

The biosynthetic pathway of potato solanidanes diverged from that of spirosolanes due to evolution of a dioxygenase

Ryota Akiyama ¹, Bunta Watanabe ², Masaru Nakayasu^{1,6}, Hyoung Jae Lee¹, Junpei Kato¹, Naoyuki Umemoto ³, Toshiya Muranaka ⁴, Kazuki Saito ^{3,5}, Yukihiro Sugimoto ¹ & Masaharu Mizutani ¹ 

Potato (*Solanum tuberosum*), a worldwide major food crop, produces the toxic, bitter tasting solanidane glycoalkaloids α -solanine and α -chaconine. Controlling levels of glycoalkaloids is an important focus on potato breeding. Tomato (*Solanum lycopersicum*) contains a bitter spirosolane glycoalkaloid, α -tomatine. These glycoalkaloids are biosynthesized from cholesterol via a partly common pathway, although the mechanisms giving rise to the structural differences between solanidane and spirosolane remained elusive. Here we identify a 2-oxoglutarate dependent dioxygenase, designated as DPS (Dioxygenase for Potato Solanidane synthesis), that is a key enzyme for solanidane glycoalkaloid biosynthesis in potato. DPS catalyzes the ring-rearrangement from spirosolane to solanidane via C-16 hydroxylation. Evolutionary divergence of spirosolane-metabolizing dioxygenases contributes to the emergence of toxic solanidane glycoalkaloids in potato and the chemical diversity in Solanaceae.

¹Graduate School of Agricultural Science, Kobe University, Hyogo, Japan. ²Institute for Chemical Research, Kyoto University, Uji, Kyoto, Japan. ³RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, Japan. ⁴Department of Biotechnology, Graduate School of Engineering, Osaka University, Osaka, Japan. ⁵Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan. ⁶Present address: Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Kyoto, Japan. ✉email: mizutani@gold.kobe-u.ac.jp

Plants produce a vast array of specialized metabolites, many of which play essential roles in plant adaptation to environment and contribute to human health¹. Steroidal glycoalkaloids (SGAs), a class of specialized metabolites, are typically found in the plant family Solanaceae. While they serve as chemical protectant against plant pathogens and herbivores^{2,3}, some are notorious as antinutritional substance exerting negative effect on quality and marketability of *Solanum* (Family: Solanaceae) staple food crops, such as potato (*Solanum tuberosum*), tomato (*S. lycopersicum*), and eggplant (*S. melongena*)⁴. Therefore, controlling the SGA level is one of important targets for breeding of Solanaceae species.

SGAs consist of two structural components; the aglycone unit composed of nitrogen containing C₂₇ steroid derived from cholesterol and oligosaccharide attached to the hydroxy group at C-3^{3,5}. Based on the skeletal structure of the aglycone, SGAs can be divided into two general classes, solanidane or spirosolane (Fig. 1)^{3,6}. Minor structural variations of these two ring types such as C-5 saturation/unsaturation or isomerization at C-22, in combination with various sugar moieties, generate the enormous structural diversity of SGAs^{7,8}. In addition, their chemical structures reflect their biological activities, for example, toxicity to animals, anti-cancer properties, and anti-microbial activities^{3,9}. Most representatives of

solanidane glycoalkaloids are potato toxins, α -solanine **9** and α -chaconine **10**, that comprise upward of 90% of the total SGAs in cultivated potatoes^{10,11}. Additionally, more than 50 different SGAs, including spirosolane glycoalkaloids, have been identified in a variety of wild potato species and cultivars^{11,12}. Tomato and eggplant, on the other hands, contain only spirosolane glycoalkaloids^{3,13}. In tomato, α -tomatine **5** and dehydrotomatine **6** are predominant in green tissues, while the non-bitter SGA esculeoside A is the major component in the red mature fruits^{14,15}. α -Solasonine and α -solamargine are the two major spirosolane glycoalkaloids produced in eggplant¹⁶.

SGA biosynthesis can be divided into two main parts: aglycone formation and glycosylation. Recent research in potato and tomato identified several SGA biosynthetic genes involved in aglycone formation. Three cytochrome P450 monooxygenases (CYPs) named as PGA2 (GAME7), PGA1 (GAME8), and PGA3 (GAME4) have been found to be involved in hydroxylation of cholesterol **1** at C-22 and C-26 and oxygenation at C-26, respectively^{17,18}. A 2-oxoglutarate-dependent dioxygenase (DOX) named as 16DOX (GAME11), and an aminotransferase were reported to be required for the C-16 α -hydroxylation and C-26 amination during SGA biosynthesis^{18,19}. These enzymes and functions are common to potato and tomato, suggesting that

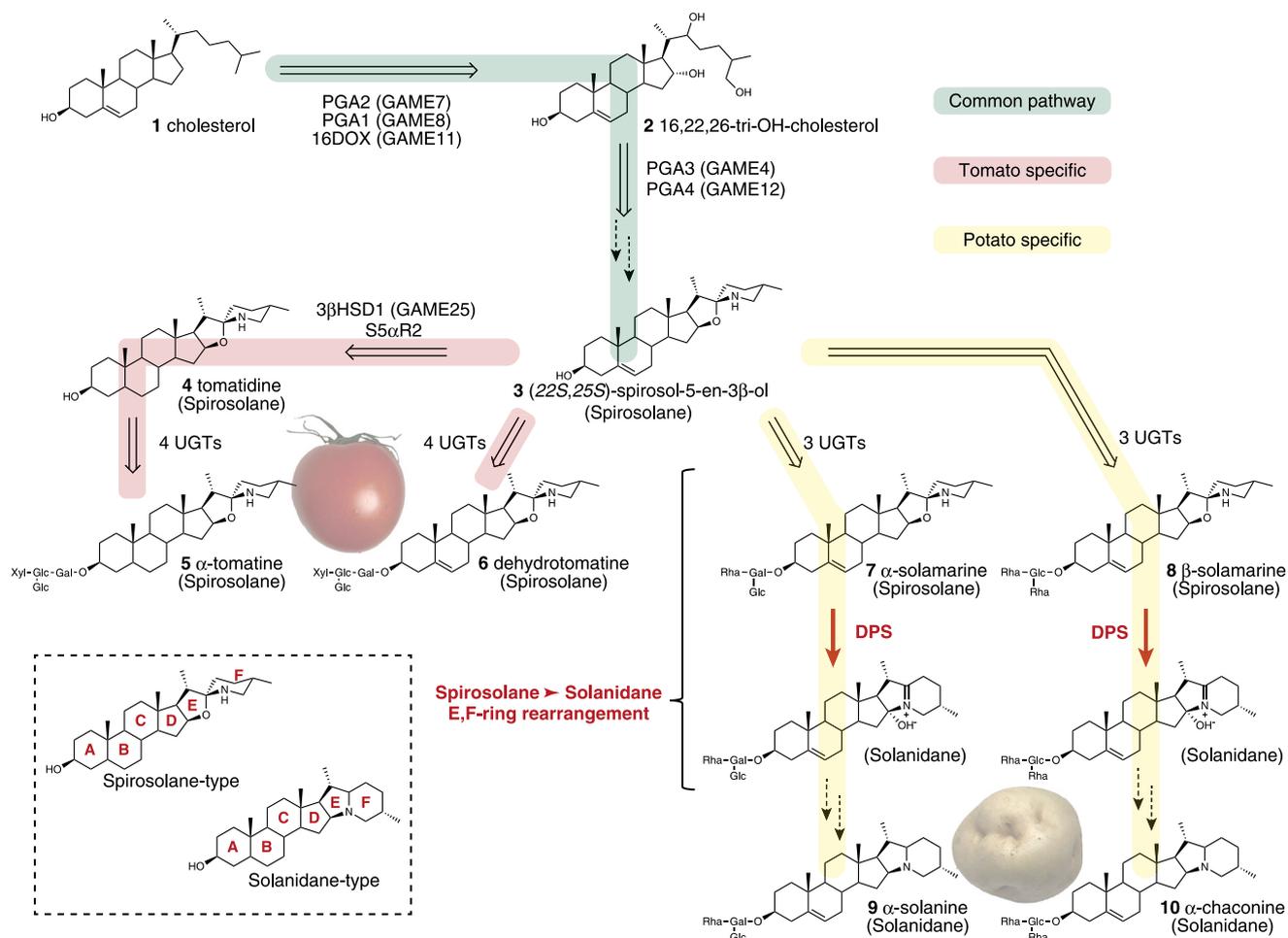


Fig. 1 Biosynthesis of solanidane and spirosolane glycoalkaloids in potato and tomato. Common steps between potato and tomato are shaded green. Steps specific to tomato or potato are indicated by red or yellow shade, respectively. White and dashed black arrows represent characterized and uncharacterized reactions, respectively. The reaction step characterized in this work, by DPS (Dioxygenase for Potato Solanidane synthesis), is shown in red arrow. PGA POTATO GLYCOALKALOID BIOSYNTHESIS, GAME GLYCOALKALOID METABOLISM, 16DOX 22,26-hydroxycholesterol 16 α -hydroxylase, 3 β HSD1 3 β -hydroxysteroid dehydrogenases 1, S5 α R2 steroid 5 α -reductase 2, UGT uridine diphosphate-dependent glycosyltransferases, Glc glucose, Gal galactose, Rha rhamnose, Xyl xylose. Photos of tomato and potato were taken by Ryota Akiyama, Graduate School of Agricultural Science, Kobe University.

they are involved in the biosynthetic steps common to solanidanes and spirosolanes (Fig. 1). In addition, several uridine diphosphate-dependent glycosyltransferases (UGTs) involved in the glycosylation steps of SGA biosynthesis have been identified in potato and tomato^{20–24}. However, the steps and enzymes involved in the metabolic branch point between solanidane-skeleton and spirosolane-skeleton formation remains an unsolved mystery.

In our efforts to elucidate the missing steps in SGAs biosynthesis, we focused on SGA biosynthesis in potato. Here, we find that solanidane glycoalkaloids are biosynthesized from spirosolane glycoalkaloids via ring rearrangement (Fig. 1). We discover DPS, a member of the DOX superfamily, is a key enzyme required for solanidane biosynthesis. DPS *in vitro* exhibits a novel activity converting the spirosolane-skeleton to a solanidane-skeleton (Fig. 1). Furthermore, studies of the genome organization of DPS homologs in a number of species from the *Solanum* genus reveal the evolutionary origin of this transformation. These results offer insights into the evolution of chemical diversity of SGAs in *Solanum* species.

Results

Discovery of the biosynthetic pathway of solanidane glycoalkaloids.

Most cultivated potatoes mostly contain only solanidane glycoalkaloids (α -solanine and α -chaconine), while some wild potato species and a few cultivars accumulate spirosolane glycoalkaloids, α -solanarine 7 and β -solanarine 8, which share the same sugar moiety with α -solanine 9 and α -chaconine 10, respectively^{12,25}. We speculated that modification of the spirosolane-skeleton would form the solanidane-skeleton in potato. To explore this hypothesis, we conducted feeding experiments using SGA-deficient potato hairy roots (16DOXko HR), in which the SGA biosynthetic gene, *St16DOX*, was disrupted²⁶. Firstly, the administration of α -solanarine 7 into 16DOXko HR resulted in restoration of the accumulation of α -solanine 9 in 16DOXko HR (Fig. 2a and Supplementary Fig. 1). Similar results were obtained from the feeding experiments with β -solanarine 8, in which the accumulation of α -chaconine 10 was detected (Supplementary Fig. 2). When feeding with (22*S*,25*S*)-spirosol-5-en-3 β -ol 3, which is an aglycone of α -solanarine 7 and β -solanarine 8, the accumulation of α -solanine 9, α -chaconine 10, and uncharacterized minor SGAs was confirmed (Supplementary Fig. 3). Additionally, the conversion of α -tomatine 6 to its corresponding solanidane, demissine 11, was confirmed with retention of the lycotetraose attached at the hydroxy group at C-3 (Supplementary Fig. 4). Furthermore, addition of the DOX inhibitor prohexadione during the α -solanarine 7 feeding assay caused the suppression of the conversion of α -solanarine to α -solanine 9 in a dose-dependent manner, whereas uniconazole-P, a CYP inhibitor, had little effects on the bioconversion (Fig. 2b). These data suggest that solanidane glycoalkaloids (i.e., α -solanine 9 and α -chaconine 10) are biosynthesized from spirosolane glycoalkaloids in potato and that a DOX family enzyme is involved in the conversion reactions.

Selection of candidate DOXs responsible for solanidane biosynthesis.

The DOX superfamily is one of the largest enzyme superfamily in plant kingdom and associated with the oxygenation reactions of various specialized metabolites²⁷. Recently, involvement of DOXs in SGA biosynthesis and modification was reported^{14,15,18,19}. In our previous study, *St16DOX*, which showed the highest fragments per kilobase of exon per million mapped fragments (FPKM) value in tuber sprout among the 256 DOX transcripts in RNA-Seq Gene Expression Data obtained from Spud DB (<http://solanaceae.plantbiology.msu.edu/>), was

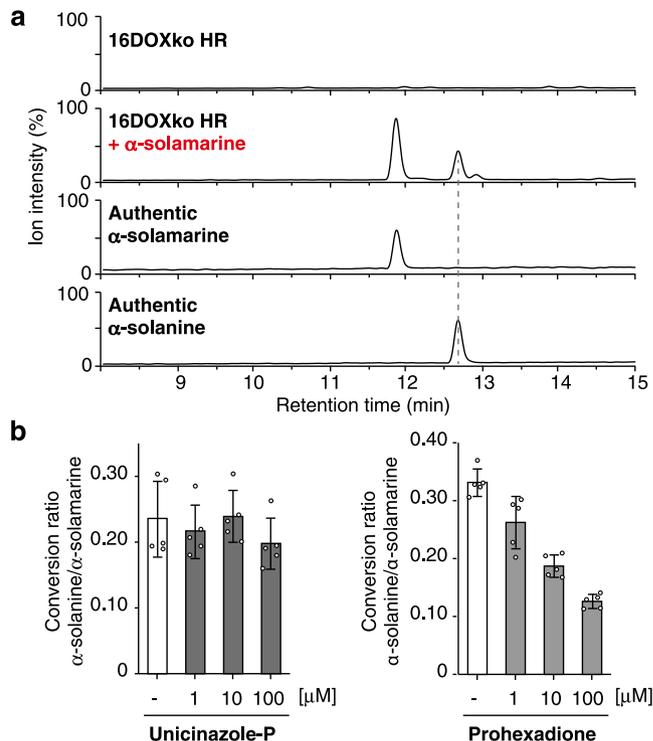


Fig. 2 Bioconversion of the solanidane-skeleton to the spirosolane-skeleton in potato hairy roots.

a LC-MS detection of conversion from administered α -solanarine to α -solanine in 16DOX-disrupted potato hairy roots (16DOXko HR). Total ion current (TIC) chromatogram obtained in positive ionization mode with a full-scan range of 350–1250 m/z are shown. Production of α -solanine was confirmed by comparison with the authentic standard of α -solanine in terms of retention time and mass spectrum provided in Supplementary Fig. 1. **b** Inhibition of the bioconversion of α -solanarine to α -solanine by uniconazole-P or prohexadione. After administering α -solanarine for 3 days in the presence of uniconazole-P or prohexadione, the SGAs accumulated in 16DOXko HR were extracted and quantified using LC-MS. Bars represent the ratio of α -solanine to α -solanarine accumulation in hairy roots. Error bars represent \pm standard deviation of mean ($n = 5$).

identified to encode the steroid 16 α -hydroxylase¹⁹. In this study, we focused on the gene *Sotub01g007130*, which showed the second highest FPKM value in the tuber sprouts among the potato DOXs (Supplementary Table 1). Quantitative reverse transcription PCR revealed that *Sotub01g007130* shared a similar expression pattern with previously identified SGA biosynthetic genes¹⁹ (Supplementary Fig. 5), and we henceforth named *Sotub01g007130* as *DPS* (dioxygenase for potato solanidane synthesis). Many SGA biosynthetic genes were reported to be clustered on chromosomes 7 and 12 in potato and tomato¹⁸, while *DPS* is positioned on chromosome 1. We found that *DPS* exists in a gene cluster (spanning about 160 kbp) with seven DOXs designated as *St7070*, *St7080*, *St7090*, *St7100*, *St7110*, *St7120*, and *St7150*, respectively. *St7090*, *St7100*, *St7120*, and *St7150* appeared to encode truncated proteins lacking canonical motifs conserved among the DOX superfamily²⁷. Among these clustered DOXs, *St7070*, *St7100*, and *St7110* also showed high FPKM values in potato tuber sprouts (Supplementary Table 1). We successfully amplified the full-length open-reading frame (ORF) sequences of *DPS* and *St7070* using the cDNAs derived from tuber sprouts of potato. *DPS* shares 85.8% amino acid sequence identity to *St7070* and a shorter C-terminus of 36-amino acid residues than *St7070*.

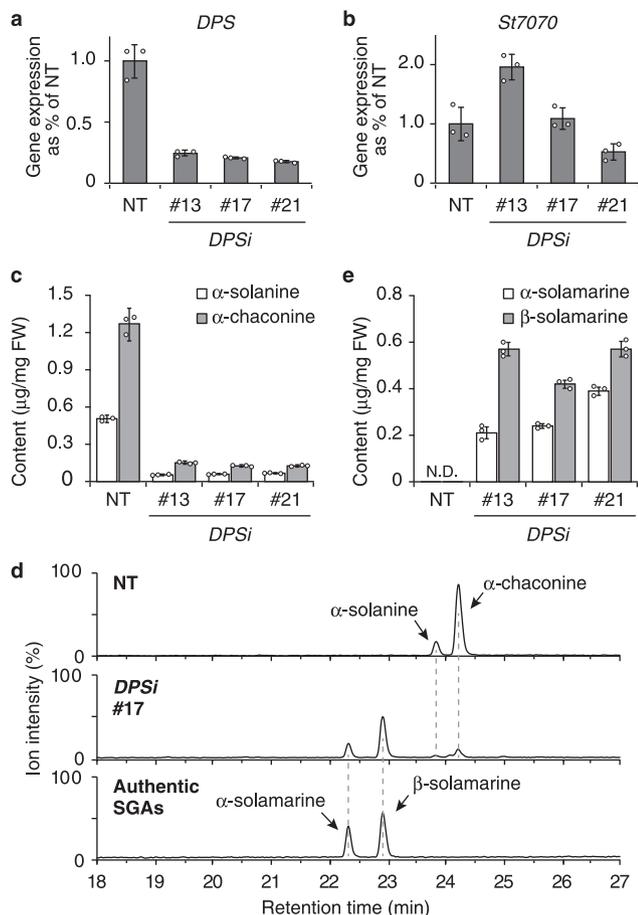


Fig. 3 Silencing of *DPS* in potato plants. Real-time quantitative RT-PCR analysis of *DPS* (a) and *St7070* (b) gene transcript levels in non-transformed and *DPS*-silenced (*DPSi*) potato plants. c Quantification of solanidane SGAs accumulated in non-transformed and *DPSi* plants using LC-MS. d SGA profiles determined by LC-MS. Total ion current chromatogram obtained in positive ionization mode with a full-scan range of 350–1250 m/z are shown. Accumulation of α -solamarine and β -solamarine was confirmed by comparison with authentic standards in terms of retention time and mass spectrum provided in Supplementary Fig. 7. e Quantification of spirosolane SGAs accumulated in non-transformed and *DPSi* potato plants. Error bars represent \pm standard deviation of mean ($n = 3$). FW fresh weight, NT non-transformed control plant, ND not detected, #13, #17, #21 independent transgenic lines.

Characterization of *DPS*-silenced transgenic potato plants. To investigate the contribution of *DPS* to solanidane biosynthesis, we generated *DPS*-silenced transgenic potato plants (i.e., *DPSi* lines) by using RNA-interference (RNAi) vector containing a stem-loop partial fragment of *DPS* cDNA (Supplementary Fig. 6). *DPSi* plants showed $\sim 80\%$ reduction in *DPS* transcript levels relative to non-transformed (NT) plants (Fig. 3a), and no off-target silencing effect on *St7070* was observed (Fig. 3b). The levels of solanidane glycoalkaloids (α -solanine 9 and α -chaconine 10) in the leaves of in vitro-grown *DPSi* plants were reduced by $\sim 90\%$ compared with those of the NT plants (Fig. 3c), suggesting that *DPS* is involved in solanidane biosynthesis in potato. Analysis of endogenous metabolites accumulated in *DPSi* plants detected two new peaks, of which the retention time (Rt) and mass spectrum matched with those of the authentic compounds, α -solamarine 7 and β -solamarine 8 (Fig. 3d and Supplementary Fig. 7). The amounts of α -solamarine 7 and β -solamarine 8 accumulated in *DPSi* plants were comparable to those of α -solanine and α -

chaconine in NT potato plants, respectively (Fig. 3e). These results implicate *DPS* as a key gene in the conversion from spirosolanes to solanidanes in potato.

Characterization of *DPS* function. Recombinant His-tagged protein of *DPS* was prepared with a bacterial expression system in *Escherichia coli* and purified by cobalt-affinity chromatography (Supplementary Fig. 8). The purified *DPS* was assayed with several spirosolanes as substrates, and the reaction products were analyzed using liquid chromatography–mass spectrometry (LC–MS). *DPS* metabolized α -solamarine 7 to form an unknown product, of which the retention time and the mass fragment pattern was not identical to α -solanine 9 (Fig. 4a, b). The product gave a parental ion mass at m/z 882.5 $[M + H]^+$ which is two mass smaller than that of the substrate α -solamarine: m/z 884.5 $[M + H]^+$ (Fig. 4b and Supplementary Fig. 9). Negative control reactions using either denatured *DPS*, lacking cofactors, or containing ethylenediaminetetraacetic acid (EDTA) as an Fe chelator failed to yield any products (Supplementary Fig. 10). The other spirosolane glycoalkaloids (β -solamarine 8 and α -tomatine 6) and the spirosolane aglycones ((2*S*,2*S*)-spirosol-5-en-3 β -ol 3 and tomatidine 4) were also accepted as substrates, and the mass decrease of 2 m/z was detected for each product (Supplementary Figs. 11–15). On the other hand, *DPS* is not active with solanidanes as substrates (Supplementary Fig. 16).

To identify the nature of the *DPS* reaction, we performed high-resolution mass spectrometry (HRMS) (Supplementary Fig. 17) and nuclear magnetic resonance (NMR) (Supplementary Data 1) analyses of the *DPS* enzymatic product obtained by reacting with α -tomatine 6, which is commercially available in enough quantities for structural analysis. The structure of the *DPS* reaction product was determined as shown in Supplementary Fig. 18 (Supplementary Fig. 19, Supplementary Tables 2, 3). The solanidane-skeleton scaffold is formed, and an iminium ion structure is constructed between C-16, the nitrogen atom, and C-22. Thus, *DPS* seems to catalyze hydroxylation at C-16 α of the spirosolane-skeleton to form a hemiacetal that then undergoes condensation with the secondary amine of the piperidine F ring to form the tertiary amine of solanidane, and hence, this enzymatic transformation by *DPS* is quite unusual (Fig. 4c). To gain insight into the reaction mechanism of dehydration and E/F ring rearrangement by *DPS*, we investigated the origin of the oxygen atom of the oxy ion in the *DPS* enzymatic reaction product by $^{18}\text{O}_2$ -labeling experiments. DOXs typically catalyze oxygenation/hydroxylation reactions, and DOX-dependent reactions in the presence of $^{18}\text{O}_2$ result in the incorporation of ^{18}O atom into the product. Accordingly, when the enzyme assay with Sl23DOX, which catalyzes C-23 hydroxylation of α -solamarine, was conducted in the presence of $^{18}\text{O}_2$, the reaction product gave 2 mass higher than the product assayed with $^{16}\text{O}_2$, indicating ^{18}O incorporation into the 23-hydroxylated product by Sl23DOX (Supplementary Fig. 20a–c). In contrast, the *DPS* enzymatic assay in the presence of $^{18}\text{O}_2$ did not yield any ^{18}O -labeled product (Supplementary Fig. 20d–f). This result indicates that the oxy ion at C-16 α in the *DPS*-reaction product is not derived from molecular oxygen and also suggests that the hydroxy moiety introduced by the *DPS*-dependent oxygenation leaves as water to form the solanidane. Based on these findings, we propose the reaction mechanism shown in Fig. 4d. First, *DPS* catalyzes hydroxylation at C-16 α of spirosolane-skeleton to form a 16-hemiacetal intermediate. Second, the lone pair on the nitrogen atom of the intermediate migrates to between C-22 and the nitrogen atom. Concomitantly, the carbon-oxygen bond at C-22 cleaves, and the hydroxy group at C-16 α introduced by *DPS* leaves as water to afford a 16-oxo species. Third, the nitrogen

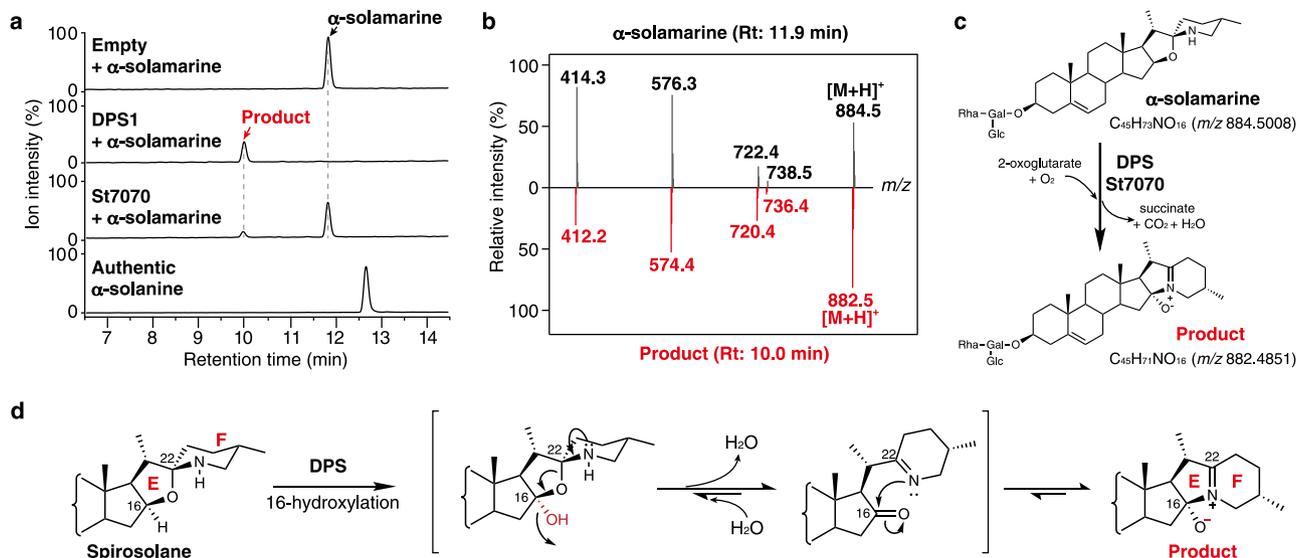


Fig. 4 The enzymatic activity of DPS and St7070. **a** LC-MS analysis of the reaction products from the recombinant DPS and St7070 proteins with α -solamarine as a substrate. Total ion current (TIC) chromatogram obtained in positive ionization mode with a full-scan range of 350–1250 m/z are shown. The negative control reaction was performed using a purified protein fraction from *E. coli* cells transformed with an empty pCold ProS2 vector. **b** Mass spectrum of the peaks at a retention time of 11.9 min (substrate, α -solamarine) and 10.0 min (reaction product). MS fragmentation pathways are shown in Supplementary Fig. 9. **c** Enzymatic reaction DPS or St7070. **d** Putative reaction mechanism of DPS or St7070.

atom attacks the C-16 carbonyl carbon to form an imine-oxy zwitterion, the DPS reaction product.

Kinetic analyses revealed that the k_{cat}/K_m value of DPS for α -solamarine 7 was 1123-fold greater than that for (2*S*,2*S*)-spirosol-5-en-3 β -ol 3, the aglycone of α -solamarine 7 (Supplementary Fig. 21 and Supplementary Table 4). The strong substrate preference of DPS for the glycosides suggests that conversion of spirosolane to solanidane occurs after the glycosylation step at the C-3 hydroxy group (Fig. 1). This is consistent with the observation that α -tomatine 6 was converted to its solanidane glycoalkaloid demissine 11 with retention of the lycotetraose at the C-3 hydroxy group (Supplementary Fig. 4). Recombinant St7070 assays with the above substrates gave the same reaction products as DPS, but the substrate preference of St7070 was significantly different from that of DPS. St7070 showed lower activity for α -solamarine 7 than DPS (Fig. 4a) and also showed only 4.5-fold higher preference for α -solamarine 7 than for its aglycone (2*S*,2*S*)-spirosol-5-en-3 β -ol 3 (Supplementary Fig. 21 and Supplementary Table 4).

Thus, DPS was found to catalyze the conversion of spirosolanes to form the corresponding solanidanes in vitro, but the resultant DPS-reaction products were not the solanidane end-products completed in planta. To resolve this inconsistency, the DPS-reaction products were fed to 16DOXko HR to examine their conversion to the corresponding end-products in vivo. The DPS-reaction product with α -solamarine 7 was converted into α -solanine 9 in 16DOXko HR (Supplementary Fig. 22). Similarly, administration of the DPS-reaction product with β -solamarine 8 to 16DOXko HR restored the accumulation of α -chaconine 10 (Supplementary Fig. 22). These results indicate that the DPS-reaction products serve as intermediates of solanidane glycoalkaloid biosynthesis in potato.

Evolution of SGA biosynthetic pathways in *Solanum* species.

With the establishment of DPS-associated metabolic pathway converting spirosolanes to solanidanes, we next linked this discovery to the evolution of SGA metabolism in the *Solanum* genus. We conducted a Basic Local Alignment Search Tool (BLAST)

search of DPS against the Tomato ITAG protein database (<https://solgenomics.net/>). This analysis identified a single distinct homolog Solyc01g006585 in the tomato genome, which shows 85% amino acid sequence identity to DPS. Recombinant Solyc01g006585 catalyzed the same reaction as DPS on the spirosolanes and was inactive with the solanidanes (Supplementary Figs. 23–29). Very low or no expression of Solyc01g006585 observed in RNA-seq data from the Tomato Functional Genomics Database (<http://ted.bti.cornell.edu>) (Supplementary Table 5) explains the lack of solanidane glycoalkaloids and accumulation of spirosolane glycoalkaloids (i.e., α -tomatine etc.) in tomato. Solyc01g006585 is embedded in the gene cluster containing three DOX genes in tomato chromosome 1 (Fig. 5a). Genes flanking this cluster showed considerable homology with genes flanking the DOX clusters harboring DPS, suggesting a common origin of these regions. Additionally, the eggplant DPS homolog, SMEL_001g151230 is present on the chromosome 1. SMEL_001g151230 protein, which shares 69% amino acid sequence identity to DPS, also had the same catalytic activity as DPS (Supplementary Fig. 30), while the expression of SMEL_001g151230 is quite low in aerial parts of eggplant²⁸ (Supplementary Table 6). These results indicate that the evolution of these orthologous DPS genes from duplication and functional divergence of the DOX family may have occurred before the speciation of *Solanum* species. Consequently, it is most likely that functional expression of the DPS gene or its orthologs is a prerequisite for the endogenous accumulation of toxic solanidane glycoalkaloids in *Solanum* species.

Based on the classification of the plant DOXs²⁷, all the clustered DOXs in potato, tomato, and eggplant are classified into the clade DOXC20. DOXC20 includes the α -tomatine 23-hydroxylase (Sl23DOX/GAME31) involved in conversion of bitter α -tomatine 6 to non-bitter esculoside A during tomato fruit ripening¹⁴. Sl23DOX/GAME31 shares 55% amino acid sequence identity to DPS, and the homologs of Sl23DOX are clustered in tomato chromosome 1 and 2 (Fig. 5a). In the phylogenetic analysis, DPS forms a distinct branch (the DPS branch) which is clearly separated from the 23DOX branch that includes Sl23DOX/GAME31 (Fig. 5b). Among the eight DOX

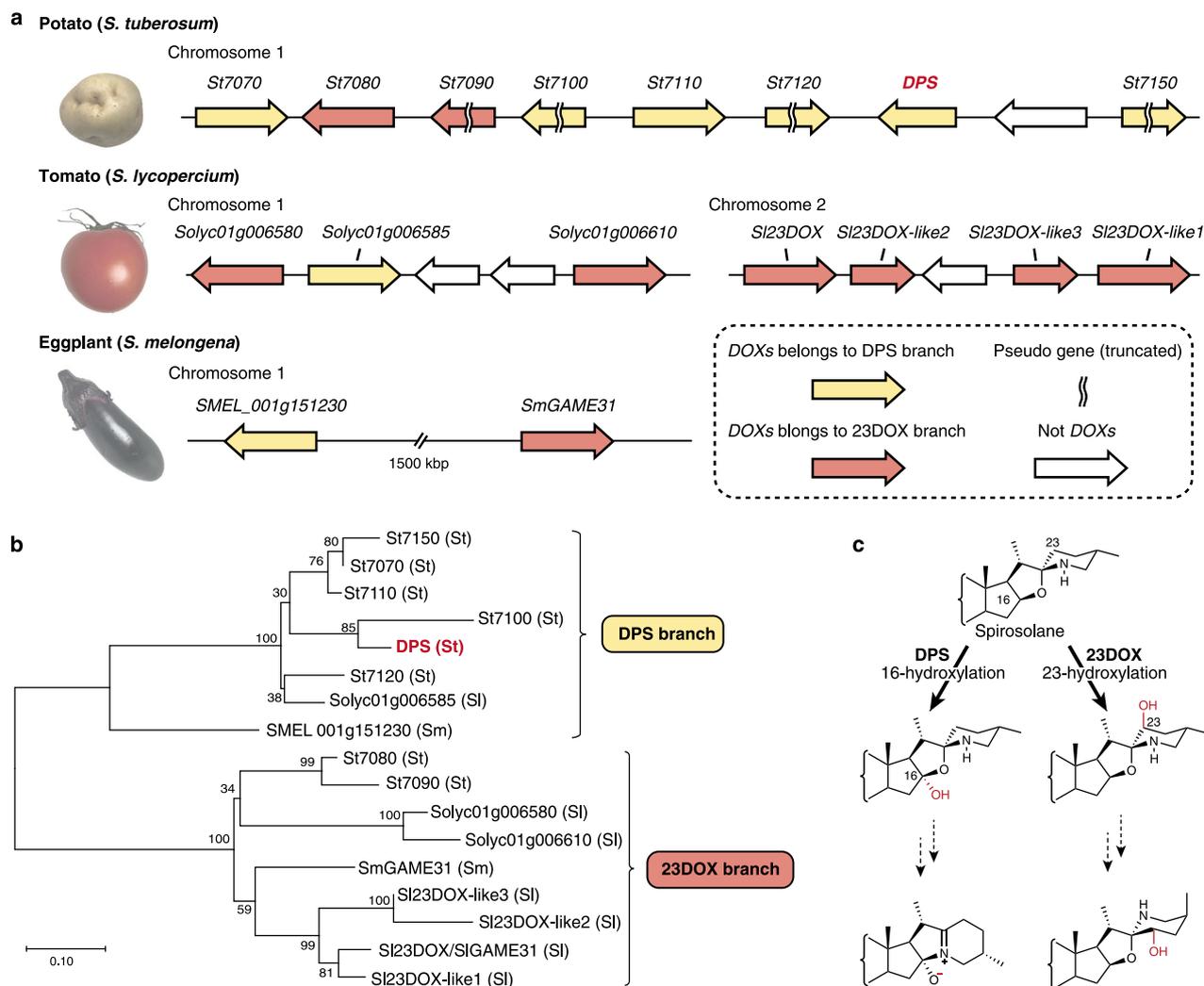


Fig. 5 Evolution of the DPS homologous genes associated with SGA metabolism in *Solanum* plants. **a** Genomic organization of the DPS homologous genes in potato, tomato, and eggplant. Genes belonging to the DPS branch and genes belonging to the 23DOX branch in phylogenetic analysis are shown in yellow arrows and red arrows, respectively. White arrows indicate genes not belonging to the DOX superfamily. **b** Phylogenetic analysis of the DPS homologous genes. A phylogenetic tree was generated using the maximum-likelihood method in MEGA X. Bootstrap values based on 1000 replicates are shown at the branching points. St *S. tuberosum* (potato), Sl *S. lycopersicum* (tomato), Sm *S. melongena* (eggplant). **c** Reactions catalyzed by DPS and 23DOX. 23DOX catalyzes C-23 hydroxylation of α -tomatine, and the product spontaneously isomerizes to neorickiioside B, which is an intermediate in α -tomatine metabolism that appears during ripening. Photos of tomato, potato, and eggplant were taken by Ryota Akiyama, Graduate School of Agricultural Science, Kobe University.

genes clustered in potato chromosome 1, *St7070*, *St7100*, *St7110*, *St7120*, *DPS*, and *St7150* are found in the DPS branch while *St7080* and *St7090* are in the 23DOX branch (Fig. 5b). *Solyc01g006585* and *SMEL_001g151230* are placed in the DPS branch, while the other DOXs clustered in tomato genome are in the 23DOX branch (Fig. 5b). These results indicate that functional diversification between *Sl23DOX/GAME31* and *DPS* may have arisen from a common ancestral DOXC20 before the speciation of *Solanum* species.

Discussion

In this study, we revealed a key reaction step diverging the biosynthetic pathways of two general classes of SGAs, solanidanes and spirosolanes. By the feeding experiments, we found that solanidane glycoalkaloids are biosynthesized from spirosolanes in potato. Gene silencing experiment and biochemical characterization revealed that potato *DPS*, belonging to the DOX superfamily, is involved in the metabolic conversion of spirosolanes to

solanidanes. Thus, *DPS* is a key enzyme in an evolutionary origin of the biosynthetic pathway of solanidane glycoalkaloids branched from spirosolanes in potato (Fig. 1).

The strong preference of *DPS* for glycosides observed in the *in vitro* assays (Supplementary Table 4) let us hypothesize that glycosylation at the C-3 hydroxy group occurs before formation of the solanidine backbone. Previously, *SGT1* and *SGT2* were reported to catalyze the initial transgalactosylation and transglucosylation of solanidine **12** at the C-3 hydroxy group, respectively. These two glycosyltransferases accept spirosolane aglycones as well, and the activities were comparable to the activity for solanidine^{21,22}. Additionally, *16DOX*-silencing which disrupts 16 α -hydroxylation of 22,26-dihydroxycholesterol resulted in the accumulation of glycosides of 22,26-dihydroxycholesterol **2**¹⁹. These previous findings support our hypothesis that conversion of spirosolane to solanidane occurs after the glycosylation step at the C-3 hydroxy group (Fig. 1).

Analysis of the catalytic activity of *DPS* clearly demonstrated that *DPS* is not sufficient to complete the conversion from

spirosolanones to the solanidane end-products accumulated in planta. The structure of the DPS-reaction product allowed us to propose a pathway for completing the solanidane-skeleton formation, which requires two imine-reduction reactions following to DPS reaction (Supplementary Fig. 31). Feeding of the DPS-reaction products to 16DOXko HR indicated their conversion to α -solanine and α -chaconine, suggesting the presence of putative enzyme(s) involved in the reduction of the imine intermediates in vivo (Supplementary Fig. 22). Analysis of transgenic tomato hairy roots constitutively expressing the *DPS* gene detected a new peak, of which the retention time (Rt. 18.7) and mass spectrum matched with those of the reaction product by DPS using α -tomatine as a substrate, but the production of demissine, which is the solanidane end-product derived from α -tomatine, was not detected (Supplementary Fig. 32). These results suggest that, in tomato, putative reductase(s) catalyzing the reduction of the imine intermediate is absent or not expressed. Thus, it is likely that the reductase(s), together with DPS, will be an important enzyme for solanidane glycoalkaloid biosynthesis in potato.

We identified the presence of DPS isoform designated as St7070 in potato genome, and biochemical characterization revealed that St7070 catalyzed the conversion of the spirosole-skeleton to the solanidane-skeleton. Gene silencing of *DPS* resulted in a significant reduction (~90%) in solanidane glycoalkaloid levels, but no off-target silencing effect on St7070 was observed (Fig. 3a–c). These results indicate that DPS is the major isoform involved in solanidane glycoalkaloid biosynthesis in potato, although there may be a minor contribution from St7070. We also identified functional DPS orthologs in the non-solanidane producing species, tomato and eggplant, and however, these orthologs were almost not expressed in these species. It has been reported that the transcription factor JRE4/GAME9 comprehensively regulate SGA biosynthetic genes in tomato^{10,29,30}. Overexpression of the transcription factor JRE4/GAME9 in potato led to significant increase of DPS transcription levels¹⁰, while the expression of *Solyc01g006585* was not induced by JRE4/GAME9 overexpression in tomato^{10,29,30}. These observations indicate that transcriptional regulation of DPS or its orthologs is a crucial factor diverging between spirosole and solanidane glycoalkaloids producers. There are two possible evolutionary scenarios; (i) gain of the DPS expression in solanidane-glycoalkaloids producing species such as domesticated potatoes and wild potatoes or (ii) loss of the DPS expression in non-solanidane glycoalkaloids producers such as tomato and eggplant. Further genetic studies are required to clarify the evolutionary events that underpin the differential expression of DPS and its orthologs in *Solanum* species.

Several studies reported that tandem gene duplication and subsequent changes in gene function drive the generation of new biosynthetic pathways or enhancement of existing metabolic flux^{31–34}. Therefore, it is noteworthy that the genes belonging to the DPS branch are highly duplicated in tandem on potato chromosome 1 (Fig. 5a). The fact that there is only one DPS ortholog in each tomato and eggplant genome indicates that the clustered genes including *DPS* and St7070 in potato have arisen from a single-copy ancestral gene via gene duplication after the speciation of *Solanum* species. *DPS* encodes a characteristic shorter polypeptide missing about 40 amino acid residues at the C-terminus as compared to typical DOXC members^{14,19,27} (Supplementary Fig. 33). Nucleic acid sequence alignment suggests that *DPS* was modified to become shorter via a single nucleotide-insertion (Supplementary Fig. 34). In addition, *DPS* and St7070 exhibited markedly differences in substrate specificity and catalytic efficiency, with *DPS* having very strong substrate preference for glycosides compared to St7070 (Supplementary Table 4). Given the major contribution of *DPS* to solanidane

glycoalkaloids biosynthesis, we presume that such improved catalytic property of *DPS*, in addition to its elevated gene expression, provide an evolutionary benefit by providing the metabolic pathways from spirosole to solanidane glycoalkaloids in cultivated potatoes. Several wild potato species contain both spirosoles and solanidanes^{12,25}, whereas domesticated potato cultivars produce almost only solanidane glycoalkaloids. To clarify the underlying mechanism of the different glycoalkaloid profiles between wild and cultivated potatoes, further experiments are needed, such as analyzing differences in gene expression, coding sequences, and catalytic properties for *DPS* and its homologous genes between wild and cultivated potatoes.

The SGA biosynthetic pathway is widely conserved among *Solanum* species^{18,28}, and evolutionary divergence of spirosole-metabolizing DOXs contributes to the chemical diversity of SGAs in the *Solanum* genus¹⁵. Recent studies identified that Sl23DOX/GAME31 in tomato catalyzes C-23 hydroxylation of α -tomatine^{14,15} (Fig. 5c), and this reaction is responsible for the first step of detoxification metabolism from toxic spirosole glycoalkaloid α -tomatine to non-toxic esculeoside A. Phylogenetic analysis indicates that functional differentiation between Sl23DOX/GAME31 and *DPS* in the DOXC20 subfamily likely has arisen from a common ancestral enzyme (Fig. 5). In tomato, spirosole-metabolizing Sl23DOX/GAME31 contributes to the domestication of cultivated tomatoes by the reduction of toxicity and bitterness in mature fruits. In contrast, metabolism of spirosole glycoalkaloids to solanidane glycoalkaloids in potato may represent evolutionary advantage to produce SGAs more toxic and effective in protection against their enemies.

In conclusion, we have revealed the genetic and enzymatic origin of solanidane glycoalkaloid biosynthesis in potato. This provides insight into the evolution of chemical diversity of SGAs in *Solanum* species and its evolution's footprints on *Solanum* genomes. Thus, our findings further elucidate understanding of how plants have evolved and developed their specialized metabolic pathways, demonstrating that evolutionary divergence of the DOX family is one of the key driving forces for phytochemical diversity.

Methods

Chemicals. Authentic samples of α -solanine **9**, α -chaconine **10**, and solanidine **12** were purchased from Sigma-Aldrich, and α -tomatine **6** was purchased from Tokyo Chemical Industry Co., Ltd. (22S,25S)-spirosole-5-en-3 β -ol **3** were isolated in our laboratory from tomatidine **4** (Chromadex), respectively, using HPLC^{35,36}. α -Solamarine **7** and β -solamarine **8** were purified from the diploid potato clone 97H32-6 and the chemical structures were confirmed using NMR in our laboratory³⁷.

Feeding experiments. St16DOX-disrupted potato hairy roots were established in our laboratory²⁶ and subcultured every month on an orbital shaker (100 rpm) in Gamborg B5 (B5) liquid medium containing 2% (w/v) sucrose and cefotaxime (250 μ g ml⁻¹) at 20 °C under 24-h dark conditions and used for the feeding experiments. α -Solamarine **7**, β -solamarine **8**, and (22S,25S)-spirosole-5-en-3 β -ol **3** (MeOH solution, final concentration 10 μ M) were added individually to the liquid medium of 3-week-old hairy roots. For the inhibition experiment, prohexadione (acetone solution) or uniconazole-P (acetone solution) was concomitantly added to the aquaculture with α -solamarine **7**. After 3 days of cultivation, SGAs contained in the harvested hairy roots were extracted three times from 100 mg fresh samples with 300 μ l methanol. The extracting solution was evaporated and the residue dissolved in 200 μ l methanol. After centrifugation, 10 μ l of supernatant was diluted with 290 μ l methanol and an aliquot (2 μ l) of the filtrate was analyzed using LC-MS. The analysis was performed using an ACQUITY UPLC H-Class System (Waters) with an SQ Detector 2 (Waters); data acquisition and analyses were performed using MassLynx 4.1 software (Waters). Each sample was injected into an ACQUITY UPLC HSS T3 chromatographic column (100 \times 2.1 mm, 1.7 μ m; Waters), with a column temperature of 40 °C and flow rate of 0.2 ml min⁻¹. The mass spectra were obtained in positive electrospray ionization mode, 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* 350–1250 was used. The mobile phases were water with 0.1% (v/v) formic acid (A) and acetonitrile (B), using gradient conditions 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 α -tomatine **6** feeding experiment, the mobile phases were water with 0.1% (v/v) formic acid (A), acetonitrile (B), and methanol (C), with a linear 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. Conversion of α -tomatine **6** to corresponding solanidane was confirmed by comparison with demissine **11** extracted from leaves of the wild potato *S. acaule*³⁸.

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 (*S. 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 first-strand cDNAs, using the ReverTra Ace[®] qPCR RT Master Mix with gDNA remover. Quantitative reverse transcriptional-polymerase chain reaction (RT-PCR) was performed with a LightCyclerNano (Roche), using FastStart Essential DNA Green Master (Roche) with the following primer sets: 1 and 2 for *DPS*, 3 and 4 for *St7070*, and 5 and 6 for *EF1 α* , which is a housekeeping gene used for real-time RT-PCR normalization³⁹ (Supplementary Table 7). Cycling was carried out as follows: 95 °C for 10 min, 45 cycles at 95 °C for 10 s, 60 °C for 10 s, amplification at 72 °C for 15 s, and holding at 95 °C for 30 s. This was followed by ramping up from 60 to 95 °C at 0.1 °C s⁻¹ to perform a melting curve analysis. Three biological replicates were analyzed in duplicate. The values obtained for the *EF1 α* gene were used as internal references in potato, and the gene expression levels were normalized against these values. Data acquisition and analysis were performed using LightCyclerNano software (Roche).

Cloning of *DPS* and *St7070* 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 frames of *DPS* and *St7070* were amplified by RT-PCR, using cDNA from tuber sprouts as a template, with primers 7 and 8 for *DPS* and primers 9 and 10 for *St7070* (Supplementary Table 7). These primers were designed from the potato unigenes sequences Sotub01g007130 and Sotub01g007070, respectively, using the Potato ITAG protein database (<http://solanaceae.plantbiology.msu.edu/index.shtml>). The PCR products were cloned into the pCR[®]4 Blunt-TOPO[®] vector (Thermo Fisher Scientific).

Generation of *DPSi* transgenic potato plants. The *DPS*-RNAi binary vector carried the inverted repeat of the partial *DPS* fragment interposing the third intron of the *Arabidopsis thaliana* *At4g14210* gene⁴⁰ under the control of the *cauliflower mosaic virus* 35S (CaMV35S) promoter in the T-DNA region. pRI201-AS (TaKaRa) was digested with *Kpn*I and *Sma*I 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 were ligated. The resulting plasmid was named pRI201- Δ MCS2 in this work. The third intron of *At4g14210* was PCR amplified from pKT258¹⁹ with primers 11 and 12, which contained restriction sites (Supplementary Table 7). The amplified fragments were cloned into the *Xba*I and *Nde*I sites of pRI201- Δ MCS2, and the resulting plasmid was named pRI201- Δ MCS2_PDS3. A 262 bp fragment of the *DPS* coding region was PCR-amplified using primers 13 and 14 (Supplementary Table 7). The fragment was inserted into the *Xba*I/*Bam*HI and *Nde*I/*Sall* sites of pRI201- Δ MCS2_PDS3 to construct a *DPSi* vector. Potatoes (*S. tuberosum* cv Sassy) were then transformed using *A. tumefaciens* EHA105 cells harboring the *DPSi* vector. Briefly, internodal stem sections were excised from in vitro-grown potato shoots and cut into ~5 mm sections. Explants were co-cultivated with agrobacterium for 48 h at 20 °C on MS medium containing 3% (w/v) sucrose, *trans*-zeatin (2 μ g ml⁻¹), indole-3-acetic acid (50 ng ml⁻¹), kanamycin (50 μ g ml⁻¹), and acetosyringone (100 μ M) under dark conditions. After the co-cultivation period, the stem pieces were transferred to shoot induction medium composed of agarose-solidified MS medium containing 3% (w/v) sucrose, *trans*-zeatin (2 μ g ml⁻¹), indole-3-acetic acid (50 ng ml⁻¹), kanamycin (50 μ g ml⁻¹), and meropenem hydrate (50 μ g ml⁻¹), and cultured at 20 °C under 16-h/8-h light/dark cycles. Stem pieces were transferred to fresh shoot induction medium every 2 weeks, until plantlets emerged. Well-developed plantlets were excised and transferred to MS medium containing 3% w/v sucrose, kanamycin (50 μ g ml⁻¹), and meropenem hydrate (50 μ g ml⁻¹). After 2–3 weeks, well-rooted shoots were selected and transformants were individually screened by genomic PCR using primer sets 15 and 16 (Supplementary Table 7), which targeted the kanamycin-resistance gene in the T-DNA region that was integrated into the genome.

Characterization of *DPSi* transgenic potato plants. Total RNA was then prepared from the leaves of four independent lines of the following in vitro-cultured plants: #13, #17, and #21. Quantitative RT-PCR analysis of *DPS* and *St7070* were performed using primers 1/2 and 3/4, respectively (Supplementary Table 7), as

described above. The SGAs that accumulated in the *DPSi* transgenic plants were extracted as described previously¹⁹ with minor modification: Briefly, 50 mg of fresh plant material was frozen and homogenized with a mixer mill at 4 °C. The homogenate was then extracted with 300 μ l MeOH. After centrifugation, the supernatant was collected. The extraction was repeated three times and the extract, dried in vacuo, was re-dissolved in 400 μ l MeOH. A 5 μ l sample of solution was diluted with 295 μ l MeOH and filtered through a 0.22- μ m polytetrafluoroethylene membrane filter. An aliquot (2 μ l) was then analyzed by UPLC-MS. LC-MS analysis was performed as described above, under “Feeding experiments”, with minor modifications. Briefly, the mobile phases were water with 0.1% (v/v) formic acid (A) and acetonitrile (B), using the following gradient conditions: solvent B ramped linearly from 10% to 32.5% over 30 min; solvent B increased linearly to 100% over 1 min, held at 100% solvent B 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 mode, 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* 350–1250 was used. Quantification of SGAs was carried out using selected ion monitoring (SIM). α -Solanine **9**, α -chaconine **10**, α -solanarine **7**, and β -solanarine **8** were monitored at *m/z* 868, 852, 884, and 868, respectively, and were quantified using the calibration curves of the authentic samples.

Expression of the recombinant proteins in *E. coli*. The coding sequences of *DPSi* and *St7070* were digested from the pCR[®]4 Blunt-TOPO[®] vector (Thermo Fisher Scientific) with *Nde*I and *Sall*. Since *Solyc01g006580* could not be amplified from cDNA in any tissue in tomato, the coding sequence of *Solyc01g006580* with *Nde*I and *Sall* sites was synthesized (Eurofins Genomics Inc.). The DNA fragments were then ligated into the *Nde*I-*Sall* sites of pClontech ProS2 (TaKaRa). *E. coli* strain BL21 (DE3) (Clontech) transformed with the constructed plasmid was grown at 37 °C in LB medium containing ampicillin (50 μ g ml⁻¹) until the OD₆₀₀ reached 0.5. Recombinant protein production was induced by adding 0.1 mM isopropyl β -D-1-thiogalactopyranoside and the culture was grown for 24 h at 15 °C. The culture was then centrifuged at 7720 \times g for 10 min at 4 °C, and the cell pellets were resuspended in 5 ml of cold sonication buffer containing 50 mM Bis-Tris-HCl (pH 7.2), 150 mM NaCl, 10% (v/v) glycerol, and 5 mM 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 20,630 \times g for 10 min at 4 °C. The recombinant His-tagged protein contained in the soluble fraction was purified using a His SpinTrap TALON column (GE Healthcare), according to the manufacturer’s instructions. After two column washes, the adsorbed proteins were eluted twice in 200 μ l 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 activity assay of the recombinant enzymes. The SGAs and their aglycones, α -solanarine **7**, β -solanarine **8**, α -tomatine **6**, α -solanine **9**, α -chaconine **10**, (2*S*,2*S*)-spirosol-5-en-3 β -ol **3**, tomatidine **4**, and solanidine **12**, were used as substrates in recombinant protein assays. An in vitro enzyme activity assay was performed using 100 μ l of reaction mixture containing 100 mM Bis-Tris-HCl (pH 7.2), 5 mM 2-oxoglutarate, 10 mM sodium ascorbate, 0.2 mM FeSO₄, 25 μ M substrate, and 1 μ g purified recombinant proteins as the enzyme. The reaction was initiated by the addition of the enzyme and was carried out at 30 °C for 30 min. The reaction was then stopped by incubation for 2 min at 90 °C. After centrifugation, 20 μ l of supernatant was diluted with 180 μ l methanol and filtered through 0.2- μ m nylon membrane filters (Acrodisc, Waters). An aliquot (2 μ l) was analyzed using LC-MS, as described in “Feeding experiments” above. We determined the kinetic parameters of the recombinant *DPS* and *St7070* in triplicate assays. The assays were conducted as described above, with minor modifications. Briefly, 0.1 μ g purified *DPS* protein or 0.02 μ g purified *St7070* were used. The activity was assayed using α -solanarine **7** or (2*S*,2*S*)-spirosol-5-en-3 β -ol **3** at concentrations ranging from 0.25 to 50 μ M, and the reaction was carried out at 30 °C for 10 min. A 20 μ l sample of assay mixture was diluted with 180 μ l methanol and filtered through 0.2- μ m nylon membrane filters (Acrodisc, Waters). An aliquot (2 μ l) was analyzed using LC-MS, as described above. The kinetic parameters were determined by nonlinear regression, using the ANEMONA program⁴¹.

¹⁸O₂-treated assay. A freshly prepared solution (1 ml) containing 100 mM Bis-Tris-HCl (pH 7.2), 5 mM 2-oxoglutarate, 10 mM sodium ascorbate, 0.2 mM FeSO₄, 25 μ M α -solanarine, was introduced into 10 ml round bottomed flask with a stir bar. Then, the flask was capped by a three-way cock which was connected to a vacuum pump line and a balloon filled with ¹⁸O₂, respectively. The reaction mixture was degassed in vacuo, and then ¹⁸O₂ was incorporated into the reaction mixture by purging once and stirring with a stirrer bar. An aliquot of the reaction mixture was quickly transferred to 1.5 ml microtube and mixed with the purified enzyme. The reaction was carried out at 30 °C for 30 min. The reaction was then stopped by incubation for 2 min at 90 °C. After centrifugation, 20 μ l of supernatant was diluted with 180 μ l methanol and filtered through 0.2- μ m nylon membrane filters (Acrodisc, Waters). An aliquot (2 μ l) was analyzed using LC-MS, as described in “Feeding experiments” above.

Structural determination of the DPS enzymatic reaction product. The DPS enzymatic reaction with α -tomatine **6** was performed using 50 ml of reaction mixture as described above, with a minor modification, to determine the structure of the product of the enzymatic reaction catalyzed by recombinant DPS: the reaction was conducted using the crude enzyme overnight. The reaction mixture was extracted with water-saturated butanol after the pH was adjusted to 3.0 by the addition of hydrochloric acid. The butanol layer was corrected and, before the dried residue was dissolved in methanol, DPS product was concentrated in vacuo. The DPS reaction product dissolved in methanol was subjected to preparative scale high performance liquid chromatography (HPLC). The sample was injected into an ODS column (COSMOSIL 5C₁₈-PAQ, 20 × 250 mm, Nacalai) at a column temperature of 40 °C, and the flow rate was set to 4 mL min⁻¹. Separation was performed with isocratic mobile phases of 22% acetonitrile in H₂O containing 0.1% trifluoroacetic acid (v/v), and the elute was detected at 203 nm. Then, the separated DPS reaction product was concentrated.

The molecular formula of the DPS enzymatic reaction product was determined to be C₅₀H₈₁NO₂₁, based on HRMS (FAB) analysis: found m/z 1032.5376 [M + H]⁺ (calcd. for C₅₀H₈₂NO₂₁, m/z 1032.5380 [M + H]⁺; Δ ppm = 0.1) (Supplementary Fig. 17). Its structure was determined by two-dimensional NMR measurements, based on correlation spectroscopy (COSY), nuclear Overhauser effect spectroscopy (NOESY), heteronuclear single-quantum correlation spectroscopy (HSQC), and heteronuclear multiple-bond correlation spectroscopy (HMBC) techniques (Supplementary Data 1). The NMR measurements in this study were carried out in CD₃SOCD₃ and CD₃OD. The ¹³C and ¹H assignments are summarized in Supplementary Tables 3 and 4, and COSY and HMBC correlations are shown in Supplementary Fig. 19. Samples were dissolved in NMR solvent immediately prior to measurement. In CD₃OD, 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 within 21 h. This observation indicated H-D exchange in CD₃OD. Accordingly, C-23 was observed as a severely broadened signal in CD₃OD by C-D coupling. In both solutions, C-16 signals were observed at 110.95 ppm (in CD₃SOCD₃) and 117.85 ppm (in CD₃OD), indicating their acetal like nature. Since C-16 of α -tomatine (the substrate of the DPS reaction) bears two carbon, one oxygen, and one hydrogen atom, the chemical shift in C-16 of the DPS reaction product indicates that oxidation occurred on C-16. In contrast, the chemical shift of C-22, within the hemiaminal structure of α -tomatine, observed at 192.21 ppm (in CD₃SOCD₃) and 196.80 ppm (in CD₃OD) in the DPS product at first suggested that C-22 exists as a carbonyl carbon of ketone or aldehyde. However, the HMBC correlation between H-26 and C-22 indicated that C-22 exists as an imine carbon instead of a carbonyl carbon. Since chemical shifts of typical imine carbons are observed around 140–170 ppm⁴², the chemical shift of C-22 of the DPS product appeared at a considerably lower field density. The observation indicated the presence of an iminium ion structure in the DPS product, and the reported ¹³C chemical shift of tri- or tetra-substituted carbon atoms of iminium ions (over 180 ppm, Supplementary Fig. 35)⁴³ is comparable to the C-22 chemical shift of the DPS reaction product. In particular, the chemical shift of the tetra-substituted iminium ion in Supplementary Fig. 35 (195.1 ppm) is in good accordance with those observed in the DPS product (192.21 or 196.80 ppm). Furthermore, the NOE correlation between H-26_b and H-15_b indicated the proximity of these hydrogen atoms. Additionally, the above-mentioned ¹³C chemical shift of C-16 indicated that not only one oxygen atom but also one nitrogen atom binds to C-16. Accordingly, the structure of the DPS product was determined to be a zwitterion species, as shown in Supplementary Fig. 18, in which an iminium ion structure is constructed between C-16, the nitrogen atom, and C-22. The previously mentioned H-D exchange of H-23 in CD₃OD, which further confirmed the iminium ionic form of the DPS product, is illustrated in Supplementary Fig. 36. The cationic nature of the nitrogen atom would increase the acidity of H-23 and accelerate the H-D exchange reaction.

Phylogenetic analysis. The amino acids sequences of St7070 (Sotub01g007070), St7080 (Sotub01g007080), St7090 (Sotub01g007090), St7100 (Sotub01g007100), St7110 (Sotub01g007110), St7120 (Sotub01g007120), DPS (Sotub01g007130), and St7150 (Sotub01g007150) were obtained from the Spud DB potato genomics resource (<http://solanaceae.plantbiology.msu.edu/>). Amino acid sequences of Sl23DOX/GAME31 (Soly02g062460), Sl23DOX/GAME31-like1 (Soly02g06250), Sl23DOX/GAME31-like2 (Soly02g062470), Sl23DOX/GAME31-like3 (Soly02g062490), Soly01g006580, Soly01g006585, and Soly01g006610 were downloaded from the Sol Genomics Network (<https://solgenomics.net>). Information about these DOX genes is summarized in Supplementary Data 2. Sequence alignments were performed using MUSCLE⁴⁴ and the maximum-likelihood tree was inferred in MEGA10⁴⁵, using 1000 bootstrap replications.

Overexpression of DPS in tomato hairy roots. A binary vector for DPS overexpression was constructed from the binary vector pBINPLUS⁴⁶; the coding region of DPS fused with the 5'-UTR of the *A. thaliana* alcohol dehydrogenase (*AtADH*) gene was inserted under the control of cauliflower mosaic virus 35S (CaMV35S) promoter and the heat shock protein terminator in the T-DNA region. The constructed binary vector was electroporated into *A. rhizogenes* ATCC15834. To generate transgenic hairy roots, tomato hypocotyls from 7-day-old seedlings were infected with *A. rhizogenes* strain ATCC15834 harboring a binary vector as

described previously²⁹. Briefly, bottom end of a hypocotyl segments (1.5 cm in length) were touched to a bacterial colony and then the segments were stood on the solidified B5 medium containing 2% (w/v) sucrose with the contacted end up. Hairy roots emerging from infected sites were excised and subcultured twice every week on solidified B5 medium containing 2% (w/v) sucrose, 300 mg l⁻¹ cefotaxime for disinfection, and 50 mg l⁻¹ kanamycin for drug resistance selection. The selected lines were maintained by subculturing every week in 100 ml glass flasks filled with 20 ml of liquid B5 medium supplemented with 2% (w/v) sucrose with shaking at 100 rpm. in the dark. The accumulated SGAs in DPS overexpressed tomato hairy roots were extracted as described above in “Feeding experiment” and then analyzed using LC-MS by the same condition as α -tomatine **6** feeding experiment in “Feeding experiment”.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequence data for DPS (LC547753), St7070 (LC547754), and Soly01g006585 (LC547755) are deposited in DDBJ. All data is available in the main text or the supplementary materials. Data supporting the findings of this study are available within the article and the supplementary materials. Source data are provided with this paper.

Received: 7 August 2020; Accepted: 22 January 2021;

Published online: 26 February 2021

References

- Chae, L., Kim, T., Nilo-Poyanco, R. & Rhee, S. Y. Genomic signatures of specialized metabolism in plants. *Science* **344**, 510–513 (2014).
- Friedman, M. Tomato glycoalkaloids: role in the plant and in the diet. *J. Agric. Food Chem.* **50**, 5751–5780 (2002).
- Milner, S. E. et al. Bioactivities of glycoalkaloids and their aglycones from solanum species. *J. Agric. Food Chem.* **59**, 3454–3484 (2011).
- Heftmann, E. Biogenesis of steroids in solanaceae. *Phytochemistry* **22**, 1843–1860 (1983).
- 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**, 3763–3774 (2014).
- Friedman, M. & McDonald, G. M. Potato glycoalkaloids: chemistry, analysis, safety, and plant physiology. *CRC Crit. Rev. Plant Sci.* **16**, 55–132 (1997).
- 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. USA* **115**, E5419–E5428 (2018).
- Sonawane, P. D., Jozwiak, A., Panda, S. & Aharoni, A. ‘Hijacking’ core metabolism: a new panache for the evolution of steroidal glycoalkaloids structural diversity. *Curr. Opin. Plant Biol.* **55**, 118–128 (2020).
- 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).
- 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).
- 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).
- Sinden, S. L. & Sanford, L. L. Origin and inheritance of solanidane glycoalkaloids in commercial potato cultivars. *Am. Potato J.* **58**, 305–325 (1981).
- Friedman, M. Chemistry and anticarcinogenic mechanisms of glycoalkaloids produced by eggplants, potatoes, and tomatoes. *J. Agric. Food Chem.* **63**, 3323–3337 (2015).
- Nakayasu, M. et al. Identification of α -tomatine 23-hydroxylase involved in the detoxification of a bitter glycoalkaloid. *Plant Cell Physiol.* **61**, 21–28 (2020).
- 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**, 5169 (2019).
- Sánchez-Mata, M. C., Yokoyama, W. E., Hong, Y. J. & Prohens, J. α -Solasonine and α -solamargine contents of gboma (*Solanum macrocarpon* L.) and scarlet (*Solanum aethiopicum* L.) eggplants. *J. Agric. Food Chem.* **58**, 5502–5508 (2010).
- Umamoto, N. et al. Two Cytochrome P450 monooxygenases catalyze early hydroxylation steps in the potato steroid glycoalkaloid biosynthetic pathway. *Plant Physiol.* **171**, 2458–2467 (2016).
- Itkin, M. et al. Biosynthesis of antinutritional alkaloids in Solanaceous crops is mediated by clustered genes. *Science* **341**, 175–179 (2013).
- Nakayasu, M. et al. A dioxygenase catalyzes steroid 16 α -hydroxylation in steroidal glycoalkaloid biosynthesis. *Plant Physiol.* **175**, 120–133 (2017).

20. 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).
21. 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).
22. McCue, K. F. et al. The primary in vivo steroidal alkaloid glucosyltransferase from potato. *Phytochemistry* **67**, 1590–1597 (2006).
23. McCue, K. F. et al. Potato glycoesterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. *Phytochemistry* **68**, 327–334 (2007).
24. Itkin, M. et al. GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell* **23**, 4507–4525 (2011).
25. Shih, M.-J. & Kuć, J. α and β -solanarine in kennebec *Solanum tuberosum* leaves and aged tuber slices. *Phytochemistry* **13**, 997–1000 (1974).
26. 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).
27. Kawai, Y., Ono, E. & Mizutani, M. Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* **78**, 328–343 (2014).
28. Barchi, L. et al. A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.* **9**, 11769 (2019).
29. Thagun, C. et al. Jasmonate-responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. *Plant Cell Physiol.* **57**, 961–975 (2016).
30. Nakayasu, M. et al. JRE4 is a master transcriptional regulator of defense-related steroidal glycoalkaloids in tomato. *Plant J.* **94**, 975–990 (2018).
31. Barco, B., Kim, Y. & Clay, N. K. Expansion of a core regulon by transposable elements promotes Arabidopsis chemical diversity and pathogen defense. *Nat. Commun.* **10**, 3444 (2019).
32. 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).
33. Fan, P., Leong, B. J. & Last, R. L. Tip of the trichome: evolution of acylsugar metabolic diversity in Solanaceae. *Curr. Opin. Plant Biol.* **49**, 8–16 (2019).
34. Des Marais, D. L. & Rausher, M. D. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765 (2008).
35. 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).
36. Akiyama, R. et al. Characterization of steroid 5 α -reductase involved in α -tomatine biosynthesis in tomatoes. *Plant Biotechnol.* **36**, 253–263 (2019).
37. Shimizu, K. et al. Hatching stimulation activity of steroidal glycoalkaloids toward the potato cyst nematode, *Globodera rostochiensis*. *Plant Biotechnol.* **37**, 1–7 (2020).
38. Kozukue, N. et al. Distribution of glycoalkaloids in potato tubers of 59 accessions of two wild and five cultivated *Solanum* species. *J. Agric. Food Chem.* **56**, 11920–11928 (2008).
39. 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).
40. Masclaux, F., Charpentreau, M., Takahashi, T., Pont-Lezica, R. & Galaud, J. P. Gene silencing using a heat-inducible RNAi system in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* **321**, 364–369 (2004).
41. Hernandez, A. & Ruiz, M. T. An EXCEL template for calculation of enzyme kinetic parameters by non-linear regression. *Bioinformatics* **14**, 227–228 (1998).
42. Hesse, M., Meier, H., Zeeh, B. *Spectroscopic Methods in Organic Chemistry* 2nd edn 222 (R. Dunmur, M. Murray, Trans.) (Thieme, 2007).
43. Volz, H. & Gartner, H. N-acetoxyammonium ions - reactive intermediates in the polonovski reaction. *Eur. J. Org. Chem.* **2007**, 2791–2801 (2007).
44. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
45. Kumar, S., Stecher, G., Li, M., Niyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
46. van Engelen, F. A. et al. pBINPLUS: an improved plant transformation vector based on pBIN19. *Transg. Res.* **4**, 288–290 (1995).

Acknowledgements

The authors are greatly indebted to K. Ohmine and A. Fujihashi (Kyoto University) for performing instrumental analyses. Parts of the experimental measurements were carried out using a Bruker AVANCEIII 800 and 600 NMR spectrometers as well as a JOEL JMS-700 spectrometer in the Joint Usage/Research Center (JURC) at the Institute for Chemical Research, Kyoto University. This work was supported in part by JSPS Grant-in-Aid for Japan Society for the Promotion of Science Research Fellow (19J10750 to R.A.); Cross-ministerial Strategic Innovation Promotion (SIP) Program, Japan (H.J.L., N.U., K. S., T.M., M.M.); The administration of individual commissioned project study (Development of new varieties and breeding materials in crops by genome editing, the Ministry of Agriculture, Forestry and Fisheries, Japan (N.U., K.S., T.M., M.M.) and the Collaborative Research Program of Institute for Chemical Research, Kyoto University (2017-83 to M.M. and B.W.).

Author contributions

R.A. and M.M. planned and designed the experiments. M.M., T.M., K.S., and Y.S. supervised the experiments. R.A. performed the feeding experiments and the DPS-silencing transgenic experiments. R.A., M.N., and J.K. performed protein expression and enzyme assay. R.A., B.W., J.K., and H.J.L. performed NMR analysis of the DPS reaction product. R.A., N.M., N.U., and M.M. performed the cloning and the sequence analyses of DOXs. R.A. and H.J.L. performed LC-MS/MS analysis of SGAs. R.A., B.W., N.U., and M. M. wrote the paper with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21546-0>.

Correspondence and requests for materials should be addressed to M.M.

Peer review information *Nature Communications* thanks Yuepeng Han, Rueben Peters, and Robert Winkler for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021