

Supplementary Information

A non-canonical vacuolar sugar transferase required for biosynthesis of antimicrobial defense compounds in oat

Anastasia Orme¹, Thomas Louveau¹, Michael J. Stephenson¹, Ingo Appelhagen¹, Rachel Melton¹, Jitender Cheema¹, Yan Li², Qiang Zhao², Lei Zhang², Danlin Fan², Qilin Tian², Robert J. Vickerstaff^{3,4}, Tim Langdon³, Bin Han² and Anne Osbourn^{1*}.

¹Department of Metabolic Biology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK.

²National Centre for Gene Research, CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Centre of Excellence for Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences (CAS), Shanghai 200032, China.

³Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth SY23 3FL, UK.

⁴Current address: East Malling Research, New Road, East Malling ME19 9BJ, UK.

*To whom correspondence should be addressed. Email: anne.osbourn@jic.ac.uk

This PDF includes:

Supplementary text

Figs. S1 to S14

Tables S1 to S16

References for SI reference citations

Additional files (to be uploaded separately):

Dataset 1: 'RSEM.isoform.TMM.EXPR.matrix.txt'. RNA-seq abundance estimates of transcripts generated in the de novo transcriptome assembly.

Dataset 2: 'SOM-analysis-results.xlsx'. Transcripts identified as co-expressed with characterised avenacin biosynthetic genes in the Self-Organising Map analysis.

Supplementary Information Text

Materials and Methods

Analysis of RNA-Seq data and generation of self-organising maps (SOMs).

RNA-seq reads were corrected using the Rcorrector method and filtered with the python script 'FilterUncorrectablePEfastq.py'

(<https://github.com/harvardinformatics/TranscriptomeAssemblyTools>) (1). Adapters and low quality bases (with a PHRED score lower than 5) were trimmed from corrected reads using TrimGalore version 0.5.0

(http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with a minimum length threshold of 36 nucleotides. The resulting filtered RNA-seq reads from the six oat tissues were pooled and used to generate a *de novo* transcriptome assembly with Trinity version 2.4.0 (2) (SI Appendix, Table S2;

<http://db.ncgr.ac.cn/oat/RNAseq.php>). Transcript abundance levels were estimated using the 'align and estimate abundance' perl script within the Trinity package.

Briefly, the filtered RNA-Seq reads for each of the tissues were aligned to the transcriptome (Bowtie-1.2.2) and abundance per gene was estimated using the RNA-Seq by Expectation-Maximization (RSEM) method using Trinity transcripts as a proxy for genes. The transcript abundance estimates were normalised to account for differences in library size by the trimmed mean of M-values (TMM) method (EdgeR) using the perl script 'abundance estimates to matrix.pl' within the Trinity package (Additional Dataset 1).

Self-organising maps were implemented and visualised in R (R version 3.5.2, 2018-12-20) using the Kohonen package (3). The analysis was carried out based on the method implemented in Jones *et al.* (4) and Payne *et al.* (5). To avoid boundary effects a toroidal map was used so that every node in the SOM had the same

number of neighbours. To represent the variation in the data, the dimensions of the SOM were chosen as the ratio between the two largest eigenvalues of the data (4-6). The number of SOM nodes was chosen based on the following ratio as in Jones (4):

$$\frac{\sum_{c=1}^S N_c (x_c - \bar{x})^2}{\sum_{g=1}^N (x_g - \bar{x})^2}$$

where N is the total number of genes; x_g is the expression vector for gene g ; \bar{x} is the global mean of all expression vectors; x_c is the expression vector for a SOM node c ; N_c is the total number of genes assigned to SOM node c and S is the total number of SOM nodes. The value of S was chosen so that the ratio was about 0.85 ($S=16$; ratio = 0.84). The RNA-Seq data were filtered by removing transcripts that did not have TMM-normalised values above 10.0 in any tissue. The filtered data were normalised to have a mean expression level of 0.0 and a variance of 1.0 across tissues and assigned to a SOM cluster based on a minimal Euclidean distance. The random presentation of data during the training of a SOM can lead to different mapping results if the training is repeated (3). The SOM training was repeated 100 times with seed values 1-100. Transcripts that mapped to the same nodes as six characterised avenacin biosynthetic genes (SI appendix, Fig S1) in SOM runs were extracted (Additional Dataset 2). To identify transcripts that encoded putative UGTs, the transcriptome was mined with a tBLASTn (TBLASTN 2.7.1) search using full-length sequences of representative UGTs from plant UGT subfamilies A-O (SI Appendix, Table S3). Transcripts that encode putative UGTs were identified for the transcripts that mapped to the same node as a characterised avenacin biosynthetic gene in 100% of SOM runs (Additional Dataset 2). The transcripts that encode full-length putative UGTs are listed in SI Appendix, Table S4. The heatmap was drawn in R using heatmap.2, <https://CRAN.R-project.org/package=gplots>).

Genomic DNA preparation. Genomic DNA for amplification and subsequent cloning of UGT genes was extracted from leaves of 5-day-old *A. strigosa* seedlings. Ground tissue was resuspended in 1.2 ml of extraction buffer (0.2 M Tris-HCl pH 8.0, 250 mM NaCl, 25 mM ethylenediaminetetraacetic acid (EDTA), 1% sodium dodecyl sulfate (SDS, Sigma-Aldrich) and centrifuged at 13 000 x g for 5 min. The supernatant (4 x 375 μ l) was removed to fresh tubes. An equal volume of

isopropanol was added and mixed by inversion, and the tubes centrifuged at 13 000 x g for 10 min. The supernatant was removed and the pellets washed with 600 μ l of 70% ethanol, centrifuged at 13 000 x g for 10 min, and the pellets dried at 35°C for 20 minutes in a vacuum dryer. Pellets were resuspended in 200 μ l water and stored at -20°C.

Expression of recombinant glucosyltransferases in *Escherichia coli*.

Chemically competent *E. coli* Rosetta strain DE3 (Novagen) was transformed with pH9GW expression vectors following the manufacturer's instructions. Selected transformants were cultured in liquid Lysogeny Broth (LB) at 37°C with shaking at 220 rpm with kanamycin (100 μ g/ml) and chloramphenicol (35 μ g/ml) selection. The cultures were diluted 100-fold into fresh medium and grown at 37°C, 200 rpm until they reached an OD₆₀₀ of 0.5-0.6. Expression of recombinant protein was induced at 16°C overnight with 0.05 mM isopropyl- β -D-thiogalactopyranoside (IPTG) (Sigma-Aldrich), after 30 min of acclimatisation. Cells were harvested by centrifugation at 3220 x g for 10 min, the supernatant discarded and the cells frozen at -80°C. The frozen cell pellets were thawed on ice and resuspended in 6 ml of chilled sonication buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.8, 20 mM imidazole, 5% glycerol, cComplete™ EDTA-free protease inhibitor cocktail (Roche) (1 tablet per 50 ml sonication buffer), 0.1% Tween 20 (Sigma- Aldrich). Resuspended cells were sonicated using a bench top ultrasonic disintegrator (Soniprep 150 plus, MSE) in iced water for 5 x 10 sec (amplitude = 7.0) with 20 sec rest. Cell lysates were centrifuged at 12000 x g for 20 min at 4°C. Supernatants were incubated with 150 μ l pre-equilibrated Ni-NTA Agarose beads (Roche) with agitation at 4°C for 1 h. Beads were transferred to 1.5 ml Eppendorf tubes and washed 3 times with 500 μ l filtered Buffer A (300mM NaCl, 50mM Tris-HCl pH 7.8, 20mM imidazole, 5% glycerol). Proteins were eluted with 3 x 200 μ l Buffer B (300mM NaCl, 50mM Tris-HCl pH 8, 500 mM imidazole, 5% glycerol). Protein elution fractions were combined, and the buffer exchanged by adding 2 x 2 ml 50 mM Tris-HCl pH 7.5 and concentrating in Amicon®Ultra-4 Centrifugal Filter Units with Ultracel-10 membranes (Merck) at 3220 x g, with a final volume of approximately 250 μ l. Protein concentrations were estimated using the Bradford assay (Bio-Rad Protein Assay Kit I, with bovine γ -globulin standard) as per the manufacturer's instructions, and protein purity assessed by SDS-PAGE.

Generation of deglycosylated avenacin A-1 substrates. To generate acceptor molecules lacking the 1-2- and 1-4-linked D-glucoses, avenacin A-1 was enzymatically deglycosylated using avenacinase, a glycosyl hydrolase from the fungus *Gaeumannomyces graminis* var. *avenae* (7-9). For the isolation of avenacin A-1, oats were grown hydroponically and the roots harvested after 1 week, and then again after a further week of regrowth. Freeze-dried ground roots were extracted overnight at 4°C with 80% methanol, filtered through Miracloth (Merck) and Whatmann filter paper, and the methanol evaporated in using a rotary evaporator. The aqueous filtrate was precipitated at 4°C overnight. The precipitate was collected, freeze-dried, resuspended in methanol with sonication, and then dried onto diatomaceous earth (Celite, Sigma-Aldrich). Compounds were separated on a Biotage®SNAP C18 30g flash chromatography column with a flow rate of 25 ml/min using a gradient of 5% methanol (Solvent A) and 95% methanol (Solvent B) as follows: 0-100% Solvent B over 21.9 column volumes and 100% Solvent B for 17.4 column volumes. Fluorescent fractions were combined and avenacin A-1 purified from the crude avenacin mix using a gradient of 0-100% Solvent B over 30.3 column volumes and 100% Solvent B for 15.9 column volumes.

Avenacinase preparations were prepared as described in Osbourn et al. (7). Blocks of mycelium from actively growing colonies of *G. graminis* var. *avenae* strain A3 (10) were placed on potato dextrose agar plates with 50 µg/ml streptomycin and 50 µg/ml ampicillin. After 5 days incubation in the dark at 22°C, colonies were scraped from the plates and homogenised with 1 ml Jermyn's medium (soluble starch: 1 g/L; K₂HPO₄: 6 g/L; NH₄Cl: 8 g/L; Yeast Extract: 1 g/L; MgSO₄·7H₂O: 1 g/L; CaCl₂: 0.02 g/L; ZnSO₄·7H₂O: 0.002 g/L; MnSO₄·7H₂O: 0.001 g/L) per colony. The homogenate was added to 2 L flasks containing 500 ml Jermyn's medium with 50 µg/ml streptomycin and 50 µg/ml ampicillin (1 colony per 100 ml), and the cultures grown for 5 days at 22°C with shaking at 200 rpm. Cultures were filtered through Miracloth (Merck) and two EDTA-free protease inhibitor tablets (Roche) per 500 ml filtrate was added. Filtrates were chilled to 4°C and ammonium sulphate added with stirring to a final concentration of 580 g/L⁻¹ of culture filtrate. The culture filtrate was centrifuged at 15000 x g at 10°C for 10 min, the supernatant discarded, and the pellet resuspended in a minimum volume of ice cold sterile water. The protein preparation

was dialysed with four changes of buffer against 20 mM Tris-HCl pH 8 at 4°C, centrifuged at 15000 x g at 10°C for ten min, and frozen at -20°C.

Avenacinase reaction conditions were adapted from Osbourn et al (7). To generate 3 β -{deglycosyl-(1->4)-} avenacin A-1 (**8**) (SI Appendix, Table S1), avenacin A-1 (0.3 mM) was incubated with an avenacinase protein preparation (3 μ l per 100 μ l reaction volume) in 100 mM sodium acetate pH 5.0 for 15 minutes at 37°C. To generate 3 β -{bisdeglycosyl-(1->2)-,(1->4)-} avenacin A-1 (**2**) (SI Appendix, Table S1) the reaction time was extended overnight. Reactions were dried down in a Genevac EZ-2 Elite centrifugal evaporator and resuspended in methanol. Samples were loaded on to 20 x 20 cm preparative silica thin layer chromatography (Merck) plates using a pipette (TLC plates were first pre-run three times in 100% methanol to 0.5 cm above the loading line). Avenacin A-1 and deglycosylated products were then separated using a mobile phase of dichloromethane:methanol:water (80:19:1; v:v:v) and visualised under ultraviolet light. Avenacin A-1 hydrolysis products were scraped off the TLC plates with a scalpel blade, resuspended in 15 ml methanol:ethyl acetate (25:75, v:v), and the suspensions filtered through filter paper. Fractions were dried down using a Genevac EZ-2 Elite centrifugal evaporator and stored at -20°C.

Glycosylation assays. Assays for UGT activity comprised 50 mM Tris-HCl pH 7.5, 0.5 mM UDP- α -D-glucose (Sigma-Aldrich) and 150 μ M acceptor 3 β -{bisdeglycosyl-(1->2)-,(1->4)-} avenacin A-1 (**2**) or 3 β -{deglycosyl-(1->4)-} avenacin A-1 (**8**) (SI Appendix, Table S1) in a total volume of 50 μ l. Reactions were initiated by addition of 4 μ g of recombinant protein preparation and incubated at 25°C overnight. AsTG1 activity reactions contained 50 mM citrate buffer pH 5.5, 5 mM *p*-nitrophenyl glucose and 150 μ M avenacin A-1 lacking the 1-4-linked D-glucose, 3 β -{deglycosyl-(1->4)-} avenacin A-1 (**8**) (SI Appendix, Table S1). AsTG1 reactions were initiated by addition of approximately 1 μ g of recombinant partially purified AsTG1-NOSIG (AsTG1 missing the predicted N-terminal signal sequence) and incubated at 30°C for 90 min. Reactions were stopped by the addition of 50 μ l methanol. Control reactions were set up as above, except with protein preparations that had been boiled at 95°C for 10 minutes. Reaction mixtures were centrifuged and product analysis was carried out by reverse phase HPLC using a 50 x 2.1mm 2.6 μ Kinetex XB-C18 column (Phenomenex) with a column oven temperature of 30°C. Detection was by UV/Vis absorbance (Shimadzu SPD-M20A), collecting spectra from 200-600nm.

Electrospray MS data (Shimadzu LC-2020 dual source MS) were collected in positive and negative modes from m/z 50-1500. The gradient was run at 0.3 ml/min with 100% water as Buffer A, 100% acetonitrile as Buffer B and was as follows: 25% Buffer B from 0-0.6 min; 25-80% Buffer B from 0.6-7 min; 80-100% Buffer B from 7-7.2 min; a linear gradient between 7.2-8 min; 100 to 25% Buffer B from 8-8.1 min, and held at 25% Buffer B until 10 min.

Analysis of metabolites from *N. benthamiana* leaves. Leaves were harvested five days after infiltration and freeze-dried. Dried leaf samples (20 mg) were ground twice at 1000 rpm for 30 sec in a Geno/Grinder SPEX Sample Prep 2010. After centrifugation at 13000 x g for 10 sec, the ground leaf material was extracted with 1 ml of 80% MeOH with 20 μ g digitoxin standard (Merck) at 18°C for 20 min with shaking at 1400 rpm. Samples were centrifuged at 20000 x g at 4°C for 2 min and 0.8 ml of supernatant was partitioned twice with 400 μ l hexane on ice. Aliquots (400 μ l) of aqueous fractions were dried in a Genevac EZ-2 Elite centrifugal evaporator maintaining the temperature below 30°C and stored at -80°C. For high performance liquid chromatography (HPLC), samples were resuspended in 75 μ l methanol and filtered through Corning®Costar®Spin-X®centrifuge tube filters (Sigma-Aldrich). The filtrate (50 μ l) was combined with 50 μ l 50% MeOH and 10 μ l was analysed by reverse phase HPLC using a 50 x 2.1mm 2.6 μ Kinetex XB-C18 column (Phenomenex). The column oven temperature was set at 30°C and detection was by charged aerosol detector (CAD, Corona Ultra RS from Dionex), as well as electrospray MS (Shimadzu LC-2020 dual source MS) collected in positive and negative modes from m/z 50 -1500. The gradient was run at 0.3 ml/min with 100% water as Buffer A, 100% acetonitrile as Buffer B as follows: 15% Buffer B from 0-1.5 min; 15-60% Buffer B from 1.5-26 min; 60-100% Buffer B from 26-26.5 min; a linear gradient between 26.5-28.5 min; 100 to 15% Buffer B from 28.5-29 min, and held at 15% Buffer B for 30 min.

Purification and structure determination of 3 β -{[β -D-glucopyranosyl-(1->2)- α -L-arabinopyranosyl]oxy}-12-keto,16 β -hydroxy- β -amyirin (4b) and 3 β -{[β -D-glucopyranosyl-(1->2)-[β -D-glucopyranosyl-(1->4)]- α -L-arabinopyranosyl]oxy}-12-keto,16 β -hydroxy- β -amyirin (12b). To generate (4b), co-expression of tHMGR, AsbAS1 (SAD1), AsCYP51H10 (SAD2), AsAAT1 and AsUGT91G16 was performed by vacuum agro-infiltration of *N. benthamiana* plants following the method of Reed et

al. (11). To generate (**12b**), AsTG1 was co-expressed with the above enzymes. After five days, leaves were harvested and freeze-dried. The dried leaf material (40 g and 33 g, respectively) was ground to a powder in a pestle and mortar. Extraction was performed using a SpeedExtractor E-914 (Büchi) as described in Reed et al. (11) with 6 cycles at 90°C and 130 bar pressure. Cycle one (ethyl acetate) had zero hold time, cycle two (ethyl acetate) had five min hold time and cycles 3-6 (methanol) had five min hold time. The run finished with a two min solvent flush and a six min N₂ flush. Crude extracts were monitored by TLC or HPLC-MS and fractions containing the compounds of interest combined and dried by rotary evaporation, resuspended in methanol, and dried onto diatomaceous earth (Celite 545 AW, Sigma-Aldrich) for dry-loading of chromatography columns.

For purification of (**4b**), extracts were separated by normal phase chromatography on a column of silica gel 60 (Material Harvest) with DCM:MeOH (90:10, v:v) over 3 L, DCM:MeOH (80:20, v:v) over 1 L and DCM:MeOH (70:30, v:v) over 1 L. Fractions containing (**4b**) were identified by TLC, pooled and dried onto diatomaceous earth (Celite, Sigma-Aldrich) and further purification was performed using an Isolera One (Biotage) automatic flash purification system. The resulting solid was separated by normal phase chromatography on a Biotage®SNAP KP-Sil 50 g column with a flow rate of 100 ml/min as follows; 100 % DCM (Solvent A) and 0% methanol (Solvent B) for 4.9 column volumes; 0-10% Solvent B over 7.7 column volumes; 10% Solvent B for 3 column volumes; 10-15% Solvent B over 3.8 column volumes; 15% Solvent B for 3.7 column volumes; 15-20% Solvent B over 3.5 column volumes and 20% Solvent B for 3.6 column volumes. Fractions containing (**4b**) were identified by TLC, combined, and dried onto diatomaceous earth (Celite, Sigma-Aldrich) for further separation by reverse phase chromatography on a Biotage®SNAP C18 30g column with a flow rate of 25 ml/min as follows: 45% water (Solvent A) and 55% methanol (Solvent B) for 5 column volumes; 55-80% Solvent B over 30 column volumes; 80-100% Solvent B over 1 column volume; 100% Solvent B for 5 column volumes. Fractions containing (**4b**) as assessed by TLC were combined and dried by rotary evaporation.

For purification of (**12b**), an Isolera One (Biotage) automatic flash purification system was used. Extracts were separated by normal phase chromatography on a Biotage®SNAP KP-Sil 100 g column with DCM (Solvent A) and 92.5% MeOH

(Solvent B) and a flow rate of 75 ml/min as follows: 95% Solvent A and 5% Solvent B over 4 column volumes; 5-65% Solvent B over 10 column volumes; 65-100% Solvent B over 0.2 column volumes and 100% Solvent B for 2 column volumes. Fractions were monitored by TLC and HPLC-MS. Fractions containing (**12b**) were combined, dried onto diatomaceous earth (Celite, Sigma-Aldrich) and separated by reverse phase chromatography on a Biotage®SNAP C18 30g column with a flow rate of 25 ml/min as follows: 50% Solvent A (water) and 50% Solvent B (MeOH) for 5 column volumes; 50-70% Solvent B over 30 column volumes; 70-100% Solvent B over 0.5 column volumes and 100% Solvent B for 5 column volumes. Fractions containing (**12b**) were identified and combined as above and further purified using a SNAP Ultra C18 12g column with a flow rate of 12 ml/min as follows: 44% Solvent A (water) and 56% Solvent B (MeOH) for 4 column volumes; 56-65% Solvent B over 30 column volumes; 65-100% Solvent B over 0.5 column volumes and 100% Solvent B for 3 column volumes. Fractions containing (**12b**) as assessed by TLC were combined and dried by rotary evaporation.

NMR spectra were recorded in Fourier transform mode at a nominal frequency of 400 MHz for ^1H NMR, and 100 MHz for ^{13}C NMR, using deuterated methanol. Chemical shifts were recorded in ppm and referenced to an internal TMS standard. Multiplicities are described as, s = singlet, d = doublet, dd; coupling constants are reported in hertz as observed and not corrected for second order effects. Where signals overlap ^1H δ is reported as the centre of the respective HSQC cross peak (see SI Appendix, Tables S7 and S13).

Characterisation of *AsUGT91G16* and *AsTG1* oat mutants. Genomic DNA for oat mutants was extracted based on the method described by Pallota et al. (12). Briefly, 50-100 mg of leaf material was harvested from two week-old seedlings, placed in 1.2 ml collection microtubes in 96 well collection microtube boxes (QIAGEN 19560) and frozen at -20°C . Extraction Buffer (0.1M Tris-HCl pH 7.5, 0.05 EDTA pH 8.0, 1.25% SDS) was pre-heated to 65°C and RNaseA (Sigma R4875) was added (27 $\mu\text{g}/\text{ml}$) just prior to use. Extraction Buffer with RNaseA (333 μl) was added to each sample and the wells were sealed with QIAGEN caps and shaken for 2 x 1.5 min at 30 Hz. Spin plates were centrifuged briefly at 3000 rpm (Sigma 4-15). Plates were incubated at 65°C for 45 min before cooling at 4°C for 15 min. Chilled 6 M ammonium acetate (167 μl per well) was added, plates were sealed and shaken

vigorously for 15 sec, chilled at -20°C for 10-15 min, and centrifuged at 5000 rpm (Sigma 4-15) for 15 min to precipitate proteins and plant tissue. The supernatant (400 μl) was transferred into pre-chilled (-20°C) 1.2 ml storage plates with 240 μl of isopropanol, and the plates were sealed and shaken vigorously for 15 sec and then centrifuged briefly at 3000 rpm (Sigma 4-15). Plates were incubated for 10-15 min at -20°C to allow DNA to precipitate. DNA was pelleted by centrifugation at 5000 rpm (Sigma 4-15) for 15 min and the supernatant removed. DNA pellets were washed in 350 μl 70% ethanol, the plates were centrifuged at 5000 rpm (Sigma 4-15) for 15 min, and the supernatant discarded. Plates were dried either overnight at room temperature or at 65°C for 30 min. DNA pellets were resuspended in 200 μl water and the plates sealed with PCR film and vortexed. DNA was allowed to dissolve at room temperature for 1 h or at 65°C for 15 min, and the plates were then vortexed, centrifuged at 5000 rpm (Sigma 4-15) for 20 min and stored at -20°C . The primers used for amplification and sequencing are listed in Table S16. The purified PCR products were sequenced by GATC Biotech.

To analyse segregating progeny, three day-old seedlings of F_2 progeny from crosses between the avenacin-deficient *A. strigosa* mutant #1139 (*sad3*) (9) and the S75 wild type were phenotyped for root fluorescence and then transferred to soil in 96-well trays and grown under glasshouse conditions. Three day-old seedlings of F_2 progeny from crosses between the uncharacterised avenacin-deficient mutants #543 and #1473 (9, 13) and the wild type were also phenotyped for root length and reduced fluorescence and then transferred to soil and grown in the glasshouse. Leaf material was harvested from two week-old seedlings and genomic DNA extracted as described above. DNA fragments spanning regions of the respective single nucleotide polymorphisms (SNPs) were amplified by PCR and sequenced (Eurofins). Amplification and sequencing primers are listed in Table S16.

A

| Gene | Gene acc. no. | Transcript I.D. in Trinity transcriptome | Reference |
|-------------------------|---------------|--|-----------|
| <i>AsbAS1/Sad1</i> | AJ311789 | TRINITY_DN35223_c0_g1_i5 | (14) |
| <i>AsCYP51H10/Sad2</i> | DQ680852 | TRINITY_DN36969_c1_g5_i2 | (13) |
| <i>AsCYP72A475/Sad6</i> | MH539812 | TRINITY_DN30959_c0_g2_i2 | (15) |
| <i>AsUGT99D1/AsAAT1</i> | MH244526 | TRINITY_DN33604_c0_g2_i5 | (16) |
| <i>AsSCPL1/Sad7</i> | FJ475130 | TRINITY_DN35775_c3_g4_i1 | (17) |
| <i>AsMT1/Sad9</i> | JQ071450 | TRINITY_DN27266_c0_g1_i1 | (18) |
| <i>AsUGT74H5/Sad10</i> | EU496509 | * | (84) |

B

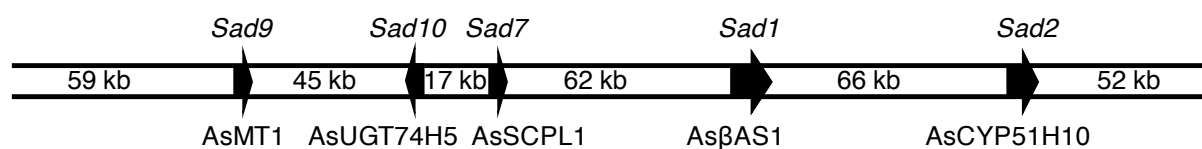


Fig. S1. Characterised biosynthetic genes of the avenacin pathway.

(A) Summary table of characterised *A. strigosa* avenacin biosynthetic genes with corresponding transcripts identified in the *de novo* Trinity transcriptome generated in this study (<http://db.ncgr.ac.cn/oat/RNAseq.php>). *Only a partial sequence was found in the *de novo* Trinity transcriptome.

(B) BAC contig showing the currently known physical distances of genes in the avenacin gene cluster. Figure adapted from Mugford et al, (18).

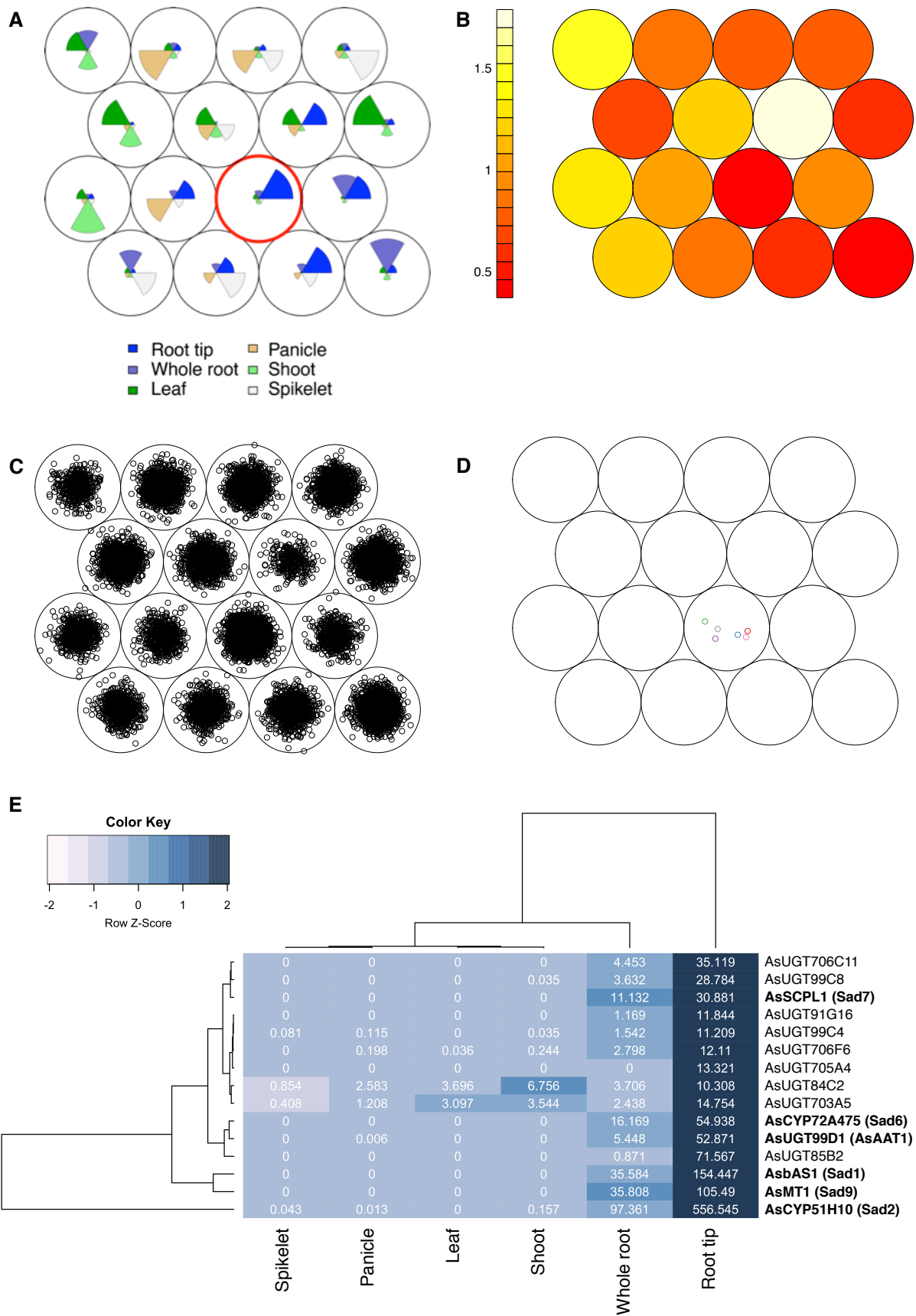


Fig. S2. Self-organising map (SOM) analysis of the *A. strigosa* RNA-seq dataset.

(A) Codes plot of the gene expression profiles for the six *A. strigosa* tissues (root, root tip, leaf, panicle, shoot and spikelet) associated with each SOM node. The nodes containing the avenacin biosynthetic genes is circled in red. (B) Mapping quality plot showing how closely the expression patterns of the genes assigned to each node match the expression profile associated with that node. Nodes are coloured according to the mean Euclidean distance of gene expression profiles to the node expression profile. (C) Mapping plot showing the distribution of *A. strigosa* genes (black circles) across the nodes of the SOM. (D) Mapping plot showing the distribution of the characterised avenacin biosynthetic pathway genes across nodes of the SOM: *AsbAS1/Sad1* (red circle) (14), *AsCYP51H10/Sad2* (blue circle) (13), *AsCYP72A475/Sad6* (green circle) (15), *AsUGT99D1/AsAAT1* (purple circle) (16), *AsSCPL1/Sad7* (pink circle) and *AsMT1/Sad9* (grey circle) (17,18). Data are shown for a representative SOM (seed = 57). (E) Heatmap showing the expression profiles of the previously characterised avenacin biosynthetic genes (shown in bold) (SI appendix, Fig. S1) and the nine full-length UGT candidates that clustered in the same node as an avenacin biosynthetic gene in 100% of SOM runs (SI appendix, Table S4).

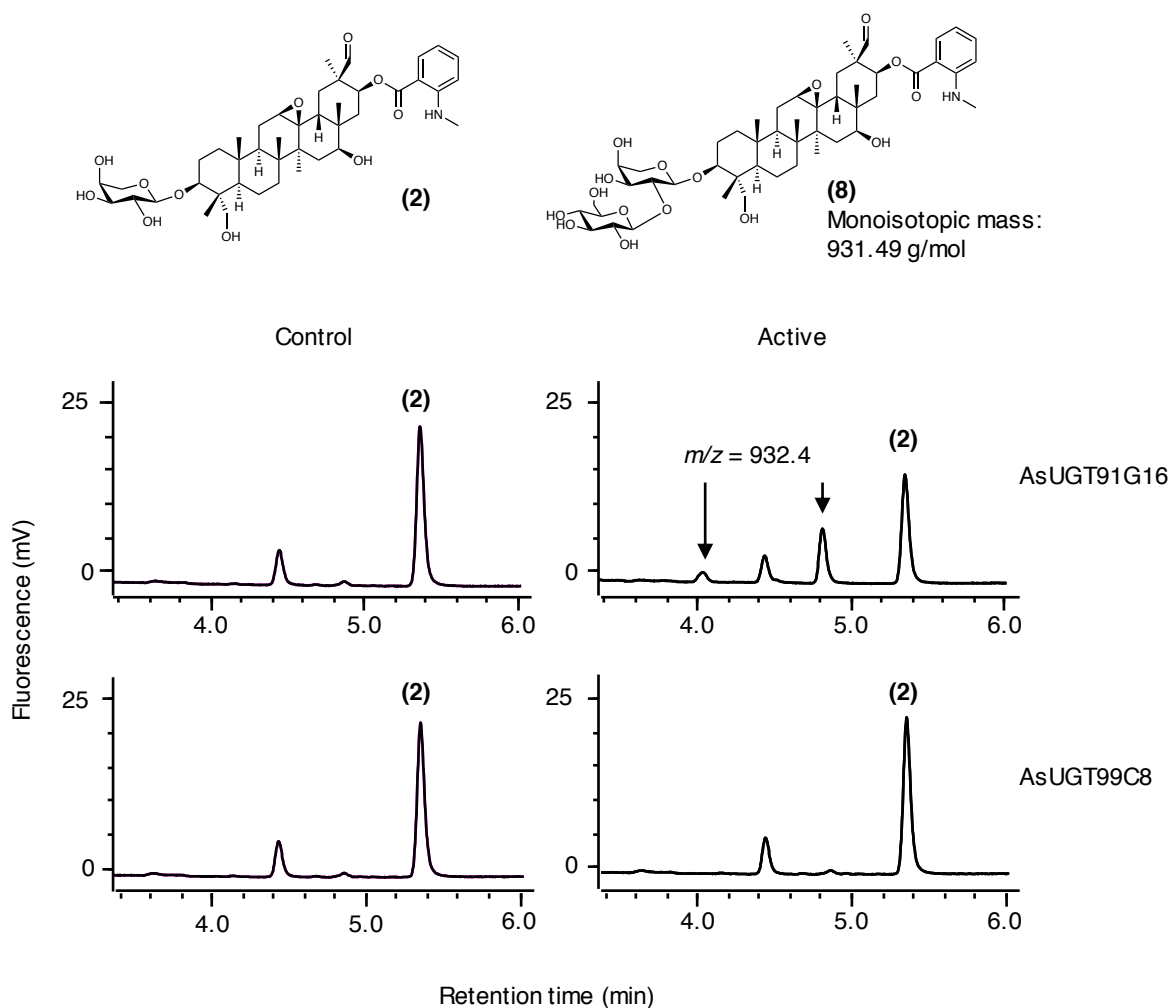
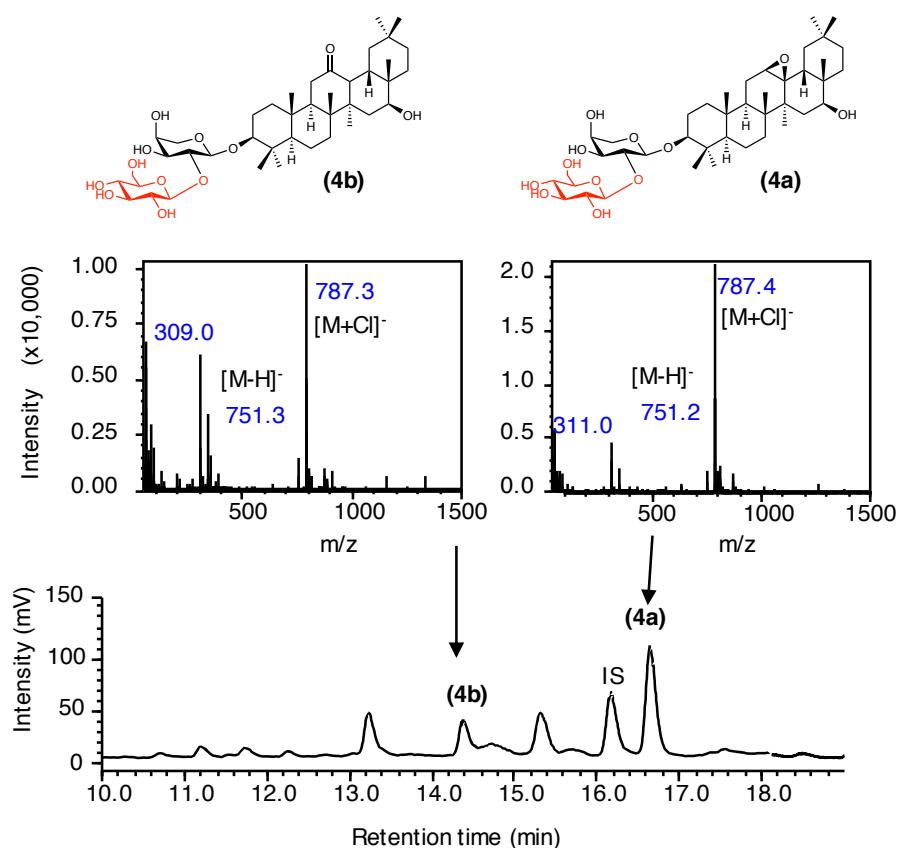


Fig. S3. Enzyme activity assays of the two recombinant *A. strigosa* UGT glucosyltransferase candidates.

Deglucosylated avenacin A-1 (2) (SI Appendix Table S1) and UDP-glucose were incubated with protein preparations of AsUGT91G16 or AsUGT99C8. Control, boiled protein. Data were analysed by HPLC-UV-MS, and fluorescence monitored at 357 nm (19). The enzymatic reaction with AsUGT91G16 resulted in new peaks at 4.1 and 4.8 minutes, both with m/z values of 932.4 consistent with the addition of D-glucose. The peak at 4.1 minutes is likely to be a degradation product of the peak at 4.8 minutes, generated by loss of the unstable C12-C13 epoxide group. The peak at 4.2 minutes is likely to be the degradation product of (2). Results are representative of two separate experiments.

A Small-scale extraction



B Large-scale extraction for structural determination by NMR

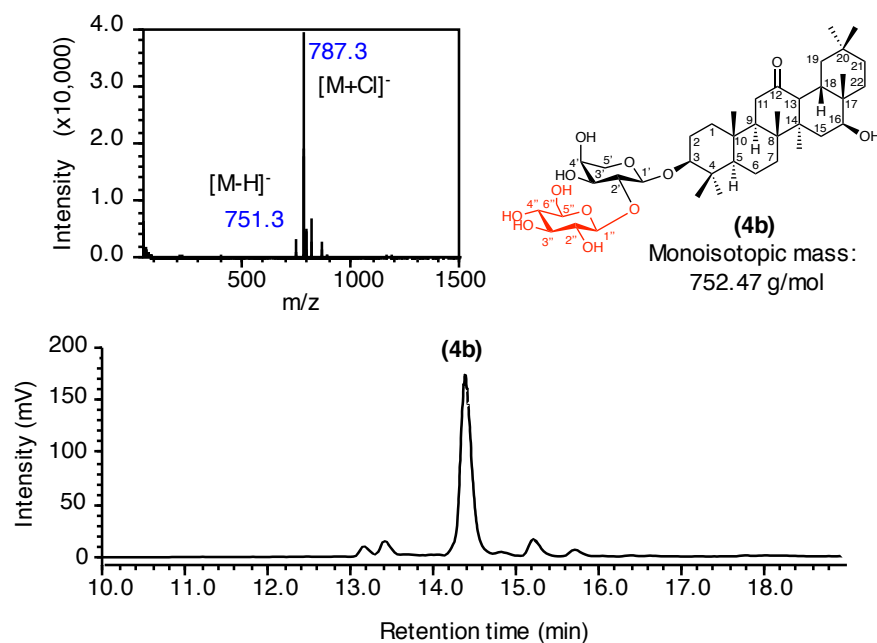


Fig. S4. Analysis of AsUGT91G16 products obtained by small and large-scale transient co-expression of AsbAS1, AsCYP51H10, AsAAT1 and AsUGT91G16 in *Nicotiana benthamiana*.

(A) HPLC-MS-CAD analysis of methanolic extracts following small-scale leaf extraction (20 mg of dried leaf material). The major peak is **(4a)** [$m/z = 751$, $R_t = 16.7$ min]. **(4a)** is likely to be the AsUGT91G16 product with an intact 12,13-epoxide. The 12,13-epoxide is known to rearrange to form the C-12 ketone and it is likely that a proportion of **(4a)** has rearranged to form **(4b)** during the extraction. IS = internal standard (digitoxin). **(B)** HPLC-MS-CAD analysis of methanolic extracts following large-scale transient expression. Leaves of 150 *N. benthamiana* plants transiently expressing *AsbAS1/Sad1*, *AsCYP51H10/Sad2*, *AsUGT99D1/AsAAT1* and *AsUGT91G16* were freeze-dried and extracted by pressurized solvent extraction. HPLC-MS with Charged Aerosol Detection (CAD) analysis showed that the major product that was obtained and purified was **(4b)** ($m/z = 751$, $R_t = 14.4$ min). The structure of **(4b)** was solved by NMR and shown to be 3β -{[β -D-glucopyranosyl-(1->2)- α -L-arabinopyranosyl]oxy}-12-keto,16 β -hydroxy- β -amyirin. As predicted, **(4b)** has the triterpene scaffold 12-keto group instead of the expected 12,13 β -epoxide, presumably because of the instability of the 12,13-epoxide during the extraction and purification process.

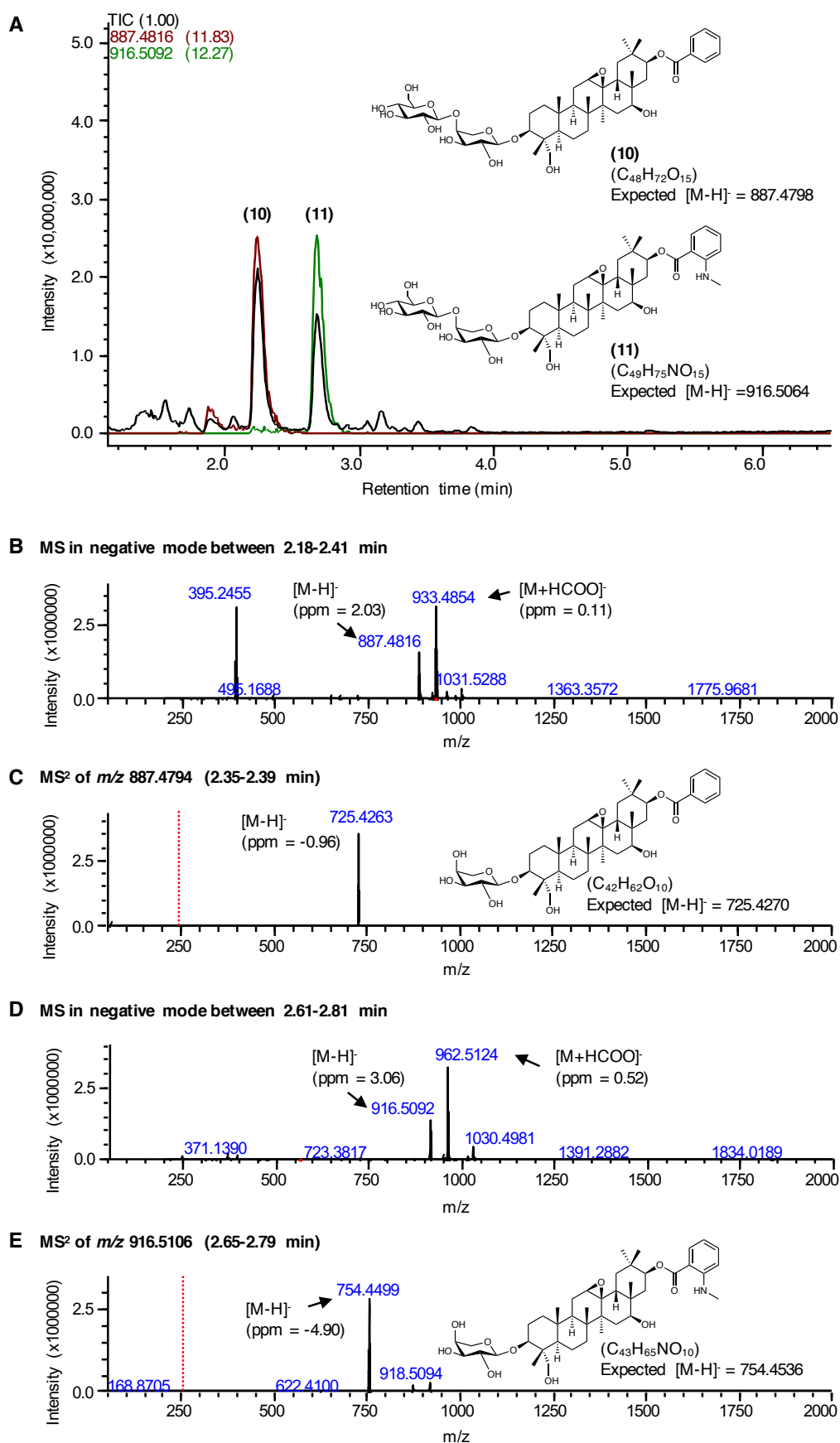


Fig. S5 HPLC-IT-ToF analysis of root extract of *ugt91g16* mutant #85.

(A) Total ion chromatogram (TIC) of root extract from the *ugt91g16* mutant #85. The two major peaks at 2.3 and 2.7 min (**10** and **11**) have masses consistent with avenacins A-2 and A-1 respectively, with loss of a D-glucose and oxidation at the C-30 position. The predicted structures for (**10**) and (**11**) are shown. (B) The mass spectrum of the peak between 2.18-2.41 minutes (A) showed signals consistent with the mass ion ($[M-H]^-$) and the formate adduct ($[M+HCOO]^-$) of (**10**). (C) The MS² spectrum of the fragmentation of the mass ion of (**10**) (precursor mass ion for fragmentation = m/z 887.4794) showed a signal consistent with the loss of a D-glucose molecule from (**10**) (proposed structure shown). (D) The mass spectrum of the peak between 2.61-2.81 minutes (A) showed signals consistent with the mass ion ($[M-H]^-$) and the formate adduct ($[M+HCOO]^-$) of (**11**). (E) The MS² spectrum of the fragmentation of the mass ion of (**11**) (precursor mass ion for fragmentation = m/z 916.5106) showed a signal consistent with the loss of a glucose molecule from (**11**) (proposed structure shown). The difference in observed and expected mass values in parts per million is indicated (ppm).

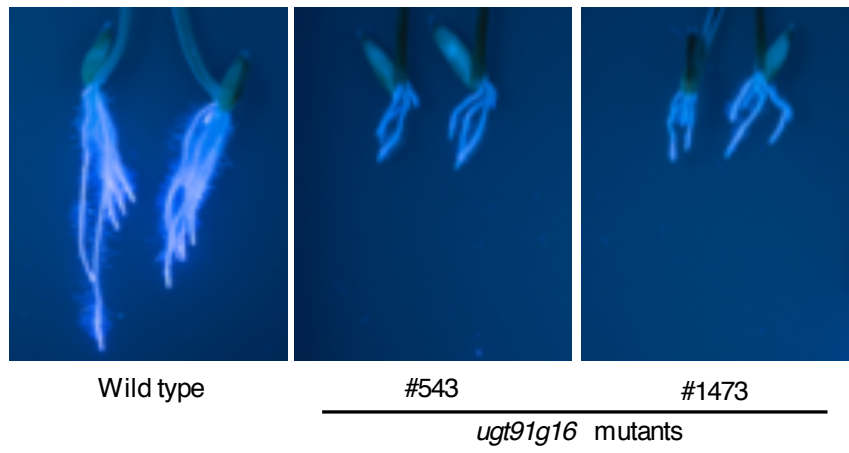


Fig. S6. Reduced fluorescence of roots of *A. strigosa asugt91g16* mutants #543 and #1473 compared to the wild type (S75).

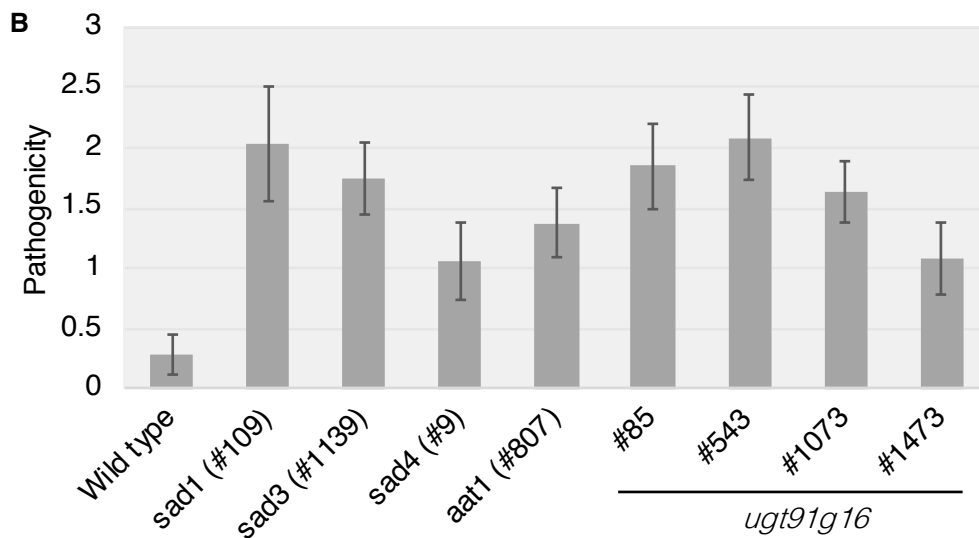
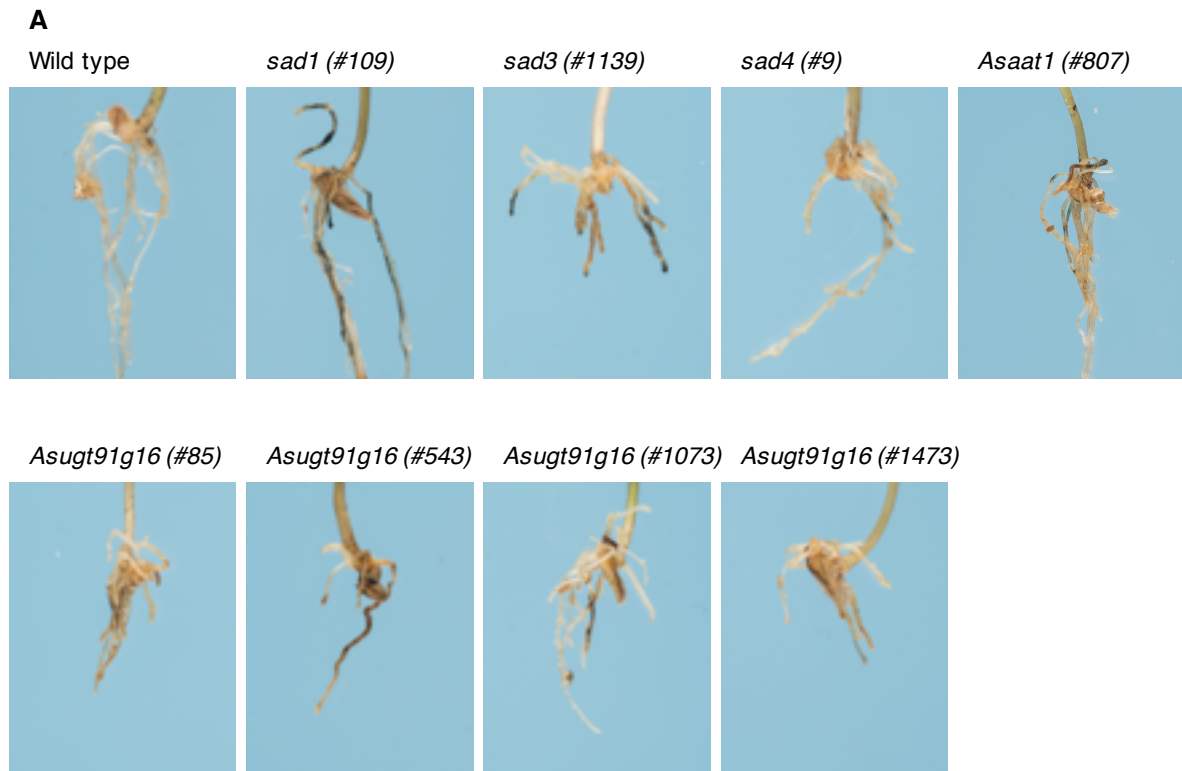


Fig. S7. Take-all disease symptoms of wild type and mutant oat lines. Seedlings were inoculated with the take-all pathogen *G. graminis* var. *tritici* isolate T5 and scored for disease symptoms 21 days later as described by Papadopoulou *et al.* (8). (A) Dark brown/black lesions on the roots are typical symptoms of infection. Images

are representative of 31-35 biological replicates from two independent experiments.

(B) Quantification of disease symptoms. Symptoms were scored using a seven-point scale (16). Mean disease scores (31-35 seedlings per line) are shown. The error bars represent 95% confidence limits.

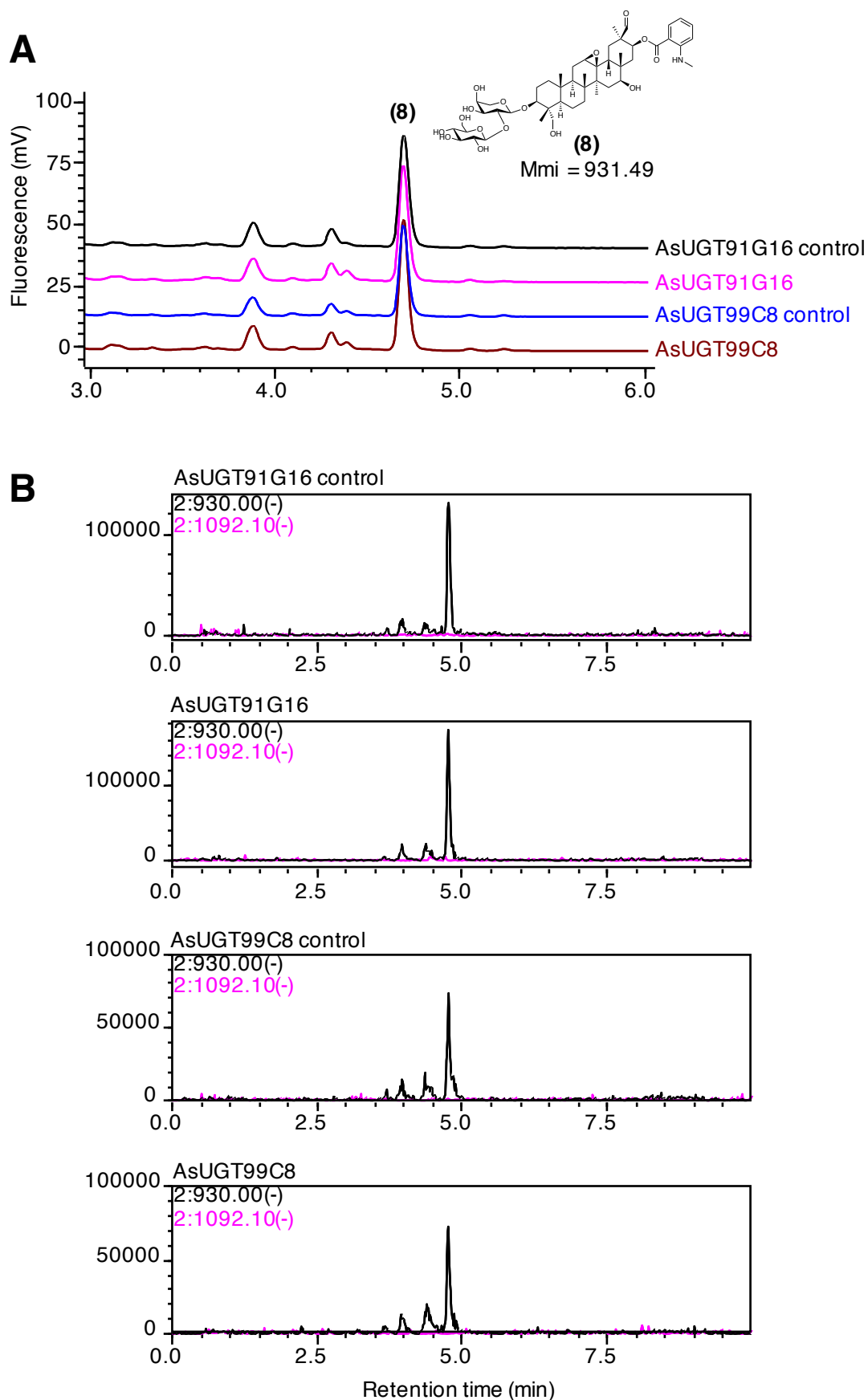


Fig. S8. *A. strigosa* UGT candidates do not show 1,4-glucosyltransferase activity. (A) Avenacin A-1 with the 1-4-linked D-glucose removed (**8**) (SI Appendix,

Table S1) and UDP-glucose were incubated with protein preparations of AsUGT91G16 or AsUGT99C8. Control, boiled protein. Data were analysed by HPLC-UV-MS and fluorescence monitored at 357 nm (19). (A) No new UV-active peaks were observed for either UGT and (B) extracted ion chromatograms did not identify any novel peaks with the expected mass of (**8**) with the addition of D-glucose ($[M-H]^- = 1092.1$).

CTGTTTGCATCGGGATTCACCGTATTAATTCAGTTCAGCCGTTCCAGTCAAACGAGAGGAAAATATGGAGGCTGTTCATCGGAATCC
 TAACAGTTTTAAATCAAGATCACTTGCATCGGGATGCTTTCAGATTTGATTCAGTTCAGTCAAACGAGACGAGAGCTTGGCAGCTTGT
 TTGTGTCAGGATTCACGTCATAAGCTTCGTTAGATCAATGGAGGCGAGAGACATCAAGATTTGTTGTTCGGTATTCATAACAACT
 AATTCATCAAAAGTCAAATGAGACGTGAGCTACTCTCAGTTTCGGTTTATAAGTCATGCGCCTATATCTAAGTTAGACAAATTTAAATATT
 AAAATGTAATAATGTAATAGTTGTTAGATTTTAAATGATATAACTTTTATGTATATAATTTATTTTTTCAGTGGTCAAATGAGAATCT
 AGGTATACGTCGAGACTTGTAAAGACTGGTCAATAGTGAAGTATCATACTAGTAGTATCATGCATATGATACTAGTGTATGATACTAGCT
 CTATAGTGTATGGCATCATATGTTAGTACTCCCTCTGTTTGGGAATACTAGTTCGGACTATTTATTTCAAATGGAGGGAGTATCATAGTA
 GTATTATATTTAATGATTTGTAAAATTTCTAGTTTTACGTGCTATGATACATAATTTAGCTATGATACCTAATCATCTCTTCTTCTT
 AATTGTTGTGCCACATCAGATTTTGGCAACATGGTATGCATAATAATTAGTTATGATACTCTCACTAGGACTAGCCTAAACTGAAATGGA
 GAGATACCAAGTTGTTTGTCTCGAGATTCAGATCATTAATAAAATTTGGTCAGACAAGAAAGATAGTTAAAACATTTATAGTTCTCAG
 ATTTTCAAATGCAATGTTTCAATAATTAGATCTACCCCGTTTGGTGTAGTCTACCCCGTTTGGTATGTTAAATAGGAGGGTTCGTTGGTGG
 CATAAAAATTTGCGGTAGGTCATGGTGGCGTGCAGGGAAGGCTCATAGGTCATGGTGGCGTGCAGGAAAGCCCATAGGGTCATTTG
 CATTGATCCCTATCTGAAGATGGTGTGTAAGAATGGAAGCGGTATCTCTGTAGCAAGTGCATATGGTGTCTGCTCGGGTCTGCTG
 GACCGGTGCACCGTTTGTGTAGTGTCTTGGTGAAGGTTAGGGCATACGGCCTTGATTCGGCCTCCCTTCAATGTGTATTAGTGT
 TTTATTCGCAAAAAGGGTGGCTTCGGGTGATCATTATAATATCTTATAAGGCTTCGAGGAAATAATTAATGAAAGCATGTTGTGTGCATC
 TATTTGATGCATAGGTCGGGTTACACCTCCATTCGAAACAAAATCGATCACATTAACACCCATTTGGTATTTGGCTTGACAGTT
 ATCTAGGTGGTAGGTCACATAAGATGACGCCACATGAAGGTTTCATATCTTTGGATTTTTTGAATTTAGTTGTGCATACCTTACCCT
 AAACACTATGGGTACACCTAATCCCTAACACTATCATATTTCTAGGAATTTTGGAAAATAATTTGATGAAATGTCATATTTGTGGACCTA
 TAAATGTCACCCCGGTGCAAGGAGTCTAGTATATACTAGACCTGCTACTTACCAGTTTTTAGTCTCATGGTGAAGTCACTGCTGGCTTAC
 GAGAGGCGATGGAAGCCGAAGATCTGAGGACGACTAAGGTTGGTTCGAGTCTGGTGTATGATGAGGTTTCGGCTAGGTGATTTGTGACTT
 GGAGCAGTGACAACAGCAAGTGGGGGTGACAACACATGAGGAATTCGCGTCAATCTTACAGGGTGAAGCCCAAGCTGCTGGCTTACG
 TTGGTTGTGCTAGCAATAACTTGTGGAGGACTTGTTTATGATTTTCAAGCCTCTCCTGGGTGAACCCCTGCTGATGATGAGG
 CGATGACGGCGTGTGCATTTCTTCTTTCGAGCGTCACTTTTAGAGAAGTTAGATTTCCGGTGTTTTTTCTGTTGGTGGTGTGTCGC
 TGCAAAAAGGAGTAGATCACTATAACAGGACTTTCTTCTTCTTCTGTAACCTTTTTTTTTTGGACCGTGTGTATCGTCCGTAGGTCGCA
 AAGCCATGCGTGTGTGCAAGATAGATGTAATTTGGTATCTTCGCGATATTAATATACTCTTTTTTAAAAAAAATCCTGGTCAITGA
 CAGGTGTAACAATGGGCACATGATGATCATATCAGCATCAGCGTGAATTAACAAAGTGAATGAAGTTAATAGAACCCATATAAGAAAGAAA
 AAAATGAAGGAATGCAATGCAACTAAAAAAATGTTAATCCAAATCGTGATCTTATTTATTTAAGGATTTGATGCTTCCAGGCTAG
 TGGGATCTTTTACGGTGTCTTTAACTTATTATACATGGTGAAGATTCACCACCGAAAAAAAATAGACGAGCTAGCTANNNNNNNNNN
 NNN
 NNN
 NNN
 NNN
 TGCATCTCCCTAGCT**CCAGTCCCGTGGCCATGGCACTGCTGCTCTGCTTGTCTTCTCAGCCTCCGGCTGGCCGCTCTCGGGAGACGT**
TGTGTTGGCGCGNN
 NNNNNNNNNNNNNNGCTAGCTACTATTTTTCGCTCCTCCATCTATCGTGAAGCCGAATAGAAAGTAGTAAATTTGTATGAGTGCATCAGGT
GGAAAGGGCGGTTGGCGAGGATGGAAAGAAAGCCAGCAATTTAGTCCAGGACACTTTCAGTCCAGGAGGTTACTCTGCCGACAATCTACCGAGA
TGTTGCGAGCAGATCAGTACCATCATTACAAGGTGCCCTCGATCGTTTACTCTTGATAATTACACTAGATGAGTTTCTTCATAATAACT
 AGATGAGTTTCTTATATATATATATAG**GAGGACGTAAGCTTATGATGAGATGGGCTAGACGCTACAGATCTCTCATCGCATGGCC**
GCGGCTTATCCAGATGGAAGAGGAGATGTGAACCCAAAGGGATGGAGTACTACAACTTGAATGATGAACCTACCGCCATGGTAA
 TATAATATAATATACCTTCACTTCTTAAATTTGAGGATATAAGCAAGCCCGCTTCAACCCGGCTGATCTGATCTGACTGACTG
 ATTGATTGCGAG**GCATCCCAACTCATGTCAAACTCACCATTTGATCTCCCTCAGTCCCTCAGGATGAATACGAGCGACTGCTGAGCCC**
GAGATTCGTGTAATTAATAAGAACTTTTCCATGCATATATCGATCGATGCTACCTACTATCTCTGCTTGTGTTGGTGTCTGATGAAT
 TGGGATGTGCTTGTATATATGATGCAG**GGATGATTACACCATGTTTCGCGGAGACGTTTTCAGAGCTTCGGGGACCGGGTGAAGCACT**
GGGTGAACCCGAAACGTCGAGGGTTFAGGCGGCTTCGACACCGCCACCATGCGCCGCGCCGCTGCTTACCCTCTCGGG
CCAACCTGCACCCCGGAGACTCCACAGGGAGCTCACAGGGAGCTACATCGGGGCGCACACACTCTCTGCTCGCGCACGCGCCGCTCCCTTACA
GGAGCAAGTACCAGGAAGCACAGAGGGGCGAGTGGAAATCACCTGCTGGCCGCTGGTACGAGCCTGCCACCGCTGCGTCCCGGACG
ACGTGGCTGCCGCTGCCAGGATGAATGACTTCGAGATCGGATGGTTCATGATCCTTTGGTGTATGGGGACTACCCGCGGTGATGAAGA
CCCGGTCGGCGGAGGCTGCGGCCATAACATGGACCTCTCCAAGATTTGACCCGGATCGTTCCAGCTTCACTCGGCTGAAACCATACC
TCATCTTAAACGCAAGCCGACGAGCCGCTTCAATCTCAAGCAGGACTACGCGCCGACACAGCAATAGCAGTAACCTAAGTAC
 TAGTATCAAATTAGTTACAAAGCCATATAGCTAAATGTATATAAACTTTCTGATTTGGATTTGGATTCAG**ATGCAATGAAAGACATCCAA**
GAGGTACGTACATACGGTGAAGCAAGCAATGCAAACTCTCTCTTTTACACAAAGTTTCTTGGGCCACGGTAAATTTATATATGTTTCT
 TATCAATGGTTGCAG**GGCCACGGTAAAGTACCTCTTGGGCTCTCGGGAGTCTACTCGACCAATCGGTTGTCAACTACGGAATCTCTCT**
ATCATCATCCATGAAAACGTACTTGTATGCAAAATTAAGGTTTGTGTTTCTTACTGTATATAAGAGATGATGATGGACTCTGCAAA
 ATAGTTTAGGGTGAATAATACATATGTTTACCTCTGTA AAAACAAAATAAATATAATTTGTTGATACTTCAATTCAG**CGGGCTGACTT**
CGTCAAAGACCCGAGCAGATCAGAACCGAGACTACCTAGATCAGAGGCTCTGCAGGACTACTTGGAGGTTCTCCAGTGTCCATACG
 GTGAGTTTGTTTACTGCTTATTTTAAAAATGCCAGAGCCACATATATATACCCAAAAATGATCTGCAAACTATCAGTTGAAACA
 CTTCCTTGAAGAAATCAAAAAGAAATCTGTGACATCAGGTGGTCTCGATCTTGGTGTAGTGTCTGATATGACAGAGCTATACA
 TGTCTTGTGTCAG**TAAACGATCGGACGCGGGGATACTTTGATGCTTCTTGGATCTACCGTTTTTCTGTTAGTTTGTGTTTCTAGCTT**
CGCTTTGGCCTGATCGGTGTCGACATGACCGTAAAGGAGAGGACGCGGTAATGTTAGGAACCTCGCCAGATGGTACTCTAGCTTCTTAAAC
GGCGCGGAGCTGCGGCAACCGGGCGGCCCAAGAAACAATAATTCAGACTCTGCGTGATGAGGCCCTAAGGATTCATGCTGCATCGCCGG
 TCTACTCCGCTCCCGCGTGGCCTTGGGGTCTTGGAGGTGTGGTGGACTATGGCTCTTTGTGCGCAAGAGGTTTATCGTTCTTACTTA
 GGCTTTTTTAGTGTCTTTTTTGAATTTCTGAGGCGGAGTTATATCTGATGTCAGAAATAAGTTATTTCCACCTTACCTGCGACATC
 GCGGATGGCGGTGGAGTTGTATCAGGCGGCTCTTTTGGATCTACCGTTTTTCTGTTAGTTTGTGTTTCTAGCTTCAAGTTTGTGCTT
 TCCAATCTATATTTGTTATCTTTGACAATGGTTGCTGCTGGTGGTGTGCTTTTCCGGTCTTACGACGATGATTTTCTGTTGTCTA
 CTATAATAAGCTCTTCTACGACAAGTTTGGCCTTCTCTGGGATGACAGGCGAGAACGGCTCGCACATATGTTATAGTCTGCTGCTAG
 GTGGTCTAACAACTATTGTAATAATTAATACCTTTAAGTTCTTTATACCACCGTTGATGAATAATTAAGATCGGTGAATTTTTGTC
 AAAACAAGTTTGGAGCCGCGGTTATTCATACCGCTTTTTTACTCTCTTATAAATAGCCATTTTACATGGGTTAGGTTAGGTTTGTGCTT
 TTTTCTGTAAGCTAATACTGCTATTTGCTTCTGACGTTGCTGACGTTAGTTAGTTTCTAGCTCAACTTTCAGTTTCTGCGAGAA

AATAAAGATATTTTATGGTATGCATATGGTGAACCTATGTATTTTCATTTCTTGTTCAGATGCTCCGAGTTTTGACAGCCGTGTAGTGTG
 GTATTGCTGGTGCATGACCAGAGCTAATGTTACACCAACCTTACTTCAGTAGTAAGGTGGTGGTAAACATTACTATTTAAATTTTAA
 GCCAACATCTTTGTCTTACTATTTAAATTTAATCTTTGGGTGTTGTTACTTGTGTTAGATGAGTGGAAACAGAGTACCAGTCCGGTGGTCA
 GAATGTGATCCACAAGGCAGAAGTGTGTGAGCACTATGCTTGCAGGGGTTTACCCTGGGCAAGACGCCGTCGGCAGGGCGAAATC
 TTTCCAGCAGAGACGCATGGCTTACGTCGACCGCCAGTGGCAAGGTGGCTCCATCGACAGGAAGATGGGCTGAGCGAAGATTACCAT
 GGGCACCACCGTGGTGAACGAGAAGGTGAAGGAGATGGACCAGAAATTCAGGTGTCGGACAAGACCAAGTCGGCGTGGCGGGCGGA
 GCAGACGGTGAACGCGGGTCCGATCATGAAAGAACCGGTACGCTTTCACGAGCGCGTGGTGGTGAACGAGCGCTTCCGCAAGGT
 GGCAAGGCTGCCACGAGCTCGGACGATGACCAAGGAGAAGATGTCGTGTCGAGGAGCAGCAGCAGAGGGGCTCCGGTTCGTCTGG
 AGGAACTCGTACACGCCCATCGTTGATTTGTCTCTTGTACTGCCTCGCTCGACCGTGTGTACTGGTGGTTCGAAACATAGATATAG
 AACACATCATATCTATCATGTGTGAACCTCAGGGGACCCAGGAAATTTGTTTCTCTCTCTCGGTAGAAGGAGAGGAGATGTGCAGATG
 ACCACAGTGTATTTGGTGGTTTTATATACGAGAAATGTACAGGTAGGATGCTGGGAATTTGGAAATGAACATTTTACTTTGGTCAA
 GTTGTGGAAGGGTGTGTAATAATCGTGTGTTCCCTTGGTGTACAATAACCTGCTCTCATCAGCCAGTCCGCAATTTGACGCGTAGG
 ACAGAGATTAATGTCCTTATAATTTAATTTTGGCCTAGAAACACATTTGTGAGCTCCAAATGATATTTAGAACATCTACCAGAT
 ACTACTGTATTTGCTGAGGTGATGAATCATGATAGTAGTACATGAATGCCAATGTGTGGAGCTCCTTATGGCAGTATGGAAGTGCATAC
 ATGAAAGCATCCATCACAATGGTGGTGTGACCATATGGTATCAAACGGTCCGACACTATTTGTTCTACTTAAACATCGGAAGATAGAT
 GGAGCATGATCCTACTCACAAGAACCATCTGCTGATACCTAAGGCCGAGCAAGGGTAAAGTGTAGGTTCCGCAATTTGCTATACCA
 CATGCAATGCGCAATGTTCAACAATAGTATTTGTTAAGTTCATATACCAATGCAACACATAGGACTCAAAGGATTTGGAAGCATCAGGAT
 TTCACAAATGGTTGAGCTGCTCTCATCATCAAAGTTCATAAATCTACTTTCAATCCGTACGTGTGCACATATGGAAGCATCAGGAT
 TGGTCACTCCAATGCCGTATGCTGGAGAGACAGGTCAGTACAACGCATTAATTTGTGTGACACGTTGTCGATAACGTCGTTGGCGCTGT
 TTTCTTTTGTGTTTGTGTTTGTCTATCGTGAATTCATAATTTTATCGTATCATTTCTAGCGCATTAGCGGCTGGGGGGAATTCCTCA
 ACATATCGTAGTAAGCAAAAATACAAAAAATACACCCGAAAAAGCAACTGGACATATAAGCTTAGCTAGTGAACAATAAAAAAC
 CAGAAATCAAAATGACACGGGTGCACAAGTACGAGAGACAAAAGCTAAGACCTGACGTTAGTCCACATTTGAAACTGAACAGCTGCC
 AAATGACAGCAAGTATCAGCACCGTCTGCATCTTACCTCAATACAGGATCCCTGATATAAATAAGAACAGTTGTTGGCAGAAGGATAAA
 CATTTATACGGCAAGGGATTAATAATAATTAATCATCAGAGCTGTGTAGTATAGTGGTAAAGTATTCGCCCTGTCAAGCGGTGACCC
 GGGTTCGATCCCGGCAACGGCGATTTTNN
 ACATATCGTAGTAAGCAAAAATACAAAAAATACACCCGAAAAAGCAACTGGACATATAAGCTTAGCTAGTGAACAATAAAAAAC
 ATCTAGCCAAACAAATACTTGGACCAAGATTAGGAACAAAAGTGACCTCATGTGCAAGTTGAAAGCACCTGGTGCATTTAAGTTTCT
 GAGTTAAACAATAACCTGTATGTTTATGTAATCTAGTCTTTCTCTCAAGGCCCGCCCTGCTTGGGAGGTAGTTTTATTTCCAGGTCCG
 TTTTGTGTTAGGGTGTATTTAGACGTGCACCTTCTTACTGCTAAATCCAGTTTGTCTCTGCTTCGTTTGTCTACAGGGGACGGCTCAGG
 CCAGTTTGTGTTTCTGTTCCGGTTTGTCTTACAGGGGAAATTTGGATACACAGTACAGCTTACAATTAACAGCATACACAGTACATCAAAA
 ATTGAATCTTTCAACATACACAGCTTAACTTACATAAAATCATCACATGTTCAATTTTAAACAGTTATCAATGTTTAACTTGCATTTCTGAT
 GAAATATAACATAAATTAATAGCATACATGANN
 TTTTCCAGGCAAAAACATGATGTTTAAAGCCATAGCATCCAGTATCTTCCATTTCCATACCAGAGCACAGGCATAAAGCGACCT
 CTTCAACTCTGCAAGAAAGATCGACACATATACAGTATGAAACGACCAAAATGATTCATTTGGTATACTACAACCTAACACACAGAA
 CTTGAATCTGTAAGATGGTCAATGCAATGCACTTGTTAACACAATAGTTAGTATGTTGCTAGCAGATAGTGTCTCATGAGAATAAAT
 AATGGCAGGTTCTCCGCTGGAATATGCTGCAATCAAGTTACAAATCCCAAAGAGGTTTACTGAATATAAAGATGAATATGTTACAAAT
 CTCATCTGATGACGACGACTCCAACCTGATTCAGTTTTCAAAACAGTAACACAGTACTACTATACAGACATAAACCGAACTAA
 AATCTGGCAGTTCCAGGCTGCTTTTCTCAACAAAGCTAGTATGTACCAGGACCAAAGCTAGTATGTTCCAGCAGCAGATCAACAGATG
 ATTTTCAAAGCTAGTATGTTCAAAGCTAGTATGTTCCAGCAAGCTAGTATGTACCAGGACCAAACAAGACCAAAGCTAGTATNNNNN
 NNN
 NNNNNNGACCATTGTTGGTCTTCCCTACAGAGCACACTACAAACCTCCTATAAGTTCCTCCTAGATATACAAGGGCAGCTGAAAA
 ACGCAGAGAAGACAACCGATTAATTAATACCAAACTCATGCAAACTAAGACAAATTAATTAGAAAACTCAACCACCCTTTTGTGTTT
 GTTCTAATAAGAAAAATAAGCCATAAAAGAAAAATCAAATATGCTAACTCGTGTATGAGAAGGATGTAAGGAATAATTCAGAAAGGTA
 CATAAGCGCCATGGAGGTTTAAATGACAAGGTAGAAAAATAAAGTAGTACTCTACAATGAATATGACAAGATGAAATGGTGGAGTTAA
 CTGTACTCTACAGGATGATGAATACCTCAGAGTGTGCAAGACATAGATGCAAAATGCATGAACAATACTATCGTGTACCCTCGCATTTTAT
 TGCAGTCTGTAGAAGGGCACCATAATGATTAAGTGGACCATCATGGTAAAGCAAAATCAATCTTGACCAGGAGAAAAAATAG
 CGCTCTAGACATTTGCCACAGTATCACACCTCATTTATGTTCCCTGACACGAAATGACTCAGTTTGTAGTATGATAAATGTTTCAATGAAT
 CTAATAGAACATCTTAGATCTGAACATTTAATCTTCTTTCCAGATAAGAAACCAACCAACCAATGAGTAACTAAAAATGTTGTAAT
 GGAGAACACTACATTTGCCAATCAAACTTATCCGGAAGTAAATATACTAGTGTCTTCAATTAATCTGGCATAATTCATGGAATCAAACTATA
 GTGCTTCAAACAATGCCACTCTGCAGGCTCTCCATCACCTACCTATCTTAGGATTAAGGTAAGAAAGGACAAAGAAACTCGTTTCAG
 AAAAGAAAAAAGAGAAGAAACAGTTTATCATTTGAGTCACTAATGGGAAGAGTCAATGTAGGCTGGCAGCATCGGTAGGAGTAATA
 TCCTTGCAAAATCCGTGCAACTGTAGGCTCGTGGCATGGCATGTCGCACATGCTCTACATCGACGTTAGAACTAGAATACTGGACA
 CCGTCCGGGACTTTGCAAGTCAAGTCTCTGACTGCTCGAACCAAGCAGCAATCTCGCAACAGTGGAGCGCAATAATACAAAGAAAGG
 CAAGAACTCTAGATACGAAGAAGGTTTCCGCTGTACAGGTAGTGTGGTGGCGGAGGAGTGCATGATCTTGGCCGCCGGACG
 GGCTGGAGTGGACCTGGAGAGCCAGACTTGSACA

Fig. S9. Sequence of the scaffold containing the *AsUGT91G16* and *AsTG1* genes in *Avena atlantica* accession Cc7651. *AsUGT91G16* is highlighted in red and *AsTG1* in blue.

| | | |
|-------------|---|----|
| DgAA7BG-GT1 | <u>MGVMKIA-YLVL--DLFVVFNSIIFIPKPANP-NQDSSAFDRNNFPVNFTFGVSSSAYQF</u> | 56 |
| DgAA7BG-GT2 | <u>MGVMKLA-YLIF--DLFVMFNPIFFIPKPADHTELDSSALNRKSFVNFTFGVASSAYQY</u> | 57 |
| Os9bglu31 | ---MTPA-----RVVFICCVLL-AAAAAASSSTAAGITRADFPPEFIFGAGSSAYQV | 50 |
| AsGH1 | -----MALLLCLFLFSLRLAALSGDVVVAALTRRDFPDGFIFGAGTSSYQV | 46 |
| DcAA5GT | -----MNMSCKFEIVLLVSWWLLLVVFGVSSMFSEFDRLDFPKHFIFGASSCAYQV | 53 |
| CmAA7GT | ---MLTONQLK--HLHLLLIVVGVCTNNW---DLTLADYSRLDFPSDFVFGAGTSAYQV | 52 |
| AaAA7GT | ---MI-S-Y-----SLFFLLAFLFLYLVEFGISQSNAPKFSRDDFSSEFVFGAGTLAYQY | 50 |
| DgAA7GT | ---MCPS-FLVT--LLLLQLSSLVVVVLVW-AEQLPEFNVRDDFPSNFVFGAGTSALQV | 53 |

Fig. S10. AsGH1 (AsTG1) and other characterised GH1 transglucosidases have predicted N-terminal targeting sequences. The N-terminal region of a Clustal Omega alignment (20) of AsGH1 (AsTG1) with other GH1 transglucosidases is shown. Predicted N-terminal signal sequences are underlined. DgAA7BG-GT1, *Delphinium grandiflorum* (GenBank accession number BAO04178) (21); DgAA7BG-GT2, *Delphinium grandiflorum* (GenBank accession number BAO04181) (21); Os9bglu31, *Oryza sativa japonica* (GenBank accession number B7F7K7) (22); DcAA5GT, *Dianthus caryophyllus* (GenBank accession number E3W9M2) (23); CmAA7GT, *Campanula medium* (GenBank accession number BAO96250) (24); AaAA7GT, *Agapanthus africanus* (GenBank accession number BAM29304) (25); DgAA7GT, *Delphinium grandiflorum* (GenBank accession number E3W9M3) (23).

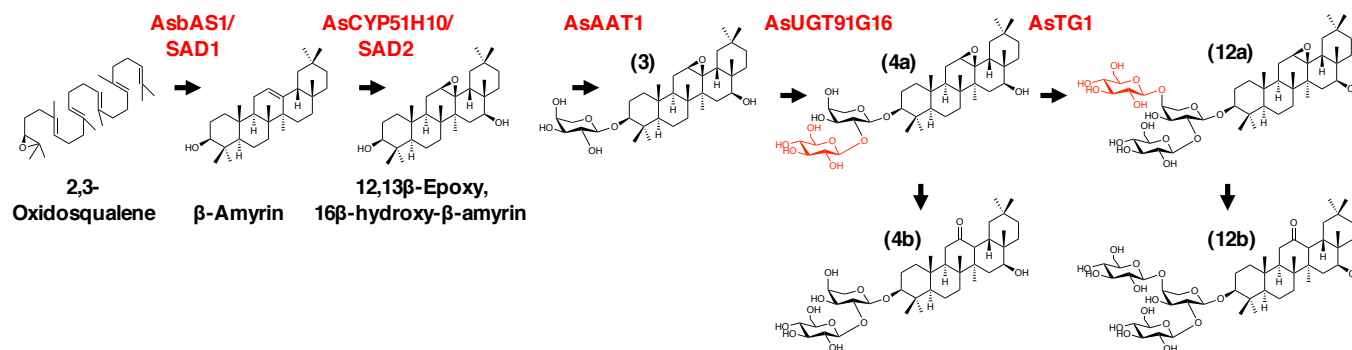
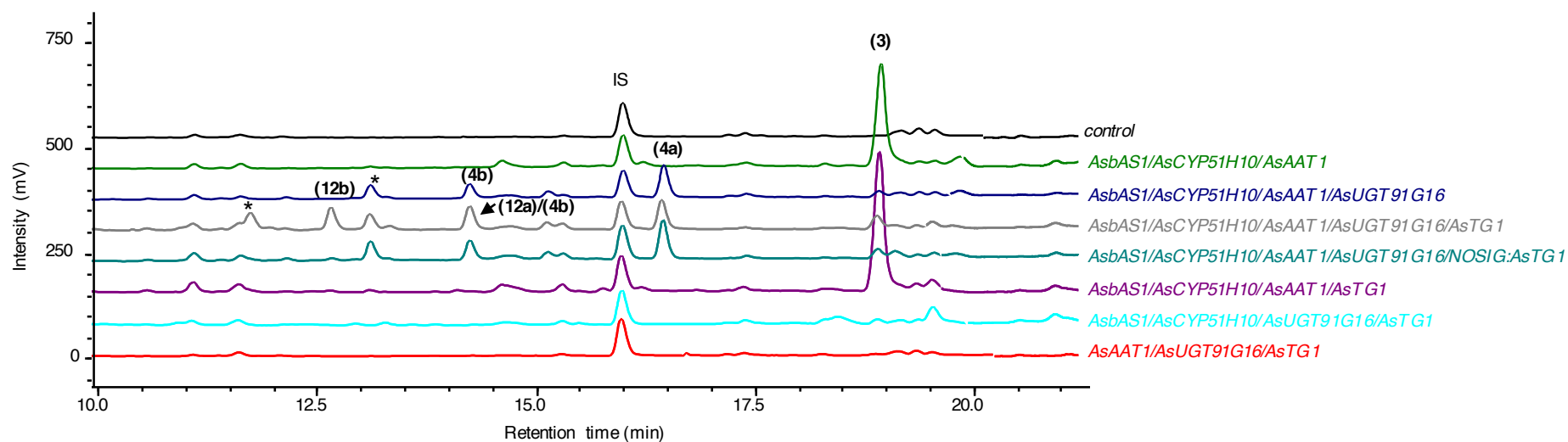


Fig. S11. Transient expression of *AsUGT91G16* and *AsTG1* in *N. benthamiana* leaves. Analysis of *N. benthamiana* leaf extracts from plants co-expressing *tHMGR*, *AsbAS1/Sad1*, *AsCYP51H10/Sad2*, *AsUGT99D1/AsAAT1*, *AsUGT91G16* and *AsTG1* in different combinations. Analysis was carried out by HPLC-MS with charged aerosol detection. *AsbAS1*, *AsCYP51H10* and

AsAAT1 together form the triterpene glycoside 3 β -(α -L-arabinopyranosyloxy)-12,13 β -epoxy,16 β -hydroxy- β -amyirin (**3**) (green trace). Addition of *AsUGT91G16* (blue trace) results in accumulation of new peaks (**4a**) and (**4b**). Co-expression of this combination together with *AsTG1* (grey trace) leads to accumulation of a new peak (**12b**) at 12.2 min and an increase in the (**4b**) peak at 14.4 min which is thought to be due to a new compound (**12a**) co-eluting. These peaks were greatly reduced when the N-terminal signal peptide of *AsTG1* was deleted (NOSIG-*AsTG1*; turquoise trace) suggesting that the signal peptide is critical for *AsTG1* activity *in planta*. *AsTG1* appears to be dependent on the activity of *AsUGT91G16*, as co-expression of *AsbAS1*, *AsCYP51H10*, *AsAAT1* and *AsTG1* without *AsUGT91G16* (purple trace) results in the accumulation of (**3**). The activities of *AsUGT91G16* and *AsTG1* are dependent on the arabinosyltransferase *AsAAT1*, as no new peaks are observed following co-expression of *AsbAS1*, *AsCYP51H10*, *AsUGT91G16* and *AsTG1* without *AsAAT1* (light blue trace). Similarly, no new peaks are observed when the arabinosyltransferase *AsAAT1* is co-expressed with *AsUGT91G16* and *AsTG1* in the absence of *AsbAS1* and *AsCYP51H10*, suggesting that these enzymes do not modify endogenous *N. benthamiana* compounds. Additional peaks (marked by asterisks) were also present that were likely due to endogenous *N. benthamiana* glycosyltransferases. IS = internal standard (digitoxin).

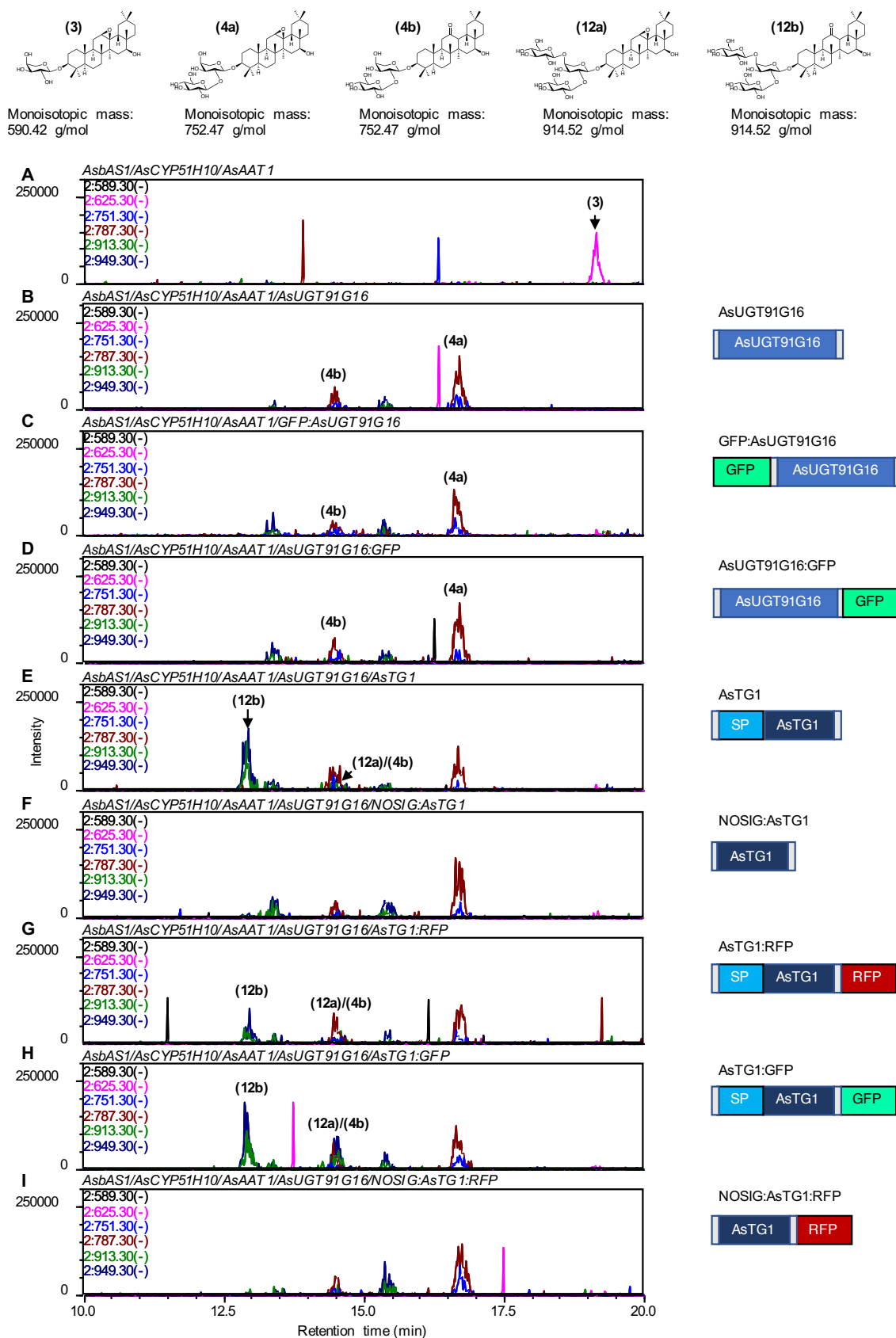


Fig. S12. Fluorescent protein fusions of AsUGT91G16 and AsTG1 are catalytically active in *N. benthamiana*.

HPLC-MS-Charged Aerosol Detection of extracts from infiltrated *N. benthamiana* leaves are shown. (A) Accumulation of the arabinosylated triterpene 3 β -(α -L-arabinopyranosyloxy)-12,13 β -epoxy,16 β -hydroxy- β -amyryn (**3**) was detected in leaves expressing *AsbAS1*, *AsCYP51H10* and *AsAAT1*. Co-expression of *AsbAS1*, *AsCYP51H10* and *AsAAT1* with wild type *AsUGT91G16* (B) or with N- or C-terminal *AsUGT91G16* GFP fusion constructs (C, D) yielded similar accumulation of peaks (**4a**) and (**4b**). Co-expression of *AsbAS1*, *AsCYP51H10*, *AsAAT1* and *AsUGT91G16* together with *AsTG1* resulted in accumulation of (**12a**) and (**12b**) (E). (F) *AsTG1* lacking the N-terminal signal sequence (NOSIG:*AsTG1*) shows reduced activity with little accumulation of (**12a/12b**), while C-terminal fusions of mRFP (*AsTG1*:RFP) (G) or GFP (*AsTG1*:GFP) (H) to *AsTG1* accumulate the same product peaks (**12a/12b**) as *AsTG1* without a fusion tag (E). (I) C-terminal RFP fusion to *AsTG1* lacking the N-terminal signal sequence also showed reduced activity, similar to that seen in (F). Schematic representation of fluorescent protein fusion constructs are shown to the right. Construct diagrams are not drawn to scale. SP, *AsTG* N-terminal signal peptide sequence; GFP, green fluorescent protein; RFP, monomeric red fluorescent protein. Grey boxes represent linkers introduced by Gateway cloning.

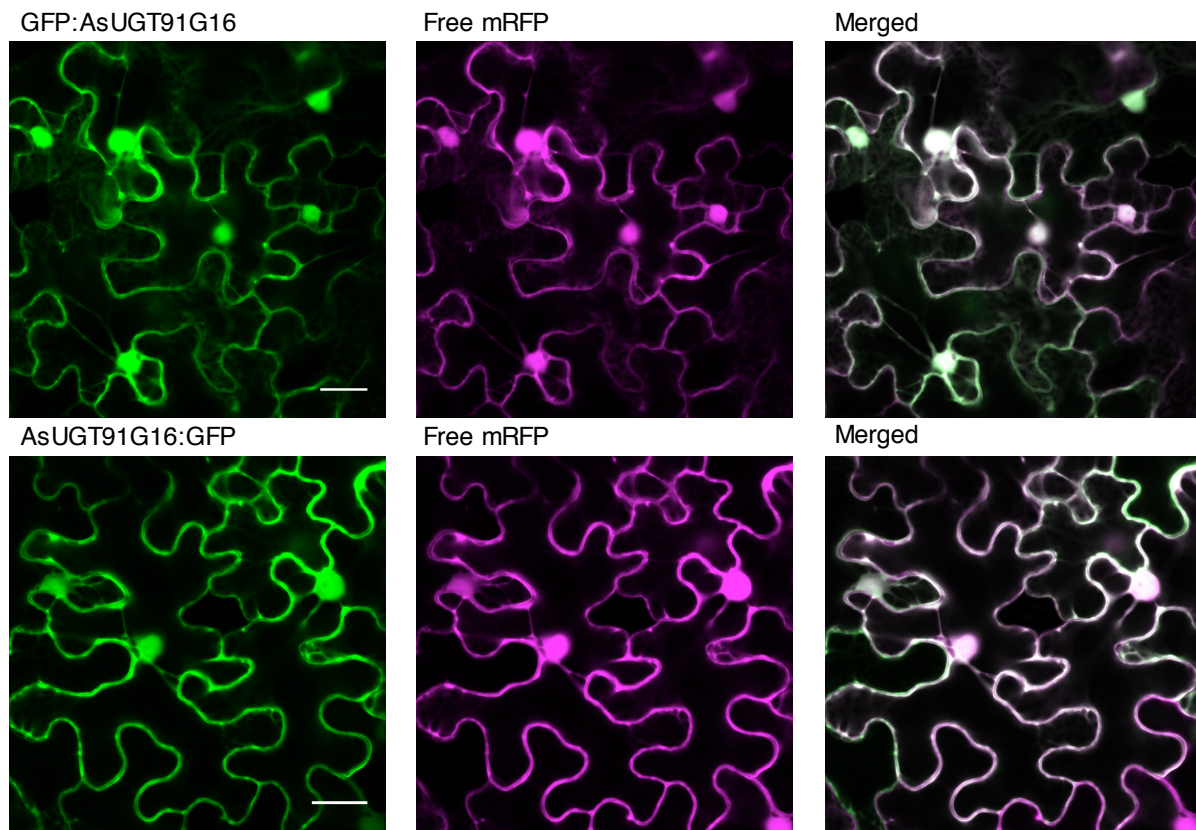


Fig. S13. AsUGT91G16 localises to the cytosol in *N. benthamiana* leaves. Co-localisation of free mRFP1 (26) with: (top panel), an N-terminal GFP fusion to AsUGT91G16 (GFP:AsUGT91G16) and (lower panel), a C-terminal GFP fusion to AsUGT91G16 (AsUGT91G16:GFP). The GFP fusions to AsUGT91G16 co-localise with free mRFP1 in the cytosol and nucleus. GFP fusions are shown in green (left); RFP is shown in magenta (middle) and merged images are shown in white (right). Images were taken two days post-infiltration. Scale bar = 20 μ m.

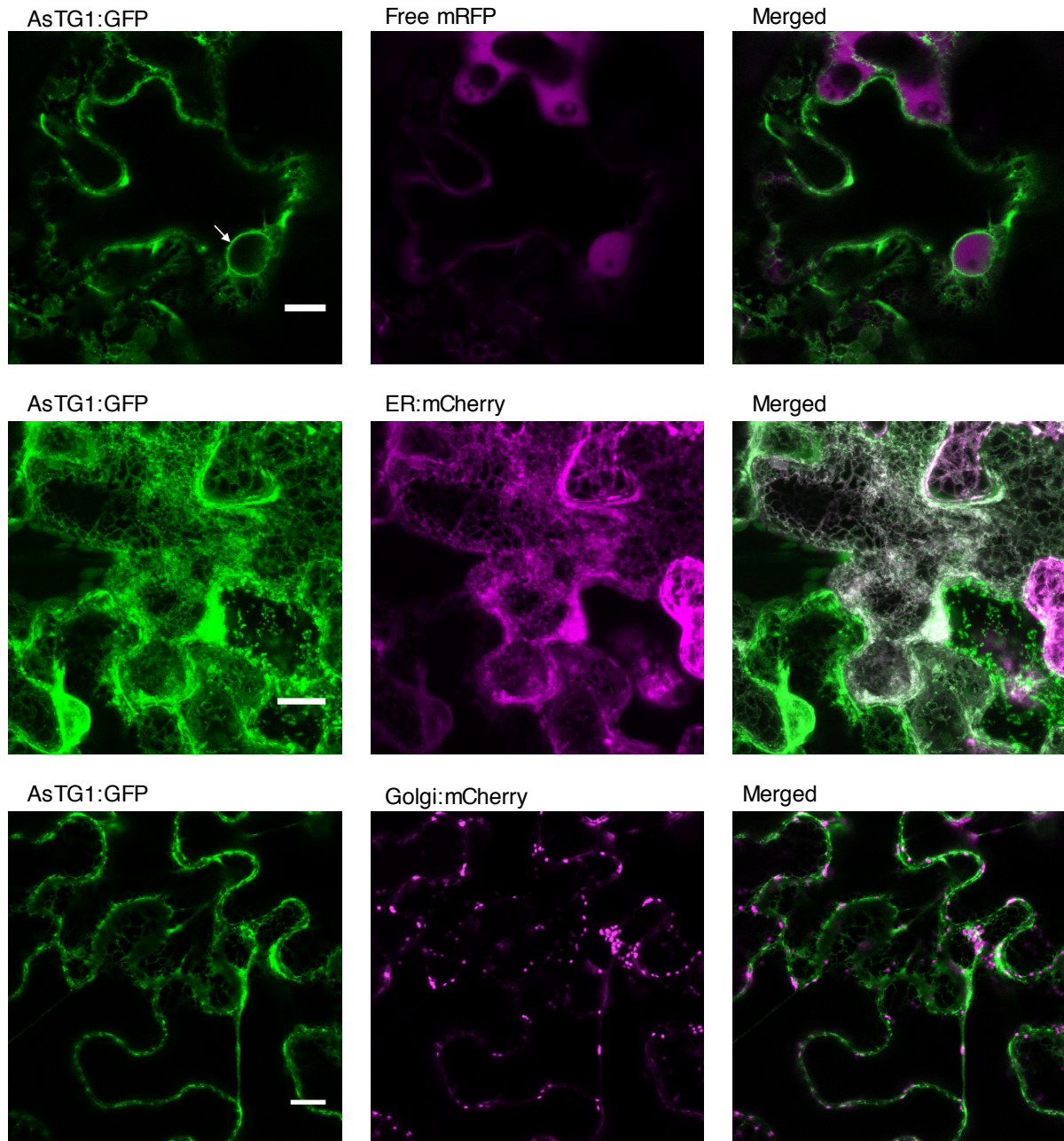
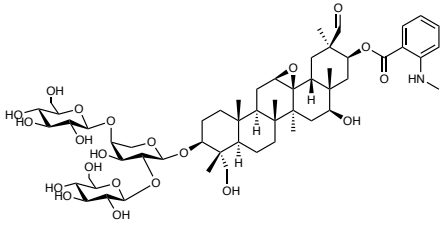
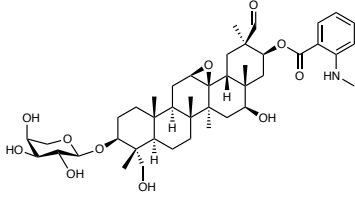
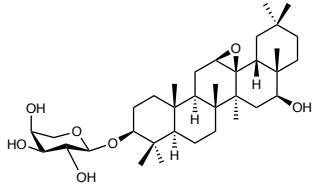
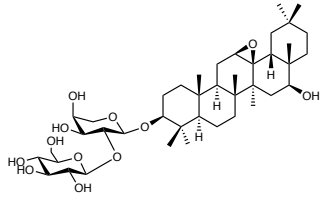
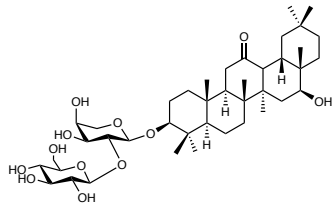
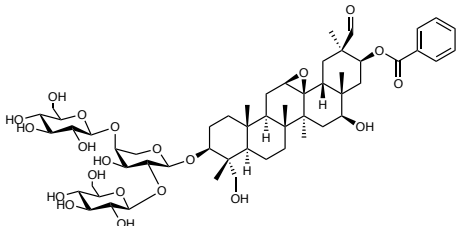
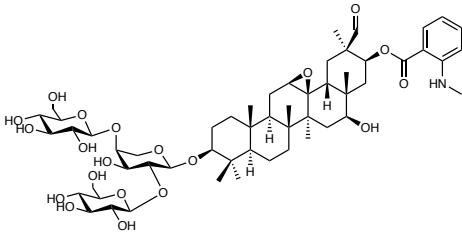
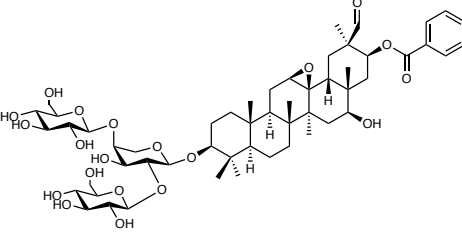
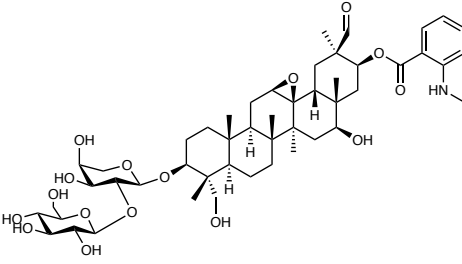
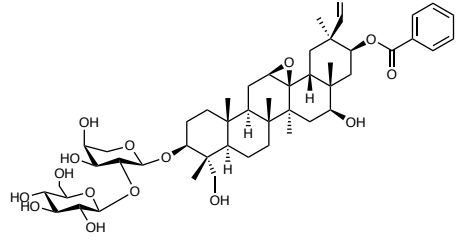
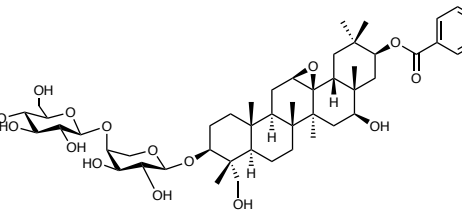
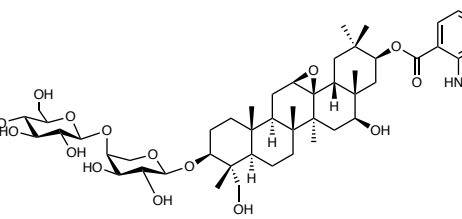


Fig. S14. AsTG1 is targeted to the endomembrane system and traffics through the ER. Co-expression of a C-terminal GFP fusion to AsTG1 (AsTG1:GFP) with: (top panel), free mRFP (35S:mRFP (26)); (middle panel), ER:mCherry (ER-rk CD3-959 (27)), and (lower panel), Golgi:mCherry (G-rk CD3-967 (27)). One day post-infiltration, a GFP fusion to AsTG1 co-localises with a marker for the ER, and does not co-localise with free RFP in the cytoplasm or nucleus. No co-localisation was seen with a marker for the Golgi. GFP fusions are shown in green (left); RFP/mCherry fusions are shown in magenta (middle) and merged images are shown in white (right). Images were taken one day post-infiltration. Scale bar = 10 μ m.

Table S1. Details of triterpene structures referred to in this work

| Compound | Full name, formula, monoisotopic mass | Structure |
|----------|--|--|
| (1) | Avenacin A-1; 3β-[[β-D-glucopyranosyl-(1->2)-[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-16β,23-dihydroxy-30-oxo-12β,13-epoxyoleanan-21β-yl 2-(methylamino)benzoate Formula: C ₅₅ H ₈₃ NO ₂₁ Monoisotopic mass: 1093.55 g/mol |  |
| (2) | Bis-deglucosyl avenacin A-1; 3β-{bisdeglucosyl-(1->2)-,(1->4)-}avenacin A-1; 3β-(α-L-arabinopyranosyloxy)-16β,23-dihydroxy-30-oxo-12β,13-epoxyoleanan-21β-yl 2-(methylamino)benzoate Formula: C ₄₃ H ₆₃ NO ₁₁ Monoisotopic mass: 769.44 g/mol |  |
| (3) | EpHβA-3-O-Ara; 3β-(α-L-arabinopyranosyloxy)-12,13β-epoxy,16β-hydroxy-β-amyryn Formula: C ₃₅ H ₅₈ O ₇ Monoisotopic mass: 590.42 g/mol |  |
| (4a) | Predicted structure: 3β-[[β-D-glucopyranosyl-(1->2)-α-L-arabinopyranosyl]oxy]-12,13β-epoxy,16β-hydroxy-β-amyryn Formula: C ₄₁ H ₆₈ O ₁₂ Monoisotopic mass: 752.47 g/mol |  |
| (4b) | 3β-[[β-D-glucopyranosyl-(1->2)-α-L-arabinopyranosyl]oxy]-12-keto,16β-hydroxy-β-amyryn Formula: C ₄₁ H ₆₈ O ₁₂ Monoisotopic mass: 752.47 g/mol |  |
| (5) | Avenacin A-2; 3β-[[β-D-glucopyranosyl-(1->2)-[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-16β,23-dihydroxy-30-oxo-12β,13-epoxyoleanan-21β-yl benzoate Formula: C ₅₄ H ₈₀ NO ₂₁ Monoisotopic mass: 1064.52 g/mol |  |

| | | |
|------|---|--|
| (6) | <p>Avenacin B-1; 3β-[[β-D-glucopyranosyl-(1->2)-[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-16β-hydroxy-30-oxo-12β, 13-epoxyoleanan-21β-yl 2-(methylamino)benzoate</p> <p>Formula: C₅₅H₈₃NO₂₀ Monoisotopic mass: 1077.55 g/mol</p> |  |
| (7) | <p>Avenacin B-2; 3β-[[β-D-glucopyranosyl-(1->2)-[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-16β-hydroxy-30-oxo-12β, 13-epoxyoleanan-21β-yl benzoate</p> <p>Formula: C₅₄H₈₀NO₂₀ Monoisotopic mass: 1048.52 g/mol</p> |  |
| (8) | <p>Mono-deglucosyl avenacin A-1; 3β-{deglucosyl-(1->4)-} avenacin A-1; 3β-[[β-D-glucopyranosyl-(1->2)-α-L-arabinopyranosyl]oxy]-16β,23-dihydroxy-30-oxo-12β, 13-epoxyoleanan-21β-yl 2-(methylamino)benzoate</p> <p>Formula: C₄₉H₇₃NO₁₆ Monoisotopic mass: 931.49 g/mol</p> |  |
| (9) | <p>Mono-deglucosyl avenacin A-2; 3β-{deglucosyl-(1->4)-} avenacin A-2; 3β-[[β-D-glucopyranosyl-(1->2)-α-L-arabinopyranosyl]oxy]-16β,23-dihydroxy-30-oxo-12β, 13-epoxyoleanan-21β-yl benzoate</p> <p>Formula: C₄₈H₇₀NO₁₆ Monoisotopic mass: 902.47 g/mol</p> |  |
| (10) | <p>Predicted structure: 3β-[[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-16β,23-dihydroxy-12β, 13-epoxyoleanan-21β-yl benzoate</p> <p>Formula: C₄₈H₇₂O₁₅ Monoisotopic mass: 888.49 g/mol</p> |  |
| (11) | <p>Predicted structure: 3β-[[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-16β,23-dihydroxy-12β, 13-epoxyoleanan-21β-yl 2-(methylamino)benzoate</p> <p>Formula: C₄₉H₇₅NO₁₅ Monoisotopic mass: 917.51 g/mol</p> |  |

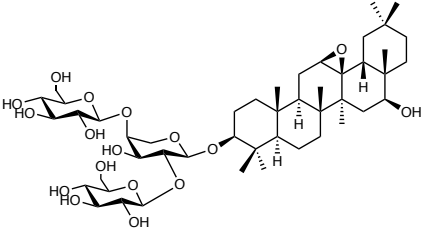
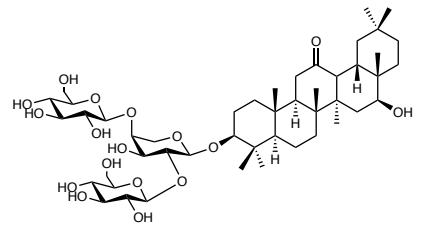
| | | |
|--------------|---|--|
| (12a) | Predicted structure: 3β-[[β-D-glucopyranosyl-(1->2)-[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-12,13β-epoxy,16β-hydroxy-β-amyirin Formula: C ₄₇ H ₇₈ O ₁₇ Monoisotopic mass: 914.52 g/mol |  |
| (12b) | 3β-[[β-D-glucopyranosyl-(1->2)-[β-D-glucopyranosyl-(1->4)]-α-L-arabinopyranosyl]oxy]-12-keto,16β-hydroxy-β-amyirin Formula: C ₄₇ H ₇₈ O ₁₇ Monoisotopic mass: 914.52 g/mol |  |

Table S2. Assembly statistics for *de novo* *A. strigosa* transcriptome (<http://db.ncgr.ac.cn/oat/RNAseq.php>).

| | |
|---------------------------------|-----------|
| Total Trinity transcripts | 196905 |
| GC content | 48.55 % |
| Contig N50 | 1763 |
| Total number of assembled bases | 210050737 |

Table S3. Protein sequences used in tBLASTn analysis of the *de novo* *A. strigosa* transcriptome

| GT1 family | UGT name | Protein acc. no. | Plant species |
|------------|------------|------------------|------------------------------|
| A | PgUGT94Q2 | AGR44632 | <i>Panax ginseng</i> |
| A | GmUGT79B30 | BAR88077 | <i>Glycine max</i> |
| A | GmUGT91H4 | BAI99585 | <i>Glycine max</i> |
| B | AtUGT89C1 | Q9LNE6 | <i>Arabidopsis thaliana</i> |
| C | PoUGT90A7 | ACB56926 | <i>Pilosella officinarum</i> |
| D | GmUGT73F2 | BAM29362 | <i>Glycine max</i> |
| E | PgUGT71A27 | AIZ00429 | <i>Panax ginseng</i> |
| E | ZmUGT708A6 | A0A096SRM5 | <i>Zea mays</i> |
| E | AtUGT72B1 | Q9M156 | <i>Arabidopsis thaliana</i> |
| E | AtUGT88A1 | AEE75831 | <i>Arabidopsis thaliana</i> |
| F | AtUGT78D1 | Q9S9P6 | <i>Arabidopsis thaliana</i> |
| G | AtUGT85A1 | AAF18537 | <i>Arabidopsis thaliana</i> |
| H | AtUGT76B1 | NP_187742 | <i>Arabidopsis thaliana</i> |
| I | AtUGT83A1 | Q9SGA8 | <i>Arabidopsis thaliana</i> |
| J | AtUGT87A1 | O64732 | <i>Arabidopsis thaliana</i> |
| K | AtUGT86A1 | Q9SJL0 | <i>Arabidopsis thaliana</i> |
| L | AtUGT84A1 | Q5XF20 | <i>Arabidopsis thaliana</i> |
| L | SgUGT74AC1 | AEM42999 | <i>Siraitia grosvenorii</i> |
| L | AtUGT75B1 | AEE27854 | <i>Arabidopsis thaliana</i> |
| M | AtUGT92A1 | Q9LXV0 | <i>Arabidopsis thaliana</i> |
| N | AtUGT82A1 | Q9LHJ2 | <i>Arabidopsis thaliana</i> |
| O | SIGAME17 | XP_004243637 | <i>Solanum lycopersicum</i> |

Table S4. *Avena strigosa* UGTs identified in the Self-Organising Map (SOM) analysis.

| UGT name | Transcript I.D. | Transcript length (nt) | Protein acc. no. | Reference |
|-----------------|--------------------------|-------------------------------|-------------------------|------------------|
| AsUGT706C11 | TRINITY_DN36645_c2_g2_i1 | 1905 | MN396759 | This study |
| AsUGT703A5 | TRINITY_DN36598_c0_g6_i3 | 1784 | ACD03255 | (28) |
| AsUGT705A4 | TRINITY_DN36376_c1_g5_i2 | 2283 | ACD03241 | (28) |
| AsUGT706F6 | TRINITY_DN36313_c4_g2_i4 | 1898 | AZQ26918 | (16) |
| AsUGT85B2 | TRINITY_DN35400_c0_g1_i1 | 1605 | ACD03248 | (28) |
| AsUGT84C2 | TRINITY_DN35339_c1_g1_i2 | 1970 | ACD03236 | (28) |
| AsUGT99C8 | TRINITY_DN34036_c1_g5_i5 | 1699 | MN396758 | This study |
| AsUGT99C4 | TRINITY_DN34036_c1_g5_i2 | 1968 | ACD03245 | (28) |
| AsUGT91G16 | TRINITY_DN33854_c2_g2_i2 | 1803 | MN396760 | This study |

Table S5. Characterised UGTs used in phylogenetic analysis

| Enzyme | Accession number | UGT family | UGT Group | Plant species | Reported activity | Reference |
|---------------|------------------|------------|-----------|-----------------------------|--|-----------|
| AtUGT79B1 | Q9LVW3 | UGT79 | A | <i>Arabidopsis thaliana</i> | Anthocyanidin 3-O-glucoside | (29) |
| AtUGT79B6 | Q9FN26 | UGT79 | A | <i>Arabidopsis thaliana</i> | [1,2]-xylosyltransferase | (30) |
| Cs1-6RhaT | ABA18631 | UGT79 | A | <i>Citrus sinensis</i> | Flavonol 3-O-galactoside [1,2]-glucosyltransferase | (31) |
| GmUGT79A6 | BAN91401 | UGT79 | A | <i>Glycine max</i> | Flavonoid 7-O/3-O-glucoside [1,6]-rhamnosyltransferase | (32) |
| LeABRT2 | BAU68118 | UGT79 | A | <i>Lobelia erinus</i> | Flavonol 3-O-glucoside/galactoside [1,6]-rhamnosyltransferase | (33) |
| GmUGT91H4 | BAI99585 | UGT91 | A | <i>Glycine max</i> | Delphinidin 3-O-glucoside [1,6]-rhamnosyltransferase | (34) |
| GmUGT91H9 | NP_001348424 | UGT91 | A | <i>Glycine max</i> | Triterpene 3-O-galactoside [1,2]-rhamnosyltransferase | (35) |
| In3GGT | Q53UH4 | UGT91 | A | <i>Ipomoea nil</i> | Triterpene 3-O-galactoside [1,2]-glucosyltransferase | (36) |
| GjUGT94E5 | F8WKW8 | UGT94 | A | <i>Gardenia jasminoides</i> | Anthocyanidin 3-O-glucoside [1,2]-glucosyltransferase | (37) |
| BpUGT94B1 | Q5NTH0 | UGT94 | A | <i>Bellis perennis</i> | Apocarotenoid glucoside [1,6]-glucosyltransferase | (38) |
| Cm1-2RhaT1 | AAL06646 | UGT94 | A | <i>Citrus maxima</i> | Anthocyanidin 3-O-glucoside [1,2]-glucuronosyltransferase | (31) |
| PgUGT94Q2 | AGR44632 | UGT94 | A | <i>Panax ginseng</i> | Flavonoid 7-O-glucoside [1,2]-rhamnosyltransferase | (39) |
| SgUGT94-289-3 | | UGT94 | A | <i>Siraitia grosvenorii</i> | Triterpene 3-O-glucoside [1,2]-glucosyltransferase | (40) |
| SIGAME18 | XP_004243636 | | A | <i>Solanum lycopersicum</i> | Triterpene 24-O-glucoside/3-O-glucoside [1,2]-/[1-6]-glucosyltransferase | (41) |
| | | | | | Steroidal alkaloid 3-O-glucoside [1,2]-glucosyltransferase | |

| Enzyme | Accession number | UGT family | UGT Group | Plant species | Reported activity | Reference |
|---------------|------------------|------------|-----------|------------------------------|---|-----------|
| VpUGT94F1 | BAI44133 | UGT94 | A | <i>Veronica persica</i> | Flavonoid 3-O-glucoside [1,2]-glucosyltransferase | (42) |
| AtUGT89C1 | AAF80123 | UGT89 | B | <i>Arabidopsis thaliana</i> | Flavonol 7-O-rhamnosyltransferase | (43) |
| UGT89A2-Col-0 | Q9LZD8 | UGT89 | B | <i>Arabidopsis thaliana</i> | Dihydroxybenzoic acid xylosyltransferase | (44) |
| PoUGT90A7 | ACB56926 | UGT90 | C | <i>Pilosella officinarum</i> | Flavonol glucosyltransferase | (45) |
| AlcUGT73G1 | AAP88406 | UGT73 | D | <i>Allium cepa</i> | Flavonoid glucosyltransferase | (46) |
| AtUGT73B3 | AAM47999 | UGT73 | D | <i>Arabidopsis thaliana</i> | Flavonoid-7-O-glucosyltransferase | (47) |
| AtUGT73C1 | AEC09294 | UGT73 | D | <i>Arabidopsis thaliana</i> | Cytokinin glucosyltransferase 1 | (48) |
| AsUGT99D1 | AZQ26921 | UGT99 | D | <i>Avena strigosa</i> | Triterpene-3-O-arabinosyltransferase | (16) |
| BavUGT73C10 | AFN26666 | UGT73 | D | <i>Barbarea vulgaris</i> | Triterpene-3-O-glucosyltransferase | (49) |
| CbBet5OGT | CAB56231 | UGT73 | D | <i>Cleretum bellidiforme</i> | Betanidin-5-O-glucosyltransferase | (50) |
| CsUGT73A20 | ALO19886 | UGT73 | D | <i>Camellia sinensis</i> | Flavonoid 7-O/3-O-glucosyltransferase | (51) |
| CsUGT73AM3 | KGN59015 | UGT73 | D | <i>Cucumis sativus</i> | Triterpene-3-O-glucosyltransferase | (52) |
| GmUGT73F2 | BAM29362 | UGT73 | D | <i>Glycine max</i> | Triterpene 22-O-arabinoside [1,3]-glucosyltransferase | (53) |
| GmUGT73F4 | BAM29363 | UGT73 | D | <i>Glycine max</i> | Triterpene 22-O-arabinoside [1,3]-xylosyltransferase | (53) |
| GmUGT73P2 | BAI99584 | UGT73 | D | <i>Glycine max</i> | Triterpene 3-O-glucuronide [1,2]-galactosyltransferase | (34) |
| GuUGAT | ANJ03631 | UGT73 | D | <i>Glycyrrhiza uralensis</i> | Triterpene 3-O-glucuronosyltransferase/ Triterpene 3-O-glucuronide [1,2]-glucuronosyltransferase | (54) |

| Enzyme | Accession number | UGT family | UGT Group | Plant species | Reported activity | Reference |
|-------------|------------------|------------|-----------|---|---|-----------|
| MtUGT73F3 | ACT34898 | UGT73 | D | <i>Medicago truncatula</i> | Triterpene 28-O-glucosyltransferase | (55) |
| SIUGT73L4 | ADQ37966 | UGT73 | D | <i>Solanum lycopersicum</i> | Steroidal alkaloid 3-O-glucoside [1,3]-xylosyltransferase | (41) |
| StSGT3 | ABB84472 | UGT73 | D | <i>Solanum tuberosum</i> | Steroidal alkaloid 3-O-glucoside/galactoside [1,2]-rhamnosyltransferase | (56) |
| CrsUGT707B1 | CCG85331 | UGT707 | E | <i>Crocus sativus</i> | Flavonol 3-O-glucoside [1,2]-glucosyltransferase | (57) |
| FcCGT | BBA18062 | UGT708 | E | <i>Citrus paradisi/Fortunella crassifolia</i> | Flavonoid 3'-C/5'-C-glucosyltransferase | (58) |
| AtUGT71B6 | NP_188815 | UGT71 | E | <i>Arabidopsis thaliana</i> | Abscisate β -glucosyltransferase | (59) |
| AtUGT71C1 | NP_180536 | UGT71 | E | <i>Arabidopsis thaliana</i> | UDP-glucosyl transferase 71C1 | (60) |
| OsUGT707A3 | BAC83989 | UGT71 | E | <i>Oryza sativa</i> | Flavonoid 3-O-glycosyltransferase | (61) |
| AtUGT72B1 | Q9M156 | UGT72 | E | <i>Arabidopsis thaliana</i> | UDP-glycosyltransferase 72B1 | (62) |
| AtUGT72E2 | AED98252 | UGT72 | E | <i>Arabidopsis thaliana</i> | Hydroxycinnamate 4- β -glucosyltransferase | (63) |
| MtUGT72L1 | ACC38470 | UGT72 | E | <i>Medicago truncatula</i> | Proanthocyanidin precursor-specific UDP-glycosyltransferase | (64) |
| AmUGT88D3 | ABR57234 | UGT88 | E | <i>Antirrhinum majus</i> | Chalcone 4'-O-glucosyltransferase | (65) |
| MdPGT1 | B3TKC8 | UGT88 | E | <i>Malus domestica</i> | Phloretin 2'-O-glucosyltransferase | (66) |
| RhGT1 | BAD99560 | UGT88 | E | <i>Rosa hybrida</i> | Anthocyanidin 5/3-O-glucosyltransferase | (67) |
| ScUGT5 | BAJ11653 | UGT88 | E | <i>Sinningia cardinalis</i> | UDP-glucose:3-deoxyanthocyanidin 5-O-glucosyltransferase | (68) |
| GmUGT708D1 | I1L3T1 | | E | <i>Glycine max</i> | Hydroxyflavanone-2-C-glucosyltransferase | (69) |
| OsCGT | CAQ77160 | | E | <i>Oryza sativa</i> | C-glucosyltransferase | (70) |

| Enzyme | Accession number | UGT family | UGT Group | Plant species | Reported activity | Reference |
|------------|------------------|------------|-----------|--------------------------------|--|-----------|
| AtUGT78D1 | Q9S9P6 | UGT78 | F | <i>Arabidopsis thaliana</i> | Flavonol 3-O-glucosyltransferase | (71) |
| Fh3GT1 | ADK75021 | UGT78 | F | <i>Freesia hybrid cultivar</i> | Anthocyanidin 3-O-glucosyltransferase | (72) |
| VmUF3GaT | BAA36972 | UGT78 | F | <i>Vigna mungo</i> | Flavonoid 3-O-galactosyltransferase | (73) |
| VvGT1 | AAB81683 | UGT78 | F | <i>Vitis vinifera</i> | Anthocyanidin 3-O-glucosyltransferase | (74) |
| AtUGT85A1 | AAF18537 | UGT85 | G | <i>Arabidopsis thaliana</i> | Cytokinin-O-glucosyltransferase 2 | (75) |
| PdUGT85A19 | ABV68925 | UGT85 | G | <i>Prunus dulcis</i> | Cyanohydrin glucoside [1,6]-glucosyltransferase | (76) |
| SbUGT85B1 | AAF17077 | UGT85 | G | <i>Sorghum bicolor</i> | Cyanohydrin glycosyltransferase UGT85B1 | (77) |
| AtUGT76D1 | AEC07843 | UGT76 | H | <i>Arabidopsis thaliana</i> | Flavonoid-7-O-glucosyltransferase | (78) |
| SrUGT76G1 | AAR06912 | UGT76 | H | <i>Stevia rebaudiana</i> | Diterpenoid 13-O-glucoside [1,3]-glucosyltransferase | (79) |
| ZmBx8 | AAL57037 | UGT76 | H | <i>Zea mays</i> | UDP-glucosyltransferase BX8 | (80) |
| AtUGT83A1 | Q9SGA8 | UGT83 | I | <i>Arabidopsis thaliana</i> | Unknown | (81) |
| AtUGT87A1 | O64732 | UGT87 | J | <i>Arabidopsis thaliana</i> | Unknown | (81) |
| AtUGT87A2 | NP_001077979 | UGT87 | J | <i>Arabidopsis thaliana</i> | Unknown | (82) |
| AtUGT86A1 | Q9SJL0 | UGT86 | K | <i>Arabidopsis thaliana</i> | Unknown | (81) |
| AtUGT74E2 | NP_172059 | UGT74 | L | <i>Arabidopsis thaliana</i> | Auxin (IBA) glycosyltransferase | (83) |
| AsUGT74H5 | ACD03250 | UGT74 | L | <i>Avena strigosa</i> | N-Methylantranilate O-glucosyltransferase | (84) |
| HvUGT13248 | ADC92550 | UGT74 | L | <i>Hordeum vulgare</i> | Epoxy-sesquiterpenoid-3-O-glucosyltransferase | (85) |
| OsUGT79 | XP_015635481 | UGT74 | L | <i>Oryza sativa Japonica</i> | Epoxy-sesquiterpene 3-O-glucosyltransferase | (86) |
| PgUGT74A1 | AGR44631 | UGT74 | L | <i>Panax ginseng</i> | Triterpene-3-O-glucosyltransferase | (39) |

| Enzyme | Accession number | UGT family | UGT Group | Plant species | Reported activity | Reference |
|------------|------------------|------------|-----------|---|---|-----------|
| VhUGT74M1 | ABK76266 | UGT74 | L | <i>Vaccaria hispanica</i> | Triterpene carboxylic acid 28-O-glucosyltransferase | (87) |
| ZmIAGT | AAA59054 | UGT74 | L | <i>Zea mays</i> | Auxin glucosyltransferase | (88) |
| AtUGT75B1 | AEE27854 | UGT75 | L | <i>Arabidopsis thaliana</i> | Indole-3-acetate b β -glucosyltransferase 1 | (89) |
| CrsGT45 | ACM66950 | UGT75 | L | <i>Crocus sativus</i> | Flavonoid 7-O-glucosyltransferase | (90) |
| GjUGT75L6 | F8WKW0 | UGT75 | L | <i>Gardenia jasminoides</i> | Apocarotenoid glucosyltransferase | (37) |
| MdUGT75L17 | AAX16493 | UGT75 | L | <i>Malus x domestica Brokh.</i> | UDP-glucose:phloretin 4'-O-glycosyltransferase | (91) |
| MpUGT75L4 | ABL85474 | UGT75 | L | <i>Maclura pomifera</i> | Flavonoid glycosyltransferase | (92) |
| Pf3R4 | BAA36421 | UGT75 | L | <i>Perilla frutescens</i> | Anthocyanin 5-O-glucosyltransferase | (93) |
| Via5GT | AHL68667 | UGT75 | L | <i>Vitis amurensis Rupr. cv. 'Zuoshanyi</i> | Anthocyanin 5-O-glucosyltransferase | (94) |
| AtUGT84A1 | Q5XF20 | UGT84 | L | <i>Arabidopsis thaliana</i> | Hydroxycinnamate glucosyltransferase 2 | (95) |
| BnUGT84A9a | CAS03354 | UGT84 | L | <i>Brassica napus</i> | Sinapate/hydroxycinnamate glucosyltransferase | (96) |
| FaGT2 | Q66PF4 | UGT84 | L | <i>Fragaria ananassa</i> | Cinnamate glucosyltransferase | (97) |
| GtUF6CGT1 | BAQ19550 | UGT84 | L | <i>Gentiana triflora</i> | Flavonoid 6-C-glucosyltransferase | (98) |
| CuLGT | BAA93039 | UGT84 | L | <i>Citrus unshui</i> | Triterpene (limonoid)-17-O-glucosyltransferase | (99) |
| AtUGT92A1 | Q9LXV0 | UGT92 | M | <i>Arabidopsis thaliana</i> | Unknown | (81) |
| CcDOPA5GT | BAD91804 | UGT92 | M | <i>Celosia cristata</i> | Cyclo-DOPA 5-O-glucosyltransferase | (100) |
| MjcDOPA5GT | BAD91803 | UGT92 | M | <i>Mirabilis jalapa</i> | Cyclo-DOPA 5-O-glucosyltransferase | (100) |
| AtUGT82A1 | Q9LHJ2 | UGT82 | N | <i>Arabidopsis thaliana</i> | Unknown | (81) |

| Enzyme | Accession number | UGT family | UGT Group | Plant species | Reported activity | Reference |
|----------------|------------------|------------|-----------|-----------------------------|--|-----------|
| SIGAME17 | XP_004243637 | UGT93 | O | <i>Solanum lycopersicum</i> | Steroidal alkaloid 3-O-galactoside [1,4]-glucosyltransferase | (41) |
| ZmcisZog1 | AAK53551 | UGT93 | O | <i>Zea mays</i> | cis-zeatin O-glucosyltransferase | (101) |
| OsUGT709A4 | BAC80066 | UGT709A4 | P | <i>Oryza sativa</i> | Isoflavonoid-7-O-glucosyltransferase | (61) |
| SgUGT720-269-1 | | UGT720 | P | <i>Siraitia grosvenorii</i> | Triterpene 24-O/3-O-glucosyltransferase | (40) |

Table S6 Characterised plant UGTs with glycoside-specific glycosyltransferase activity

| Enzyme | UGT family | Group | Activity | Plant species | Reference |
|------------|------------|-------|--|-----------------------------|-----------|
| AtUGT79B6 | UGT79 | A | Flavonol 3- <i>O</i> -galactoside [1,2]-glucosyltransferase | <i>Arabidopsis thaliana</i> | (30) |
| BpUGT94B1 | UGT94 | A | Anthocyanidin 3- <i>O</i> -glucoside [1,2]-glucuronosyltransferase | <i>Bellis perennis</i> | (38) |
| CaUGT3 | UGT94 | A | Flavonol 3- <i>O</i> -glucoside [1,6]-glucosyltransferase (processive) | <i>Catharanthus roseus</i> | (102) |
| Cm1-2RhaT1 | UGT94 | A | Flavonol 7- <i>O</i> -glucoside [1,2]-rhamnosyltransferase | <i>Citrus maxima</i> | (31) |
| Cs1-6RhaT | UGT91 | A | Flavonol 7- <i>O</i> /3- <i>O</i> glucoside [1,6]-rhamnosyltransferase | <i>Citrus sinensis</i> | (31) |
| GjUGT94E5 | UGT94 | A | Apocarotenoid glucoside [1,6]-glucosyltransferase | <i>Gardenia jasminoides</i> | (37) |
| GmUGT79A6 | UGT79 | A | Flavonol 3- <i>O</i> -glucoside/galactoside [1,6]-rhamnosyltransferase | <i>Glycine max</i> | (32) |
| GmUGT79A7 | UGT79 | A | Flavonol 3- <i>O</i> -glucoside/galactoside [1,6]-glucosyltransferase | <i>Glycine max</i> | (103) |
| GmUGT79B30 | UGT79 | A | Flavonol 3- <i>O</i> -glucoside/galactoside [1,2]-glucosyltransferase | <i>Glycine max</i> | (104) |
| GmUGT91H4 | UGT91 | A | Triterpene 3- <i>O</i> -galactoside [1,2]-rhamnosyltransferase | <i>Glycine max</i> | (34) |
| GmUGT91H9 | UGT91 | A | Triterpene 3- <i>O</i> -galactoside [1,2]-glucosyltransferase | <i>Glycine max</i> | (35) |
| LeABRT2 | UGT79 | A | Flavonol 3- <i>O</i> -glucoside [1,6]-rhamnosyltransferase | <i>Lobelia erinus</i> | (33) |
| LeABRT4 | UGT79 | A | Flavonol 3- <i>O</i> -glucoside [1,6]-rhamnosyltransferase | <i>Lobelia erinus</i> | (33) |
| PgUGT94Q2 | UGT94 | A | Triterpene 3- <i>O</i> -glucoside [1,2]-glucosyltransferase | <i>Panax ginseng</i> | (39) |

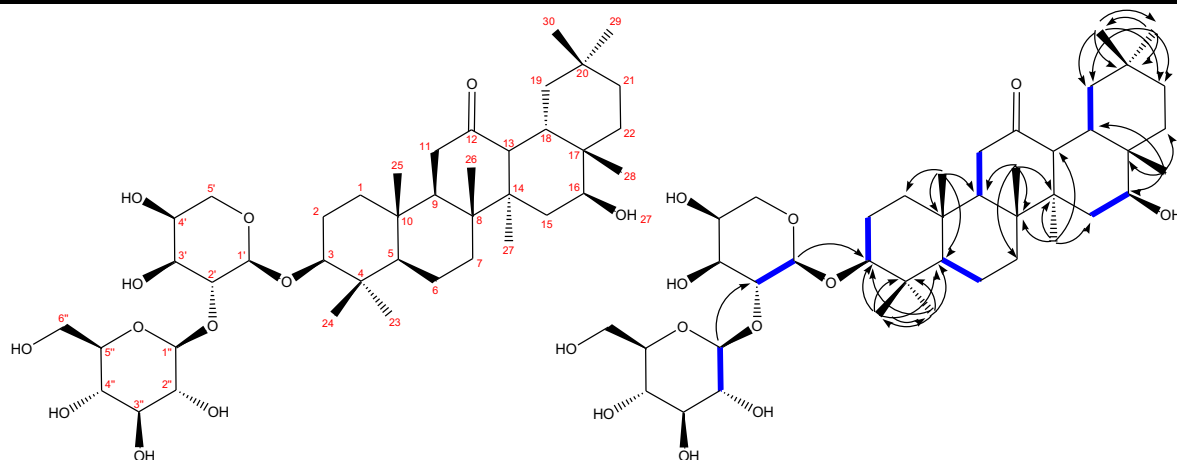
| Enzyme | UGT family | Group | Activity | Plant species | Reference |
|---------------|------------|-------|---|------------------------------|-----------|
| PhA3G1-6RhaT | UGT79 | A | Anthocyanidin 3- <i>O</i> -glucoside [1,6]-rhamnosyltransferase | <i>Petunia hybrida</i> | (105) |
| In3GGT | UGT91 | A | Anthocyanidin 3- <i>O</i> -glucoside [1,2]-glucosyltransferase | <i>Ipomoea nil</i> | (36) |
| SiUGT94D1 | UGT94 | A | Lignan 2'- <i>O</i> -glucoside [1,6]-glucosyltransferase | <i>Sesamum indicum</i> | (106) |
| SgUGT94-289-3 | UGT94 | A | Triterpene 24- <i>O</i> -glucoside/3- <i>O</i> -glucoside [1,2]-/[1,6]-glucosyltransferase | <i>Siraitia grosvenorii</i> | (40) |
| SgUGT94-289-2 | UGT94 | A | Triterpene 24- <i>O</i> -glucoside [1,2]-/[1,6]-glucosyltransferase | <i>Siraitia grosvenorii</i> | (40) |
| SgUGT94-289-1 | UGT94 | A | Triterpene 24- <i>O</i> -glucoside/3- <i>O</i> -glucoside [1,6]-glucosyltransferase | <i>Siraitia grosvenorii</i> | (40) |
| SIGAME18 | UGT94 | A | Steroidal alkaloid 3- <i>O</i> -glucoside [1,2]-glucosyltransferase | <i>Solanum lycopersicum</i> | (41) |
| VpUGT94F1 | UGT94 | A | Flavonol 3- <i>O</i> -glucoside [1,2]-glucosyltransferase | <i>Veronica persica</i> | (42) |
| GmUGT73F2 | UGT73 | D | Triterpene 22- <i>O</i> -arabinoside [1,3]-glucosyltransferase | <i>Glycine max</i> | (53) |
| GmUGT73F4 | UGT73 | D | Triterpene 22- <i>O</i> -arabinoside [1,3]-xylosyltransferase | <i>Glycine max</i> | (53) |
| GmUGT73P2 | UGT73 | D | Triterpene 3- <i>O</i> -glucuronide [1,2]-galactosyltransferase | <i>Glycine max</i> | (34) |
| GuUGAT | UGT73 | D | Triterpene 3- <i>O</i> -glucuronosyltransferase/ Triterpene 3- <i>O</i> -glucuronide [1,2]-glucuronosyltransferase | <i>Glycyrrhiza uralensis</i> | (54) |
| SIUGT73L4 | UGT73 | D | Steroidal alkaloid 3- <i>O</i> -glucoside [1,3]-xylosyltransferase | <i>Solanum lycopersicum</i> | (41) |
| StSGT3 | UGT73 | D | Steroidal alkaloid 3- <i>O</i> -glucoside/galactoside [1,2]-rhamnosyltransferase | <i>Solanum tuberosum</i> | (56) |

| Enzyme | UGT family | Group | Activity | Plant species | Reference |
|------------|------------|-------|---|-----------------------------|-----------|
| CsUGT707B1 | UGT707 | E | Flavonol 3- <i>O</i> -glucoside [1,2]-glucosyltransferase | <i>Crocus sativus</i> | (57) |
| PdUGT85A19 | UGT85 | G | Cyanohydrin glucoside [1,6]-glucosyltransferase | <i>Prunus dulcis</i> | (76) |
| SrUGT76G1 | UGT76 | H | Diterpene 13- <i>O</i> -glucoside [1,3]-glucosyltransferase | <i>Stevia rebaudiana</i> | (79) |
| SIGAME17 | UGT93 | O | Steroidal alkaloid 3- <i>O</i> -galactoside [1,4]-glucosyltransferase | <i>Solanum lycopersicum</i> | (41) |

Triterpene and steroidal glycoside UGTs are indicated in blue.

Table S7. ^{13}C & ^1H δ assignments for the product 3β -[[β -D-glucopyranosyl-(1->2)- α -L-arabinopyranosyl]oxy]-12-keto,16 β -hydroxy- β -amyrin (**4b**) generated by co-expression of AsbAS1/SAD1, AsCYP51H10/SAD2, AsAAT1 and AsUGT91G16 in *N. benthamiana*.

Carbon numbering scheme and selected COSY and HMBC correlations



| Carbon Number | ^{13}C δ | ^1H δ | Carbon Number | ^{13}C δ | ^1H δ |
|---------------|--------------------------|------------------------------|---------------|--------------------------|-------------------------------|
| 12 | 217.46 | / | 1 | 39.98 | 1.61 (1H, m) 1.13 (1H, m) |
| 1' | 105.47 | 4.51 (1H, d, $J=6.1$) | 11 | 39.39 | 2.24 (2H, m) |
| 1'' | 104.82 | 4.59 (1H, d, $J=7.7$) | 10 | 38.23 | / |
| 3 | 90.89 | 3.21 (1H, m) | 17 | 37.77 | / |
| 2' | 79.14 | 3.89 (1H, m) | 15 | 36.91 | 1.86 (1H, m) 1.22 (1H, m) |
| 5'' | 78.25 | 3.26 (1H, m) | 18 | 36.02 | 2.64 (1H, dd, $J=13.6, 3.9$) |
| 3'' | 78.00 | 3.35 (1H, m) | 7 | 35.74 | 1.84 (1H, m) 1.42 (1H, m) |
| 2'' | 76.02 | 3.21 (1H, m) | 21 | 35.25 | 1.34 (1H, m) 1.15 (1H, m) |
| 3' | 73.78 | 3.77 (1H, m) | 29 | 33.19 | 0.91 (3H, s) |
| 4'' | 71.79 | 3.25 (1H, m) | 22 | 32.21 | 1.86 (1H, m) 1.32 (1H, m) |
| 6'' | 71.57 | 3.64 (2H, m) | 20 | 32.02 | / |
| 4' | 69.02 | 3.84 (1H, m) | 27 | 29.58 | 1.40 (3H, s) |
| 16 | 66.76 | 3.92 (1H, m) | 23 | 28.72 | 1.09 (3H, s) |
| 5' | 62.96 | 3.82 (1H, m) 3.64 (1H, m) | 2 | 27.00 | 1.87 (1H, m) 1.75 (1H, m) |
| 13 | 60.02 | 1.92 (1H, m) | 30 | 25.16 | 0.91 (3H, s) |
| 5 | 56.85 | 0.88 (1H, m) | 28 | 22.19 | 0.79 (3H, s) |
| 9 | 48.52 | 1.94 (1H, m) | 26 | 20.22 | 0.94 (3H, s) |
| 14 | 47.32 | / | 6 | 19.20 | 1.59 (1H, m) 1.42 (1H, m) |
| 19 | 46.73 | 1.67 (1H, m) 1.00 (1H, m) | 24 | 17.21 | 0.88 (3H, s) |
| 8 | 43.28 | / | 25 | 16.51 | 1.02 (3H, s) |
| 4 | 40.38 | / | | | |

MeOH-*d*4 [referenced to TMS]. Coupling constants are reported as observed and not corrected for second order effects. Assignments were made via a combination of

^1H , ^{13}C , COSY, DEPT-edited HSQC, and HMBC experiments. Where signals overlap ^1H δ is reported as the centre of the respective HSQC crosspeak. A HMBC correlation was clearly observed between the C1'-H of the arabinose moiety and the C3 of the triterpene scaffold confirming the initial site of glycosylation. This is consistent with our previously reported functional characterisation of AsAAT1 (16). The C2' could be confidently assigned from a COSY correlation between the C1'-H and C2'-H. Assignment of C2' allowed confirmation of a 1,2 linkage of the second saccharide moiety, through observation of a clear HMBC correlation between the C1''-H and C2'. This clearly supports the proposed function of AsUGT91G1 as being a triterpene 3-O-arabinose- β -[1,2]-glucosyltransferase. The analysis also revealed that the compound possessed the triterpene scaffold of 12-keto,16 β -hydroxy- β -amyirin, and not the expected product of AsbAS1 and CYP51H10, 12,13 β -epoxy,16 β -hydroxy- β -amyirin. This degradation of the avenacin epoxide has been previously observed (19, 107-109) and mild acid conditions that leave the sugar chain intact are sufficient to cause the opening of the epoxide ring and the formation of the ketone (110).

Table S8. Single nucleotide mutations in the *AsUGT91G16* gene of avenacin-deficient *A. strigosa* mutants.

| Mutant | Mutation event | Predicted outcome |
|---------------|-----------------------|--------------------------|
| #85 | G963A | W321-STOP |
| #543 | G375A | W125-STOP |
| #1073 | G776A | G259E |
| #1473 | G775A | G259R |

Table S9. F₂ progeny phenotype ratios of ugt91g16 mutant lines crossed to the *A. strigosa* wild type parent (S75).

| ugt91g16 mutant | Wild type root phenotype | Mutant short root phenotype | X² Wild type:Mutant = 3:1 |
|------------------------|---------------------------------|------------------------------------|---|
| #543 | 134 | 58 | 2.778 (P> 0.05) |
| #1473 | 144 | 48 | 0 (P> 0.05) |

Table S10. Predicted full length AsGH1 transcript identified by BLASTn searches with the *Avena atlantica* AsGH1 gene region in a previously generated *A. strigosa* accession S75 root tip transcriptome database*.

| Name of transcript | Transcript length | Predicted protein sequence | Sequence identity | Transcriptome database |
|--------------------|-------------------|----------------------------|-------------------|------------------------|
| contig17777 | 1901 | Full length | 99.8% | (111) |

*Only partial sequences were found in the *de novo* transcriptome generated in this study.

Table S11. Glycosyl hydrolase family 1 (GH1) sequences used in phylogenetic analysis.

| Enzyme name | Accession | GH1 group | Plant species | Activity | Publication |
|-------------------------|-----------|----------------------|-----------------------------|--|-------------|
| AaAA7GT | BAM29304 | At/Os 6 | <i>Agapanthus africanus</i> | Acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase | (25) |
| AtBGLU35 | Q3ECS3 | At I | <i>Arabidopsis thaliana</i> | Myrosinase | (112) |
| AtBGLU38 | P37702 | At I | <i>Arabidopsis thaliana</i> | Myrosinase | (113) |
| AtBGLU33 | O48779 | At II | <i>Arabidopsis thaliana</i> | Not determined | (114) |
| AtBGLU30 | Q9M1C9 | At II | <i>Arabidopsis thaliana</i> | Not determined | (114) |
| AtBGLU31 | Q9FLU9 | At II | <i>Arabidopsis thaliana</i> | Not determined | (114) |
| AtBGLU26 | O64883 | At II | <i>Arabidopsis thaliana</i> | Myrosinase | (115) |
| AtBGLU24 | Q9LKR7 | At II | <i>Arabidopsis thaliana</i> | Not determined | (114) |
| AtBGLU40 | Q9FZE0 | At/Os 1 | <i>Arabidopsis thaliana</i> | Not determined | (114) |
| AtBGLU41 | Q9FIU7 | At/Os 2 | <i>Arabidopsis thaliana</i> | Not determined | (114) |
| AtBGLU42 | Q9FIW4 | At/Os 3 | <i>Arabidopsis thaliana</i> | β -Glucosidase | (116) |
| AtBGLU44 | Q9LV33 | At/Os 4 | <i>Arabidopsis thaliana</i> | β -Mannosidase | (114) |
| AtBGLU45 | O80689 | At/Os 5 | <i>Arabidopsis thaliana</i> | β -Glucosidase | (117) |
| AtBGLU6 | AJW76497 | At/Os 6 | <i>Arabidopsis thaliana</i> | Predicted flavonol 3-O-glucoside 1,6-glucosyltransferase | (118) |
| AtBGLU10 | Q93ZI4 | At/Os 6 | <i>Arabidopsis thaliana</i> | Predicted anthocyanin A9 glucosyltransferase | (119) |
| AtBGLU12 | Q9FH03 | At/Os 7 | <i>Arabidopsis thaliana</i> | Not determined | (114) |
| AtSFR2 | Q93Y07 | At/Os 8 | <i>Arabidopsis thaliana</i> | SENSITIVE TO FREEZING 2 | (120) |
| As-Glu1 | Q38786 | plastid β -Glu | <i>Avena sativa</i> | Avenacosidase I | (121) |
| AsTG1 | | At/Os 6 | <i>Avena strigosa</i> | Triterpene 3-O-arabinoside 1,4-glucosyltransferase | This study |
| Bn_zeatin-O-glucosidase | CAA57913 | At II | <i>Brassica napus</i> | Zeatin β -glucosidase | (122) |

| Enzyme name | Accession | GH1 group | Plant species | Activity | Publication |
|----------------------|-----------|-----------|--------------------------------|--|-------------|
| Cs_β-primeverosidase | BAC78656 | At/Os 7 | <i>Camellia sinensis</i> | β-Primeverosidase | (123) |
| CmAA7GT | BAO96250 | At/Os 6 | <i>Campanula medium</i> | Acyl-glucose dependent anthocyanin 7-O-glucosyltransferase | (24) |
| DgAA7GT | E3W9M3 | At/Os 6 | <i>Delphinium grandiflorum</i> | Acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase | (23) |
| DgAA7BG-GT1 | BAO04178 | At/Os 7 | <i>Delphinium grandiflorum</i> | Acyl-glucose-dependent anthocyanin glucosyltransferase | (21) |
| DgAA7BG-GT2 | BAO04181 | At/Os 7 | <i>Delphinium grandiflorum</i> | Acyl-glucose-dependent anthocyanin glucosyltransferase | (21) |
| DcAA5GT | E3W9M2 | At/Os 6 | <i>Dianthus caryophyllus</i> | Acyl-glucose-dependent anthocyanin 5-O-glucosyltransferase | (23) |
| Gm HIUHase | Q8S3J3 | At/Os 6 | <i>Glycine max</i> | Hydroxyisourate hydrolase | (124) |
| Os3bglu6 | Q8L7J2 | At/Os 1 | <i>Oryza sativa japonica</i> | β-Glucosidase | (125) |
| Os6bglu25 | Q0DA21 | At/Os 2 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| Os1bglu4 | Q5N863 | At/Os 3 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| Os12bglu38 | Q2QSR8 | At/Os 4 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| Os3bglu7 | Q75I93 | At/Os 4 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| Os4bglu18 | Q7XSK0 | At/Os 5 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| Os9bglu31 | B7F7K7 | At/Os 6 | <i>Oryza sativa japonica</i> | Acyl-glucose-dependent flavonol/phenylpropanoid/phytohormone glucosyltransferase | (22) |
| Os5bglu21 | Q60DY1 | At/Os 6 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| Os1bglu5 | Q5JK35 | At/Os 6 | <i>Oryza sativa japonica</i> | Not determined | (126) |

| Enzyme name | Accession | GH1 group | Plant species | Activity | Publication |
|-----------------------------------|-----------|----------------------|------------------------------|---------------------------------|-------------|
| Os4bglu10 | Q7F9K4 | At/Os 7 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| Os11bglu35 | Q53NF0 | At/Os 7 | <i>Oryza sativa japonica</i> | Not determined | (126) |
| OsSFR2 | Q8L6H7 | At/Os 8 | <i>Oryza sativa japonica</i> | SENSITIVE TO FREEZING 2 | (126) |
| Pc Coniferin β -glucosidase | Q9ZT64 | At/Os 5 | <i>Pinus contorta</i> | Coniferin β -glucosidase | (127) |
| Sa Myrosinase MB3 | P29092 | At I | <i>Sinapis alba</i> | Myrosinase | (128) |
| Sl_ β -mannosidase | AAL37714 | At/Os 4 | <i>Solanum lycopersicum</i> | β -Mannosidase | (129) |
| Sb_dhurrinase | AAC49177 | plastid β -Glu | <i>Sorghum bicolor</i> | Dhurrinase | (130) |
| Tr_linamarinase | P26205 | At/Os 7 | <i>Trifolium repens</i> | Cyanogenic β -glucosidase | (131) |
| ZmGlu1 | P49235 | plastid β -Glu | <i>Zea mays</i> | β -Glucosidase | (132) |

Table S12. Characterised GH1 plant transglycosidases (TGs)

| Enzyme | Accession | GH1 group | Full name | Subcellular Localisation | Plant species | Reference |
|-------------|-----------|-----------|--|--------------------------|--------------------------------|-----------|
| AtSFR2 | Q93Y07 | At/Os 8 | SENSITIVE TO FREEZING 2 | chloroplast membrane | <i>Arabidopsis thaliana</i> | (120) |
| AaAA7GT | BAM29304 | At/Os 6 | Acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase | not determined | <i>Agapanthus africanus</i> | (25) |
| CmAA7GT | BAO96250 | At/Os 6 | Acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase | not determined | <i>Campanula medium</i> | (24) |
| DgAA7GT | E3W9M3 | At/Os 6 | Acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase | vacuolar | <i>Delphinium grandiflorum</i> | (23) |
| DgAA7BG-GT1 | BAO04178 | At/Os 7 | Acyl-glucose-dependent anthocyanin glucosyltransferase | vacuolar | <i>Delphinium grandiflorum</i> | (21) |
| DgAA7BG-GT2 | BAO04181 | At/Os 7 | Acyl-glucose-dependent anthocyanin glucosyltransferase | vacuolar | <i>Delphinium grandiflorum</i> | (21) |
| DcAA5GT | E3W9M2 | At/Os 6 | Acyl-glucose-dependent anthocyanin 5-O-glucosyltransferase | vacuolar | <i>Dianthus caryophyllus</i> | (23) |
| Os9bglu31 | B7F7K7 | At/Os 6 | Acyl-glucose-dependent flavonol/phytohormone/phenylpropanoid glucosyltransferase | vacuolar | <i>Oryza sativa japonica</i> | (22) |

Table S13: ^{13}C & ^1H δ assignments for the product 3 β -{[β -D-glucopyranosyl-(1->2)-[β -D-glucopyranosyl-(1->4)]- α -L-arabinopyranosyl]oxy}-12-keto,16 β -hydroxy- β -amyrin (12b)

generated by co-expression of AsbAS1/SAD1, AsCYP51H10/SAD2, AsAAT1, AsUGT91G16 and AsTG1 in *N. benthamiana*.

| Carbon numbering scheme and selected COSY and HMBC | | | | | |
|--|------------------------------------|--|----------|------------------------------------|---------------------------------|
| | | | | | |
| Carbon # | ^{13}C δ (100 MHz) | ^1H δ (400 MHz) | Carbon # | ^{13}C δ (100 MHz) | ^1H δ (400 MHz) |
| 12 | 217.36 | / | 19 | 46.68 | 1.66 (1H, m) 0.99 (1H, m) |
| 1''' | 105.67 | 4.47 (1H, d, $J=7.7$) | 8 | 43.24 | / |
| 1' | 105.19 | 4.52 (1H, d, $J=5.7$) | 4 | 40.35 | / |
| 1'' | 104.49 | 4.62 (1H, d, $J=7.7$) | 1 | 39.93 | 1.62 (1H, m) 1.12 (1H, m) |
| 3 | 91.01 | 3.20 (1H, m) | 11 | 39.36 | 2.25 (2H, m) |
| 2' | 78.70 | 3.91 (1H, m) | 10 | 38.20 | / |
| 4' | 78.34 | 3.98 (1H, m) | 17 | 37.73 | / |
| 3''' | 78.19 | 3.27 (1H, m) | 15 | 36.87 | 1.87 (1H, m) 1.22 (1H, m) |
| 5' & 5'' | 78.06 | 3.35 (2H, m) | 18 | 35.99 | 2.65 (1H, dd, $J=13.5, 4.0$) |
| 3'' | 77.92 | 3.27 (1H, m) | 7 | 35.69 | 1.84 (1H, m) 1.43 (1H, m) |
| 2'' | 75.81 | 3.20 (1H, m) | 21 | 35.22 | 1.34 (1H, m) 1.16 (1H, m) |
| 2''' | 75.44 | 3.24 (1H, m) | 30 | 33.15 | 0.91 (3H, s) |
| 3' | 73.37 | 3.89 (1H, m) | 22 | 32.17 | 1.86 (1H, m) 1.31 (1H, m) |
| 4''' | 71.91 | 3.22 (1H, m) | 20 | 31.97 | / |
| 4'' | 71.50 | 3.26 (1H, m) | 27 | 29.53 | 1.41 (3H, s) |
| 15 | 66.75 | 3.91 (1H, m) | 24 | 28.68 | 1.09 (3H, s) |
| 5' | 64.63 | 4.12 (1H, dd, $J=12.0, 4.7$) 3.55 (1H, dd, $J=12.0, 2.0$) | 2 | 26.88 | 1.87 (1H, m) 1.74 (1H, m) |
| 6''' | 63.12 | 3.83 (1H, m) 3.61 (1H, m) | 29 | 25.15 | 0.91 (3H, s) |
| 6'' | 62.77 | 3.88 (1H, m) 3.66 (1H, m) | 28 | 22.14 | 0.79 (3H, s) |
| 13 | 59.99 | 1.92 (1H, m) | 26 | 20.18 | 0.94 (3H, s) |
| 5 | 56.82 | 0.88 (1H, m) | 6 | 19.16 | 1.59 (1H, m) 1.42 (1H, m) |
| 9 | 48.49 | 1.93 (1H, m) | 23 | 17.18 | 0.88 (3H, s) |
| 14 | 47.27 | / | 25 | 16.47 | 1.02 (3H, s) |

MeOH-*d*₄ [referenced to TMS]. Coupling constants are reported as observed and not corrected for second order effects. Assignments were made via a combination of ¹H, ¹³C, DEPT-135, COSY, DEPT-edited HSQC, HMBC. Where signals overlap ¹H δ is reported as the centre of the respective HSQC crosspeak. * anomeric ¹H δ are in close agreement with those reported for avenecin A-1: [1''] 4.63 (1H, d, *J*=7.8), [1'] 4.53 (1H, d, *J*=4.7), [1'''] 4.47 (1H, d, *J*=7.5), providing further support for the correct trisaccharide moiety being present (109).

Table S14. Single nucleotide mutations in the *astg1* gene of *sad3* mutants.

| Mutant | Mutation event | Predicted outcome |
|--------|----------------|-------------------|
| #1139 | G1800A | Intron/exon |
| #105 | G1705A | Intron/exon |
| #368 | G216A | Val-29 Met |
| #891 | C481T | Ala-88 Val |

Table S15. F₂ progeny phenotype ratio of *sad3* mutant #1139 crossed to the *A. strigosa* wild type parent (S75).

| <i>sad3</i> mutant | Root fluorescence phenotype | | χ^2 Wild type:mutant = 3:1 |
|--------------------|-----------------------------|---------|---------------------------------|
| | Wild type | Reduced | |
| #1139 | 142 | 48 | 0.007 (P> 0.05) |

Table S16. Primer sequences

| Name | Sequence | Annealing temperature | PCR cycles |
|-----------------------------------|---------------------------------|-----------------------|------------|
| UGT expression profiling primers: | | | |
| GAPDH-RT-PCR-F | GGTGGTCATTTTCAGCCCCTA | 58°C | 27 |
| GAPDH-RT-PCR-R | CTCCCACCTCTCCAGTCCTT | 58°C | 27 |
| SAD10FWD | GAGGGAGGAGTTGGAGAGGT | 58°C | 30 |
| SAD10REV | GGGCCACAGATCGATCCATT | 58°C | 30 |
| Frt-UGT99D1 (AsAAT1) | CGAGCACAACGTCCACGAGT | 58°C | 35 |
| Rrt-UGT99D1 (AsAAT1) | TTCGCCTCTACAGGTGGTGG | 58°C | 35 |
| Futr-UGT91G16like | AGGAGAGAGGGGTGGGACTA | 58°C | 35 |
| Rutr-UGT91G16bis | GGAACCATATTGAAAAATCGCTTA | 58°C | 35 |
| Fep-UGT99C4 | AGGCTGCCCTTGAAATAGT | 58°C | 30 |
| Rep-UGT99C4 | ACGTGTCCTTGGTCATCTCC | 58°C | 30 |
| Frp-UGT99C8 | CTGTGGTGCGTGAACACATT | 58°C | 30 |
| Rrp-UGT99C8 | TCCGATCTTGGGTAGTCTGC | 58°C | 30 |
| Futr-UGT703A5 | TGTGATACGCGGTGAGGTAA | 58°C | 30 |
| Rutr-UGT703A5 | AAAGCGAGCGAGGTACAAAA | 58°C | 30 |
| TG expression profiling primers: | | | |
| GAPDH-RT-PCR-F | GGTGGTCATTTTCAGCCCCTA | 55°C | 30 |
| GAPDH-RT-PCR-R | CTCCCACCTCTCCAGTCCTT | 55°C | 30 |
| Sad1-1-5 | ATGTGGAGGCTAACAAATAGG | 55°C | 30 |
| Sad1-2-3 | TATCTCATGACGATGTTCCG | 55°C | 30 |
| F-AsTG-8 | CTCGGGAGTCTACTCGACCA | 55°C | 30 |
| R-AsTG-8 | GGGTGTTTCCATTTGCGAGC | 55°C | 30 |
| Cloning primers: | | | |
| attB1F-1 | GGGGACAAGTTTGTACAAAAAAGCAGGCTTA | 50°C | 25 |

| Name | Sequence | Annealing temperature | PCR cycles |
|--|---|-----------------------|------------|
| attB2R-1 | GGGGACCACTTTGTACAAGAAAGCTGGGTA | 50°C | 25 |
| attLF-1 | TCGCGTTAACGCTAGCATGGATCTC | 50°C | 25 |
| attLR-2 | GTAACATCAGAGATTTTGAGACAC | 50°C | 25 |
| Fgw-UGT91G16 | AAAAAGCAGGCTTAATGGCCGCCTCTGCTTCC | 60°C | 18 |
| Rgw-UGT91G16 | GAAAGCTGGGTATCAGTCCATGTAAGACGTGAGCTGCTG | 60°C | 18 |
| Fgw-AsTG | AAAAAGCAGGCTTAATGGCACTGCTGCTCTGC | 60°C | 18 |
| Rgw-AsTG | AGAAAGCTGGGTATCACGCAGAGTCGTAATATTGT | 60°C | 18 |
| Fgw-nosigAsTG | AAAAAGCAGGCTTAATGGGAGACGTTGTGGTGGCG | 60°C | 18 |
| Primers used to amplify <i>AsUGT91G16</i> gene in avenacin-deficient mutants: | | | |
| F-UGT91G16-0816 | GCCCGCTACCTATTTGAATGGTGG | 67°C | 40 |
| RoutUGT91G16 | GTGTTGACCATGCACGAATCTCC | 67°C | 40 |
| Primers used to sequence <i>AsUGT91G16</i> gene in avenacin-deficient mutants: | | | |
| F-UGT91G16-0816 | GCCCGCTACCTATTTGAATGGTGG | | |
| Rutr-UGT91G16-bis | GGAACCATATTGAAAATCGCTTA | | |
| <i>AsUGT91G16</i> gene primers for F ₂ analysis: | | | |
| FUTR-UGT91G16-0516 | TGTTTTTGTAAAGCAGCGGGC | 67°C | 40 |
| RUTR-UGT91G16-0516 | AGGTAGTACTCGCTCGCT | 67°C | 40 |
| <i>AsUGT91G16</i> gene sequencing primers for F ₂ analysis: | | | |
| Rrt-UGT91G16 | ACGACCAGCTGAAGCTTGCC | | |
| Primers used to amplify <i>AsTG1</i> gene in avenacin-deficient mutants: | | | |
| R-AsTG-tot-1 | GCGCGGTCTCAAACCTTGTTT | 65°C | 40 |
| F-AsTG-3 | TGTCTTCCAGGCTAGTGGGA | 65°C | 40 |
| Primers used to sequence <i>AsTG1</i> gene in avenacin deficient mutants: | | | |

| Name | Sequence | Annealing temperature | PCR cycles |
|--|--|-----------------------|------------|
| R-AsTG-3 | TGCTGCAACATCTCCGGTAG | | |
| F-AsTG-4 | TGTTTCTCTTCAGCCTCCGG | | |
| F-AsTG-5 | GATTATAAGCAAGCCGCCGC | | |
| R-AsTG-6 | GCTTGAGATTGAAGGCGTGC | | |
| F-AsTG-7 | GGACTACCCTCCGGTGATGA | | |
| R-AsTG-7 | CAGCCCGTCCTGAATGAAGT | | |
| F-AsTG-8 | CTCGGGAGTCTACTCGACCA | | |
| R-AsTG-9 | TCTTGCCGACAAAGAGCCAT | | |
| F-AsTG-10 | ACTCCGCCAGATGGTACTCT | | |
| R-AsTG-10 | GTTGTTGGACCACCTAGCGA | | |
| <i>AsTG1</i> gene primers for F ₂ analysis: | | | |
| F-AsTG-7 | GGACTACCCTCCGGTGATGA | 66°C | 40 |
| R-AsTG-7 | CAGCCCGTCCTGAATGAAGT | 66°C | 40 |
| <i>AsTG1</i> sequencing primer for F ₂ analysis: | | | |
| R-AsTG-7 | CAGCCCGTCCTGAATGAAGT | | |
| C-terminal <i>AsTG1</i> :fluorescent protein fusion construct primers: | | | |
| Fgw-AsTGsig-FULL | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCACTGCTGCTCTGC | | |
| Fgw-AsTG-NOSIG-FULL | GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGGAGACGTTGTGGTGGCG | | |
| Rgw-AsTG-NOSTOP-FULL | GGGGACCACTTTGTACAAGAAAGCTGGGTACGCAGAGTCGTAATATTGTTTC | | |
| N-terminal GFP <i>AsUGT91G16</i> fusion construct primers: | | | |
| AsUGT91-NTGW | AAAAAGCAGGCTTATGGCCGCCTCTGCTTCC | | |
| Rgw-UGT91G16 | GAAAGCTGGGTATCAGTCCATGTAAGACGTGAGCTGCTG | 60°C | 18 |
| C-terminal GFP fusion <i>AsUGT91G16</i> construct primers: | | | |

| Name | Sequence | Annealing temperature | PCR cycles |
|--|---------------------------------------|-----------------------|------------|
| Fgw-UGT91G16 | AAAAAGCAGGCTTAATGGCCGCCTCTGCTTCC | 60°C | 18 |
| Rgw-UGT91-NOSTOP | AGAAAGCTGGGTAGTCCATGTAAGACGTGAGCTGCTG | | |
| Fluorescent fusion construct sequencing primers: | | | |
| midRFP-Rev | GAGCCGTA CTGGA ACTGAGG | | |
| midGFP-Rev | GTAGTTCCCGTCGTCCTTGA | | |
| midGFP-For | TCAAGGAGGACGGAAACATC | | |
| midRFP-For | CATCCCCGACTACTTGAAGC | | |

References

1. L. Song, L. Floria, Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* **4**, 48 (2015).
2. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
3. R. Wehrens, L. M. C. Buydens, Self- and Super-organizing Maps in R: The kohonen package. *J. Stat. Softw.* **21**(i5), 1-19 (2007).
4. D. M. Jones *et al.*, Spatio-temporal expression dynamics differ between homologues of flowering time genes in the allopolyploid *Brassica napus*. *Plant J.* **96**, 103–118 (2018).
5. R. M. E. Payne *et al.*, An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nat. Plants* **3**, 16208 (2017).
6. T. Kohonen, Essentials of the self-organizing map. *Neural Net.* **37**, 52–65 (2013).
7. A. Osbourn, B. Clarke, J. Dow, M. Daniels, Partial characterization of avenacinase from *Gaeumannomyces graminis* var. *avenae*. *Physiol. Mol. Plant Pathol.* **38**, 301–312 (1991).
8. K. Papadopoulou, R. E. Melton, M. Leggett, M. J. Daniels, A. E. Osbourn, Compromised disease resistance in saponin-deficient plants. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12923–12928 (1999).
9. P. Mylona *et al.*, *Sad3* and *Sad4* are required for saponin biosynthesis and root development in oat. *Plant Cell* **20**, 201–212 (2008).
10. G. T. Bryan *et al.*, DNA polymorphism and host range in the take-all fungus *Gaeumannomyces graminis*. *Mycol. Res.* **103**, 319–327 (1999).
11. J. Reed *et al.*, A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* **42**, 185–193 (2017).
12. M. A. Pallotta *et al.*, Marker assisted wheat breeding in the southern region of Australia. Proceedings of the Tenth International Wheat Genetics Symposium,

- Istituto Sperimentale per la Cerealicoltura, edited by N. E. Pogna, M. Romano, E. A. Pogna, Z. Galterio, 789-791 (2003).
13. X. Qi *et al.*, A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18848–18853 (2006).
 14. K. Haralampidis *et al.*, A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13431–13436 (2001).
 15. A. Leveau *et al.*, Towards take-all control: a C-21 β oxidase required for acylation of triterpene defence compounds in oat. *New Phytol.* **221**, 1544–1555 (2019).
 16. T. Louveau *et al.*, Analysis of Two New Arabinosyltransferases Belonging to the Carbohydrate-Active Enzyme (CAZY) Glycosyl Transferase Family1 Provides Insights into Disease Resistance and Sugar Donor Specificity. *Plant Cell* **30**, 3038-3057 (2018).
 17. S. T. Mugford *et al.*, A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell* **21**, 2473-2484 (2009).
 18. S. T. Mugford *et al.*, Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat. *Plant Cell* **25**, 1078-1092 (2013).
 19. M. J. Begley, L. Crombie, W. M. L. Crombie, D. A. Whiting, The isolation of avenacins A-1, A-2, B-1, and B-2, chemical defences against cereal 'take-all' disease. Structure of their 'aglycones', the avenestergenins, and their anhydro dimers. *J. Chem. Soc. Perkin Trans.* **1**, 1905–1915 (1986).
 20. S. Chojnacki, A. Cowley, J. Lee, A. Foix, R. Lopez, Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Res.* **45**, W550-W553 (2017).
 21. Y. Nishizaki *et al.*, *p*-hydroxybenzoyl-glucose is a zwitter donor for the biosynthesis of 7-polyacylated anthocyanin in *Delphinium*. *Plant Cell* **25**, 4150–4165 (2013).

22. S. Luang *et al.*, Rice Os9BGlu31 is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. *J. Biol. Chem.* **288**, 10111–10123 (2013).
23. Y. Matsuba *et al.*, A novel glucosylation reaction on anthocyanins catalyzed by acyl-glucose-dependent glucosyltransferase in the petals of carnation and delphinium. *Plant Cell* **22**, 3374–3389 (2010).
24. T. Miyahara *et al.*, Isolation of anthocyanin 7-O-glucosyltransferase from canterbury bells (*Campanula medium*). *Plant Biotechnology* **31**, 555–559 (2014).
25. T. Miyahara, M. Takahashi, Y. Ozeki, N. Sasaki, Isolation of an acyl-glucose-dependent anthocyanin 7-O-glucosyltransferase from the monocot *Agapanthus africanus*. *J. Plant Physiol.* **169**, 1321–1326 (2012).
26. A. Moglia *et al.*, Dual catalytic activity of hydroxycinnamoyl-coenzyme A quinate transferase from tomato allows it to moonlight in the synthesis of both mono- and dicaffeoylquinic acids. *Plant Physiol.* **166**, 1777–1787 (2014).
27. B. K. Nelson, X. Cai, A. Nebenführ, A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J.* **51**, 1126–1136 (2007).
28. B. J. Townsend, A. E. Osbourn, Root-expressed glycosyltransferases as candidates for glycosylation of triterpenoid saponins in a monocot. Direct Submission. (2008).
29. K. Yonekura-Sakakibara *et al.*, Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in *Arabidopsis thaliana*. *Plant J.* **69**, 154–167 (2012).
30. K. Yonekura-Sakakibara *et al.*, A flavonoid 3-O- glucoside:2"-O- glucosyltransferase responsible for terminal modification of pollen-specific flavonols in *Arabidopsis thaliana*. *Plant J.* **79**, 769–782 (2014).
31. A. Frydman *et al.*, The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. *Plant J.* **73**, 166–178 (2013).
32. F. Rojas Rodas *et al.*, Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-O-glucoside (1 → 6) rhamnosyltransferase. *Plant Mol. Biol.* **84**, 287–300 (2014).

33. Y.-H. Hsu *et al.*, Functional characterization of UDP-rhamnose-dependent rhamnosyltransferase involved in anthocyanin modification, a key enzyme determining blue coloration in *Lobelia erinus*. *Plant J.* **89**, 325–337 (2017).
34. M. Shibuya, K. Nishimura, N. Yasuyama, Y. Ebizuka, Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin i in *Glycine max*. *FEBS Lett.* **584**, 2258–2264 (2010).
35. R. Yano *et al.*, Isolation and characterization of the soybean Sg-3 gene that is involved in genetic variation in sugar chain composition at the C-3 position in soyasaponins. *Plant Cell Physiol.* **59**, 792–805 (2018).
36. Y. Morita *et al.*, Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose:anthocyanidin 3-O-glucoside-2"-O-glucosyltransferase, due to 4-bp insertions in the gene. *Plant J.* **42**, 353–363 (2005).
37. M. Nagatoshi *et al.*, UGT75L6 and UGT94E5 mediate sequential glucosylation of crocetin to crocin in *Gardenia jasminoides*. *FEBS Lett.* **586**, 1055–1061 (2012).
38. S. Sawada *et al.*, UDP-glucuronic acid:anthocyanin glucuronosyltransferase from red daisy (*Bellis perennis*) flowers. Enzymology and phylogenetics of a novel glucuronosyltransferase involved in flower pigment biosynthesis. *J. Biol. Chem.* **280**, 899–906 (2005).
39. S.-C. Jung *et al.*, Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. *Plant Cell Physiol.* **55**, 2177–2188 (2014).
40. M. Itkin *et al.*, The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7619–E7628 (2016).
41. M. Itkin *et al.*, Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* **341**, 175–179 (2013).
42. E. Ono, M. Ruike, T. Iwashita, K. Nomoto, Y. Fukui, Co-pigmentation and flavonoid glycosyltransferases in blue *Veronica persica* flowers. *Phytochemistry* **71**, 726–735 (2010).
43. K. Yonekura-Sakakibara, T. Tohge, R. Niida, K. Saito, Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in

- Arabidopsis* by transcriptome coexpression analysis and reverse genetics. *J. Biol. Chem.* **282**, 14932–14941 (2007).
44. H.-Y. Chen, X. Li, Identification of a residue responsible for UDP-sugar donor selectivity of a dihydroxybenzoic acid glycosyltransferase from *Arabidopsis* natural accessions. *Plant J.* **89**, 195–203 (2017).
 45. S. Witte, S. Moco, J. Vervoort, U. Matern, S. Martens, Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from *Hieracium pilosella* L. *Planta* **229**, 1135–1146 (2009).
 46. C. M. Kramer *et al.*, Cloning and regiospecificity studies of two flavonoid glucosyltransferases from *Allium cepa*. *Phytochemistry* **64**, 1069–1076 (2003).
 47. J. H. Kim *et al.*, Characterization of flavonoid 7-O-glucosyltransferase from *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* **70**, 1471–1477 (2006).
 48. F. Gandia-Herrero *et al.*, Detoxification of the explosive 2,4,6-trinitrotoluene in *Arabidopsis*: discovery of bifunctional O- and C-glucosyltransferases. *Plant J.* **56**, 963–974 (2008).
 49. J. M. Augustin *et al.*, UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze saponin 3-O-glucosylation in saponin-mediated insect resistance. *Plant Physiol.* **160**, 1881–1895 (2012).
 50. T. Vogt, R. Grimm, D. Strack, Cloning and expression of a cDNA encoding betanidin 5-O-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the *Solanaceae*. *Plant J.* **19**, 509–519 (1999).
 51. X. Zhao *et al.*, Functional characterisation of a new tea (*Camellia sinensis*) flavonoid glycosyltransferase. *J. Agric. Food Chem.* **65**, 2074–2083 (2017).
 52. Y. Zhong *et al.*, Developmentally regulated glucosylation of bitter triterpenoid in cucumber by the UDP-glucosyltransferase UGT73AM3. *Mol. Plant* **10**, 1000–1003 (2017).
 53. T. Sayama *et al.*, The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *Plant Cell* **24**, 2123–2138 (2012).

54. G. Xu, W. Cai, W. Gao, C. Liu, A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetic acid to yield glycyrrhizin. *New Phytol.* **212**, 123–135 (2016).
55. M. A. Naoumkina *et al.*, Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Cell* **22**, 850–866 (2010).
56. K. F. McCue *et al.*, Potato glycoesterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. *Phytochemistry* **68**, 327–334 (2007).
57. A. Trapero *et al.*, Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in *Crocus sativus*. *Plant Physiol.* **159**, 1335–1354 (2012).
58. T. Ito, S. Fujimoto, F. Suito, M. Shimosaka, G. Taguchi, C-glycosyltransferases catalyzing the formation of di-C-glucosyl flavonoids in citrus plants. *Plant J.* **91**, 187–198 (2017).
59. D. M. Priest *et al.*, Use of the glucosyltransferase UGT71B6 to disturb abscisic acid homeostasis in *Arabidopsis thaliana*. *Plant J.* **46**, 492–502 (2006).
60. C. E. Lim *et al.*, Improved resistance to oxidative stress by a loss-of-function mutation in the *Arabidopsis* UGT71C1 gene. *Mol. Cells* **25**, 368–375 (2008).
61. J. H. Ko *et al.*, Four glucosyltransferases from rice: cDNA cloning, expression, and characterization. *J. Plant Physiol.* **165**, 435–444 (2008).
62. M. Brazier-Hicks *et al.*, Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20238–20243 (2007).
63. A. Lanot *et al.*, The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in *Arabidopsis thaliana*. *Plant J.* **48**, 286–295 (2006).
64. Y. Pang, G. J. Peel, S. B. Sharma, Y. Tang, R. A. Dixon, A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14210–14215 (2008).
65. E. Ono *et al.*, Yellow flowers generated by expression of the aurone biosynthetic pathway. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11075–11080 (2006).

66. H. Jugde, D. Nguy, I. Moller, J. M. Cooney, R. G. Atkinson, Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. *FEBS J.* **275**, 3804–3814 (2008).
67. J. Ogata, Y. Kanno, Y. Itoh, H. Tsugawa, M. Suzuki, Plant biochemistry: anthocyanin biosynthesis in roses. *Nature* **435**, 757–758 (2005).
68. T. Nakatsuka, M. Nishihara, UDP-glucose:3-deoxyanthocyanidin 5-O-glucosyltransferase from *Sinningia cardinalis*. *Planta* **232**, 383–392 (2010).
69. Y. Hirade *et al.*, Identification and functional analysis of 2-hydroxyflavanone C-glucosyltransferase in soybean (*Glycine max*). *FEBS Lett.* **589**, 1778–1786 (2015).
70. M. Brazier-Hicks *et al.*, The C-glycosylation of flavonoids in cereals. *J. Biol. Chem.* **284**, 17926–17934 (2009).
71. P. Jones, B. Messner, J.-I. Nakajima, A. R. Schaffner, K. Saito, UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana*. *J. Biol. Chem.* **278**, 43910–43918 (2003).
72. W. Sun *et al.*, Biochemical and molecular characterization of a flavonoid 3-O-glucosyltransferase responsible for anthocyanins and flavonols biosynthesis in *Freesia hybrida*. *Front. Plant Sci.* **7**, 410 (2016).
73. M. Mato *et al.*, Isolation and characterization of a cDNA clone of UDP-galactose: flavonoid 3-O-galactosyltransferase (UF3GaT) expressed in *Vigna mungo* seedlings. *Plant Cell Physiol.* **39**, 1145–1155 (1998).
74. C. M. Ford, P. K. Boss, P. B. Hoj, Cloning and characterization of *Vitis vinifera* UDP-glucose:flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize *Bronze-1* locus that may primarily serve to glucosylate anthocyanidins *in vivo*. *J. Biol. Chem.* **273**, 9224–9233 (1998).
75. B. Hou, E.-K. Lim, G. S. Higgins, D. J. Bowles, N-glycosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. *J. Biol. Chem.* **279**, 47822–47832 (2004).
76. T. K. Franks *et al.*, A seed coat cyanohydrin glucosyltransferase is associated with bitterness in almond (*Prunus dulcis*) kernels. *Funct. Plant Biol.* **35**, 236–246 (2008).

77. K. S. Hansen *et al.*, The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. *Phytochemistry* **64**, 143–151 (2003).
78. E.-K. Lim, D. A. Ashford, B. Hou, R. G. Jackson, D. J. Bowles, *Arabidopsis* glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol. Bioeng.* **87**, 623–631 (2004).
79. A. Richman *et al.*, Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. *Plant J.* **41**, 56–67 (2005).
80. U. von Rad, R. Huttli, F. Lottspeich, A. Gierl, M. Frey, Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *Plant J.* **28**, 633–642 (2001).
81. J. Ross, Y. Li, E. Lim, D. J. Bowles, Higher plant glycosyltransferases. *Genome Biol.* **2**, 3004 (2001).
82. B. Wang *et al.*, UGT87A2, an *Arabidopsis* glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. *New Phytol.* **194**, 666–675 (2012).
83. V. B. Tognetti *et al.*, Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase UGT74E2 modulates *Arabidopsis* architecture and water stress tolerance. *Plant Cell* **22**, 2660–2679 (2010).
84. A. Owatworakit *et al.*, Glycosyltransferases from oat (*Avena*) implicated in the acylation of avenacins. *J. Biol. Chem.* **288**, 3696–3704 (2013).
85. W. Schweiger *et al.*, Validation of a candidate deoxynivalenol- inactivating UDP-glucosyltransferase from barley by heterologous expression in yeast. *Mol. Plant. Microbe Interact.* **23**, 977–986 (2010).
86. H. Michlmayr *et al.*, Biochemical characterization of a recombinant UDP-glucosyltransferase from rice and enzymatic production of deoxynivalenol-3-O- β -d-glucoside. *Toxins* **7**, 2685–2700 (2015).
87. D. Meesapyodsuk, J. Balsevich, D. W. Reed, P. S. Covello, Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding β -amyrin synthase and a triterpene carboxylic acid glucosyltransferase. *Plant Physiol.* **143**, 959–969 (2007).

88. J. B. Szerszen, K. Szczyglowski, R. S. Bandurski, *iaglu*, a gene from *Zea mays* involved in conjugation of growth hormone indole-3-acetic acid. *Science* **265**, 1699–1701 (1994).
89. A. Eudes *et al.*, Metabolism of the folate precursor *p*-aminobenzoate in plants: glucose ester formation and vacuolar storage. *J. Biol. Chem.* **283**, 15451–15459 (2008).
90. A. R. Moraga, A. T. Mozos, O. Ahrazem, L. Gomez-Gomez, Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation. *BMC Plant Biol.* **9**, 109 (2009).
91. M. Yahyaa *et al.*, Identification and characterization of UDP-glucose:phloretin 4'-O-glycosyltransferase from *Malus x domestica* Borkh. *Phytochemistry* **130**, 47–55 (2016).
92. L. Tian, J. W. Blount, R. A. Dixon, Phenylpropanoid glycosyltransferases from osage orange (*Maclura pomifera*) fruit. *FEBS Lett.* **580**, 6915–6920 (2006).
93. M. Yamazaki *et al.*, Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. *J. Biol. Chem.* **274**, 7405–7411 (1999).
94. F. He *et al.*, Molecular and biochemical characterization of the UDP-glucose: Anthocyanin 5-O-glucosyltransferase from *Vitis amurensis*. *Phytochemistry* **117**, 363–372 (2015).
95. C. Milkowski, A. Baumert, D. Strack, Identification of four *Arabidopsis* genes encoding hydroxycinnamate glucosyltransferases. *FEBS Lett.* **486**, 183–184 (2000).
96. J. Mittasch, S. Mikolajewski, F. Breuer, D. Strack, C. Milkowski, Genomic microstructure and differential expression of the genes encoding UDP-glucose:sinapate glucosyltransferase (UGT84A9) in oilseed rape (*Brassica napus*). *Theor. Appl. Genet.* **120**, 1485–1500 (2010).
97. S. Lunkenbein *et al.*, Cinnamate metabolism in ripening fruit. Characterization of a UDP-glucose:cinnamate glucosyltransferase from strawberry. *Plant Physiol.* **140**, 1047–1058 (2006).
98. N. Sasaki *et al.*, Identification of the glucosyltransferase that mediates direct flavone C-glucosylation in *Gentiana triflora*. *FEBS Lett.* **589**, 182–187 (2015).

99. M. Kita *et al.*, Molecular cloning and characterization of a novel gene encoding limonoid UDP-glucosyltransferase in *Citrus*. *FEBS Lett.* **469**, 173–178 (2000).
100. N. Sasaki *et al.*, Isolation and characterization of cDNAs encoding an enzyme with glucosyltransferase activity for cyclo-DOPA from four o'clocks and feather cockscombs. *Plant Cell Physiol.* **46**, 666–670 (2005).
101. R. C. Martin, M. C. Mok, J. E. Habben, D. W. Mok, A maize cytokinin gene encoding an O-glucosyltransferase specific to *cis*-zeatin. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5922–5926 (2001).
102. S. Masada *et al.*, Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from *Catharanthus roseus*. *Plant Cell Physiol.* **50**, 1401–1415 (2009).
103. F. Rojas Rodas *et al.*, Cloning and characterization of soybean gene *Fg1* encoding flavonol 3-O-glucoside/galactoside (1→6) glucosyltransferase. *Plant Mol. Biol.* **92**, 445–456 (2016).
104. S. Di *et al.*, Linkage mapping, molecular cloning and functional analysis of soybean gene *Fg3* encoding flavonol 3-O-glucoside/galactoside (1 → 2) glucosyltransferase. *BMC Plant Biol.* **15**, 126 (2015).
105. F. Brugliera *et al.*, Isolation and characterization of a cDNA clone corresponding to the *Rt* locus of *Petunia hybrida*. *Plant J.* **5**, 81–92 (1994).
106. A. Noguchi *et al.*, Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases. *Plant J.* **54**, 415–427 (2008).
107. L. Crombie, W. M. L. Crombie, D. A. Whiting, Isolation of avenacins A-1, A-2, B-1, and B-2 from oat roots: structures of their 'aglycones', the avenestergenins. *J. Chem. Soc. Chem. Commun.* **4**, 244–246 (1984).
108. L. Crombie, W. M. L. Crombie, D. A. Whiting, Structures of the four avenacins, oat root resistance factors to 'take-all' disease. *J. Chem. Soc., Chem. Commun.* **4**, 246–248 (1984).
109. L. Crombie, W. M. L. Crombie, D. A. Whiting, Structures of the oat root resistance factors to 'take-all' disease, avenancins A-1, A-2, B-1 and B-2 and their companion substances. *J. Chem. Soc. Perkin Trans.* **1**, 1917–1922 (1986).

110. K. Geisler *et al.*, Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3360–3367 (2013).
111. A. C. Kemen *et al.*, Investigation of triterpene synthesis and regulation in oats reveals a role for β -amyrin in determining root epidermal cell patterning. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8679–8684 (2014).
112. D. Andersson *et al.*, Myrosinases from root and leaves of *Arabidopsis thaliana* have different catalytic properties. *Phytochemistry* **70**, 1345-1354 (2009).
113. M. M. Islam *et al.*, Myrosinases, TGG1 and TGG2, redundantly function in ABA and MeJA signaling in *Arabidopsis* guard cells. *Plant Cell Physiol.* **50**, 1171-1175 (2009).
114. Z. Xu *et al.*, Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* **55**, 343–367 (2004).
115. P. Bednarek *et al.*, A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* **323**, 101–106 (2009).
116. C. Zamioudis, J. Hanson, C. M. Pieterse, beta-Glucosidase BGLU42 is a MYB72-dependent key regulator of rhizobacteria-induced systemic resistance and modulates iron deficiency responses in *Arabidopsis* roots. *New Phytol.* **204**, 368-379 (2014).
117. L. L. Escamilla-Treviño *et al.*, *Arabidopsis thaliana* β -Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry* **67**, 1651-1660 (2006).
118. H. Ishihara *et al.*, Natural variation in flavonol accumulation in *Arabidopsis* is determined by the flavonol glucosyltransferase BGLU6. *J. Exp. Bot.* **67**, 1505–1517 (2016).
119. T. Miyahara, R. Sakiyama, Y. Ozeki, N. Sasaki, Acyl-glucose-dependent glucosyltransferase catalyzes the final step of anthocyanin formation in *Arabidopsis*. *J. Plant Physiol.* **170**, 619–624 (2013).
120. E. R. Moellering, B. Muthan, C. Benning, Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science* **330**, 226–228 (2010).

121. S. Gus-Mayer, H. Brunner, H. A. Schneider-Poetsch, W. Rüdiger, Avenacosidase from oat: purification, sequence analysis and biochemical characterization of a new member of the BGA family of β -glucosidases. *Plant Mol. Biol.* **26**, 909–921 (1994).
122. A. Falk, L. Rask, Expression of a zeatin-*O*-glucoside-degrading beta-glucosidase in *Brassica napus*. *Plant Physiol.* **108**, 1369-1377 (1995).
123. M. Mizutani *et al.*, Cloning of β -primeverosidase from tea leaves, a key enzyme in tea aroma formation. *Plant Physiol.* **130**, 2164–2176 (2002).
124. A. Raychaudhuri, P. A. Tipton, Cloning and expression of the gene for soybean hydroxyisourate hydrolase. Localization and implications for function and mechanism. *Plant Physiol.* **130**, 2061–2068 (2002).
125. Y. Hua, S. Sansenya, C. Saetang, S. Wakuta, J.R. Ketudat Cairns, Enzymatic and structural characterization of hydrolysis of gibberellin A4 glucosyl ester by a rice beta-D-glucosidase. *Arch. Biochem. Biophys.* **537**, 39-48 (2013).
126. R. Opassiri *et al.*, Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 beta-glucosidase. *BMC Plant Biol.* **6**, 33 (2006).
127. D. P. Dharmawardhana, B. E. Ellis, J. E. Carlson, cDNA cloning and heterologous expression of coniferin β -glucosidase. *Plant Mol. Biol.* **40**, 365–372 (1999).
128. J.P. Xue, M. Lenman, A. Falk, L. Rask, The glucosinolate-degrading enzyme myrosinase in *Brassicaceae* is encoded by a gene family. *Plant Mol. Biol.* **18**, 387-398 (1992).
129. B. Mo, J.D. Bewley, Beta-mannosidase (EC 3.2.1.25) activity during and following germination of tomato (*Lycopersicon esculentum* Mill.) seeds. Purification, cloning and characterization. *Planta* **215**, 141-152 (2002).
130. M. Cicek, A. Esen, Cloning and Sequencing of a cDNA Coding for Beta-Glucosidase (Dhurrinase) from *Sorghum bicolor* (L.) Moench 1 (Accession No. U33817) (PGR 95-097). *Plant Physiol.* **109**, 1497 (1995).
131. E. Oxtoby, M. A. Dunn, A. Pancoro, M. A. Hughes, Nucleotide and derived amino acid sequence of the cyanogenic β -glucosidase (linamarase) from white clover (*Trifolium repens* L.). *Plant Mol. Biol.* **17**, 209-219 (1991).
132. A. Esen, Purification and partial characterization of maize (*Zea mays* L.) beta-glucosidase. *Plant Physiol.* **98**, 174-182 (1992).

