます.

植物機能化学研究室のみなさまには，大変お世話になりました．特に，共に研究を進めてきました，中安大 博士（京都大学生存圏研究所)，李榮宰 博士（神戸大学農学研究科)，加藤純平 君（神戸大学農学研究科）には多くの示唆と刺激をいただき，感謝の念に堪えません．お一人お一人のお名前を挙げることができませんが，それぞれのお立場から研究から私生活においてまで様々な形でお力添えをいただきました．本当にありがとうございました.

Regerences

1. Harrison, D. M. Steroidal alkaloids. *Nat. Prod. Rep.* **7**, 139 (1990).
2. Petersen, H. W., Mølgaard, P., Nyman, U. & Olsen, C. E. Chemotaxonomy of the tuber-bearing Solanum species, subsection Potatoe (Solanaceae). *Biochem. Syst. Ecol.* **21**, 629–644 (1993).
3. Heftmann, E. Biogenesis of steroids in solanaceae. *Phytochemistry* **22**, 1843–1860 (1983).
4. Friedman†, M. Potato Glycoalkaloids and Metabolites: Roles in the Plant and in the Diet. (2006). doi:10.1021/JF061471T
5. Friedman†, M. Tomato Glycoalkaloids: Role in the Plant and in the Diet. (2002). doi:10.1021/JF020560C
6. Milner, S. E. *et al.* Bioactivities of glycoalkaloids and their aglycones from solanum species. *J. Agric. Food Chem.* **59**, 3454–3484 (2011).
7. Roddick, J. G. The acetylcholinesterase-inhibitory activity of steroidal glycoalkaloids and their aglycones. *Phytochemistry* **28**, 2631–2634 (1989).
8. Al Sinani, S. S. S. & Eltayeb, E. A. The steroidal glycoalkaloids solamargine and solasonine in Solanum plants. *South African Journal of Botany* **112**, 253–269 (2017).
9. Sonawane, P. D. *et al.* Short-chain dehydrogenase/reductase governs steroidal specialized metabolites structural diversity and toxicity in the genus Solanum. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5419–E5428 (2018).
10. Lee, H. J. *et al.* Identification of a 3β-Hydroxysteroid Dehydrogenase/ 3-Ketosteroid Reductase Involved in α-Tomatine Biosynthesis in Tomato. *Plant Cell Physiol.* **60**, 1304–1315 (2019).

11. Cárdenas, P. D. *et al.* Pathways to defense metabolites and evading fruit bitterness in genus Solanum evolved through 2-oxoglutarate-dependent dioxygenases. *Nat. Commun.* **10**, (2019).

12. Nakayasu, M. *et al.* Identification of α-Tomatine 23-Hydroxylase Involved in the Detoxification of a Bitter Glycoalkaloid. *Plant Cell Physiol.* (2019). doi:10.1093/pcp/pcz224

13. Friedman, M. *et al.* Tomatine-Containing Green Tomato Extracts Inhibit Growth of Human Breast, Colon, Liver, and Stomach Cancer Cells. *J. Agric. Food Chem.* **57**, 5727–5733 (2009).

14. Itkin, M. *et al.* Biosynthesis of Antinutritional Alkaloids in Solanaceous Crops Is Mediated by Clustered Genes. *Science (80-.).* **341**, (2013).

15. Shakya, R. & Navarre, D. A. LC-MS analysis of solanidane glycoalkaloid diversity among tubers of four wild potato species and three cultivars (Solanum tuberosum). *J. Agric. Food Chem.* **56**, 6949–6958 (2008).

16. Itkin, M. *et al.* GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell* **23**, 4507–25 (2011).

17. Iijima, Y. *et al.* Steroidal glycoalkaloid profiling and structures of glycoalkaloids in wild tomato fruit. *Phytochemistry* **95**, 145–157 (2013).

18. Schwahn, K., de Souza, L. P., Fernie, A. R. & Tohge, T. Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. *J. Integr. Plant Biol.* **56**, 864–875 (2014).

19. Wu, S. B., Meyer, R. S., Whitaker, B. D., Litt, A. & Kennelly, E. J. A new liquid chromatography-mass spectrometry-based strategy to integrate chemistry,

morphology, and evolution of eggplant (Solanum) species. *J. Chromatogr. A* **1314**, 154–172 (2013).

20. Sawai, S. *et al.* Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato. *Plant Cell* **26**, 3763–74 (2014).

21. Umemoto, N. *et al.* Two Cytochrome P450 Monooxygenases Catalyze Early Hydroxylation Steps in the Potato Steroid Glycoalkaloid Biosynthetic Pathway. *Plant Physiol.* **171**, 2458–67 (2016).

22. Ginzberg, I., Tokuhisa, J. G. & Veilleux, R. E. Potato steroidal glycoalkaloids: Biosynthesis and genetic manipulation. *Potato Research* **52**, 1–15 (2009).

23. Ohyama, K., Okawa, A. & Fujimoto, Y. *Biosynthesis of steroidal alkaloids in Solanaceae plants: Incorporation of 3β-hydroxycholest-5-en-26-al into tomatine with tomato seedlings*. *Bioorganic & Medicinal Chemistry Letters* **24**, (2014).

24. Nakayasu, M. *et al.* A dioxygenase catalyzes steroid 16α-hydroxylation in steroidal glycoalkaloid biosynthesis. *Plant Physiol.* **175**, 120–133 (2017).

25. McCue, K. F. *et al.* Potato glycosterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. *Phytochemistry* **68**, 327–334 (2007).

26. McCue, K. F. *et al.* The primary in vivo steroidal alkaloid glucosyltransferase from potato. *Phytochemistry* **67**, 1590–1597 (2006).

27. McCue, K. F. *et al.* Metabolic compensation of steroidal glycoalkaloid biosynthesis in transgenic potato tubers: Using reverse genetics to confirm the in vivo enzyme function of a steroidal alkaloid galactosyltransferase. *Plant Sci.* **168**, 267–273 (2005).

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

29. Umemoto N, Sasaki K. June 2, 2013. Protein having glycoalkaloid biosynthetic enzyme activity and gene encoding the same. US Patent Application No. 20130167271 A1

30. Van Eck, J. Genome editing and plant transformation of solanaceous food crops. *Current Opinion in Biotechnology* **49**, 35–41 (2018).

31. Nicolia, A. *et al.* Targeted gene mutation in tetraploid potato through transient TALEN expression in protoplasts. *J. Biotechnol.* **204**, 17–24 (2015).

32. Clasen, B. M. *et al.* Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol. J.* **14**, 169–176 (2016).

33. Butler, N. M., Atkins, P. A., Voytas, D. F. & Douches, D. S. Generation and Inheritance of Targeted Mutations in Potato (Solanum tuberosum L.) Using the CRISPR/Cas System. *PLoS One* **10**, e0144591 (2015).

34. Andersson, M. *et al.* Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Rep.* **36**, 117–128 (2017).

35. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3570–5 (2015).

36. Ueta, R. *et al.* Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Sci. Rep.* **7**, (2017).

37. Xiao, A. *et al.* CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* **30**, 1180–1182 (2014).

38. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **3**, e3647 (2008).

39. Thagun, C. *et al.* Jasmonate-Responsive ERF Transcription Factors Regulate Steroidal Glycoalkaloid Biosynthesis in Tomato. *Plant Cell Physiol.* **57**, 961–975 (2016).

40. Abdelkareem, A. *et al.* Jasmonate-induced biosynthesis of steroidal glycoalkaloids depends on COI1 proteins in tomato. *Biochem. Biophys. Res. Commun.* **489**, 206–210 (2017).

41. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).

42. Wang, Z. *et al.* The positive effects of secreting cytokines IL-17 and IFN-γ on the early-stage differentiation and negative effects on the calcification of primary osteoblasts in vitro. *Int. Immunopharmacol.* **57**, 1–10 (2018).

43. Qi, X., Ye, C., Hou, Y. & Guo, X. A large spontaneous intrahepatic portosystemic shunt in a cirrhotic patient. *Intractable Rare Dis. Res.* **5**, 58–60 (2016).

44. Arora, L. & Narula, A. Gene editing and crop improvement using CRISPR-cas9 system. *Frontiers in Plant Science* **8**, (2017).

45. Li, J.-F. *et al.* Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat. Biotechnol.* **31**, 688–691 (2013).

46. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).

47. Ron, M. *et al.* Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. *Plant Physiol.* **166**, 455–69 (2014).

48. Porter, J. R. Host range and implications of plant infection by agrobacterium rhizogenes. *CRC. Crit. Rev. Plant Sci.* **10**, 387–421 (1991).

49. Noguchi, T. *et al.* Arabidopsis det2 is defective in the conversion of (24R)-24-methylcholest-4-en-3-one to (24R)-24-methyl-5α-cholestan-3-one in brassinosteroid biosynthesis. *Plant Physiol.* **120**, 833–839 (1999).

50. Rosati, F. *et al.* 5α-Reductase activity in Lycopersicon esculentum: Cloning and functional characterization of LeDET2 and evidence of the presence of two isoenzymes. *J. Steroid Biochem. Mol. Biol.* **96**, 287–299 (2005).

51. Ohnishi, T., Watanabe, B., Sakata, K. & Mizutani, M. CYP724B2 and CYP90B3 function in the early C-22 hydroxylation steps of brassinosteroid biosynthetic pathway in tomato. *Biosci. Biotechnol. Biochem.* **70**, 2071–2080 (2006).

52. Nakayasu, M. *et al.* Generation of α-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiol. Biochem.* **131**, 70–77 (2018).

53. Hashimoto, R., Ueta, R., Abe, C., Osakabe, Y. & Osakabe, K. Efficient Multiplex Genome Editing Induces Precise, and Self-Ligated Type Mutations in Tomato Plants. *Front. Plant Sci.* **9**, (2018).

54. Russell, D. W. & Wilson, J. D. Steroid 5alpha-Reductase: Two Genes/Two Enzymes. *Annu. Rev. Biochem.* **63**, 25–61 (1994).

55. Roddick, J. G. The steroidal glycoalkaloid α-tomatine. *Phytochemistry* **13**, 9–25 (1974).

56. Sonawane, P. D. *et al.* Corrigendum: Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. plants* **3**, 17101 (2017).

57. Christ, B. *et al.* Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* **10**, (2019).

58. Thomas, J. L., Boswell, E. L., Scaccia, L. A., Pletnev, V. & Umland, T. C. Identification of key amino acids responsible for the substantially higher affinities of human type 1 3β-hydroxysteroid dehydrogenase/isomerase (3β-HSD1) for substrates, coenzymes, and inhibitors relative to human 3β-HSD2. *J. Biol. Chem.* **280**, 21321–21328 (2005).

59. Obata, T. Metabolons in plant primary and secondary metabolism. *Phytochem. Rev.* (2019). doi:10.1007/s11101-019-09619-x

60. Nicot, N., Hausman, J.-F., Hoffmann, L. & Evers, D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J. Exp. Bot.* **56**, 2907–2914 (2005).

61. Hernandez, A. & Ruiz, M. T. An EXCEL template for calculation of enzyme kinetic parameters by non- linear regression. *Bioinformatics* **14**, 227–228 (1998).

62. Hesse, M.; Meier, H.; Zeeh, B. Spectroscopic Methods in Organic Chemistry, Translated by Dunmur, R. and Murray, M, 2nd Ed.; Thieme: New York, p 222 (2008).

63. Volz, H. & Gartner, H. N-acetoxyammonium ions - Reactive intermediates in the polonovski reaction. *European J. Org. Chem.* **2007**, 2791–2801 (2007).

64. Sinden, S. L. & Sanford, L. L. Origin and inheritance of solarmarine glycoalkaloids in commercial potato cultivars. *Am. Potato J.* **58**, 305–325 (1981).

65. Shih, M.-J. & Kuć, J. α and β-solamarine in kennebec Solanum tuberosum leaves and aged tuber slices. *Phytochemistry* **13**, 997–1000 (1974).

66. Kawai, Y., Ono, E. & Mizutani, M. Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* **78**, 328–343 (2014).

67. Barco, B. & Clay, N. K. Evolution of Glucosinolate Diversity via Whole-Genome Duplications, Gene Rearrangements, and Substrate Promiscuity. *Annu. Rev. Plant Biol* **70**, 585–604 (2019).

68. Cárdenas, P. D. *et al.* GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat. Commun.* **7**, 10654 (2016).