A willow UDP-glycosyltransferase involved in salicinoid biosynthesis

Satish Kulasekaran, Sergio Cerezo-Medina, Claudia Harflett, Charlotte Lomax, Femke de Jong[1], Amelie Rendour, Gianluca Ruvo, Steven J. Hanley, Michael H. Beale and Jane L. Ward*

Computational and Analytical Sciences Department, Rothamsted Research, West Common, Harpenden, Hertfordshire, AL5 2JQ, UK. [1]Present Section Plant Cell Biology, Swammerdam Institute for Life Sciences (SILS), University of Amsterdam, Amsterdam, Netherlands.

Jane.ward@rothamsted.ac.uk; Mike.beale@rothamsted.ac.uk

Emails:

Satish.kulasekaran@rothamsted.ac.uk; Sergio.cerezo-medina@rothamsted.ac.uk; Claudia.harflett@rothamsted.ac.uk; Charlotte.lomax@rothamsted.ac.uk; f.dejong@uva.nl, Amelie.rendour@rothamsted.ac.uk; Gianluca.ruvo@rothamsted.ac.uk; Steve.hanley@rothamsted.ac.uk; Mike.beale@rothamsted.ac.uk;

*Jane.ward@rothamsted.ac.uk.

Highlight: We report the cloning, and functional characterisation of willow glycosytransferase enzymes involved in the biosynthesis of salicinoids, metabolites with an historic role in the development of aspirin.

Abstract

The salicinoids are phenolic glycosides that are characteristic secondary metabolites of the Salicaceae, particularly willows and poplars. Despite the well-known pharmacology of salicin, that led to the development of aspirin over 100 years ago, the biosynthetic pathways leading to salicinoids have yet to be defined. Here, we describe the identification, cloning and biochemical characterisation of SpUGT71L2 and SpUGT71L3, - isozymic glycosyl transferases from *Salix purpurea* – that function in the glucosylation of *ortho*-substituted phenols. The best substrate *in vitro* was salicyl-7-benzoate. Its product, salicyl-7-benzoate glucoside, was shown to be endogenous in poplar and willow. Together they are inferred to be early intermediates in the biosynthesis of salicortin and related metabolites *in planta*. The role of this UGT was confirmed *via* the metabolomic analysis of transgenic plants produced by RNAi knockdown of the poplar orthologue (UGT71L1) in the hybrid clone *Populus tremula x P. alba*, INRA 717 1-B4.

Key words: Poplar; salicin; salicinoid; salicortin; salicyl-7-benzoate; Salix; tremulacin; UGT

Introduction

Phenolic glycosides are products of phenylalanine and tyrosine metabolism that are widely distributed in the plant kingdom, with particular abundance in the Salicaceae family. *Salix* sp. are well-known for salicin (Figure 1), a phenolic glucoside that formed the basis of pharmaceutical development leading to aspirin (Desborough and Keeling, 2017). The salicinoids form a sub-group of the phenylalanine derived phenolics and are structurally characterised by the salicyl alcohol (saligenin) base unit that is glucosylated at C-2 to form salicin (Figure 1). The expanding palette of known salicinoids comprise more complex structures that are further decorated with acyl groups (e.g. acetyl, benzoyl, 4-coumaroyl) on the glucose moiety and/or esterification at C-7 of the salicyl alcohol base (e.g. salicyloyl, feruloyl, 4-coumaroyl and HCH) (Boeckler *et al* 2011). HCH, 1-hydroxy-6-oxocyclohex-2-ene carboxylate, is a biosynthetically intriguing module that occurs in salicortin and tremulacin (Figure 1), compounds that are often abundant and generally widespread across the Salicaceae. Moreover, the HCH group can also occur on the glucose moiety (e.g. 6′-HCH-salicortin; 6′-HCH-tremulacin) (Boeckler *et al* 2011; Keefover-Ring *et al*, 2014) but has also been reported as a substituent on flavonoids (e.g. HCH-catechin)(Hsu *et al* 1985).

The biosynthetic pathway to salicinoids, particularly the formation of the HCH-group is currently unknown. Early work by Zenk in *Salix purpurea* indicated that salicyl alcohol and salicin arise from phenylalanine *via* cinnamate, *ortho*-coumarate and salicylaldehyde. (Zenk 1967). More recently ^{13}C labelling studies in *Populus nigra* suggested that both rings of salicortin arise from phenylalanine *via* the benzoate pathway and that salicin is not a precursor of salicortin (Babst *et al* 2010). A network of pathways to salicortin was put forward, including the possible intermediacy of benzylbenzoate, a volatile metabolite of benzoate metabolism that has been extensively studied in petunia (Boatright *et al* 2004). More evidence for the involvement of benzylbenzoate and/or the analogue salicyl-7-benzoate (Figure 1), was the demonstration (Chedgy *et al* 2015) that two acyltransferase enzymes, PtBEBT (benzoyl-CoA: benzyl alcohol O-benzoyltransferase) and PtSABT

(benzoyl-CoA: salicyl alcohol O-benzoyltransferase), cloned from *Populus trichocarpa*, and expressed in *E. coli*, were capable of producing benzylbenzoate and salicyl-7-benzoate from benzoylCoA condensing with benzyl alcohol and salicyl alcohol respectively. Although no direct evidence for the role of these acyltransferases *in planta* has yet been presented, the observed down-regulation of these genes in co-expression analysis of transgenic poplar plants with upregulated proanthocyanin biosynthesis, in concert with reduced salicortin and tremulacin, (Chedgy *et al*, 2015) suggested that they may operate in the salicinoid pathway. The salicinoid pathway requires a glucosylation step to generate the majority of known metabolites. To study this, we have been investigating the role of UGTs (Uridine diphosphate dependent Glucosyl Transferases) in the biosynthesis of salicortin in willow species. Recently, we have characterised a family of cyclodimeric salicinoids that contain glucosyl and acylglucosyl moieties (Ward *et al* 2020). These compounds apparently arise from a spontaneous Diels-Alder coupling of potential transient intermediates in salicortin/tremulacin biosynthesis, indicating that the glucosyl addition and further acylation of that glucose may be earlier steps in the pathway, preceding the formation of the HCH ring.

Tsai *et al* (2011) had previously suggested a role for a glycosyl transferase, PtUGT71L1 in salicinoid production in poplar roots, and very recently Fellenberg *et al* (2020) have demonstrated, via gene-editing, that this enzyme plays a key role in salicortin biosynthesis in roots, and, via *E.coli* expression, that salicyl-7-benzoate was a likely substrate. This new report has prompted us to describe our own parallel studies, that have independently led us to the willow homologues of this poplar UGT and also the demonstration, using RNAi knockdown of *PtUGT71L1* in transgenic poplar, of the role of these UGTs in salicinoid biosynthesis in the aerial parts of these species.

Materials and Methods

Amino acid sequence analysis

The sequence data for SapurV1A.1999s0070 (Sp*UGT71L2*), SapurV1A.0398s0160 (Sp*UGT71L3*), , Potri.016G014500 (Pt*UGT71L1*) were obtained from Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html). Multiple sequence alignment and percent identity matrices were constructed with Geneious v. 10.1.3 using the ClustalW alignment algorithm. Sequence analysis for intracellular targeting was carried out with SignalP-5.0, TargetP-2.0 and TMHMM software (http://www.cbs.dtu.dk/services/).

Protein expression and purification

The UGT's were expressed as a 6 X His fusion protein from the expression vector pPROEX HTa. Sp*UGT71L2* and Sp*UGT71L3* were single exon genes and were PCR amplified from genomic DNA prepared from *Salix purpurea* 'Uralensis' (NWC844) leaves. All the primers used in this study are listed in Table S1. The 1.4-kb PCR products of Sp*UGT71L2* and Sp*UGT71L3* were cloned between *Nco1* and *EcoR1* sites. The recombinant plasmids encoded modified fusion protein having a 6 X His tag at the N- terminus and were then transformed in *E. coli* BL21 (DE3) for expression. The transformed cells were grown in LB medium at 37°C until they reached an OD 600 of 0.6 and then 0.1 mM IPTG was added to induce protein production overnight at 16°C. Induction of protein was confirmed by SDS-polyacrylamide gel electrophoresis. Cells were harvested by centrifuging at 5000 rpm for 20 min at 4°C and the pellet was resuspended in buffer comprising 20 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM PMSF and a tablet of protease inhibitor cocktail (Roche), followed by disruption with a sonicator. The lysate was then centrifuged at 20,000 g for 30 min and the supernatant was loaded onto a pre-equilibrated gravity flow cartridge containing Ni-NTA resin (GE Life Sciences) equilibrated with buffer containing 15 mM imidazole, 50 mM sodium phosphate and 300 mM NaCl. The loaded column was then washed with 10 column volumes

of 15 mM imidazole 50 mM sodium phosphate and 300 mM NaCl. Proteins were eluted with buffer containing 250 mM imidazole. To remove unspecific protein contaminants a second identical purification step was performed using cobalt resin (GE Life Sciences) and protein was eluted in 150 mM imidazole. The eluted protein was concentrated using Vivaspin® 6 Centrifugal Concentrator and snap frozen. The protein aliquots were then stored at -80°C. Protein concentration was estimated by Nanodrop using the molar extinction coefficients and molecular weights which were calculated using Protparam (https://web.expasy.org/protparam/).

Synthesis of salicyl-7-benzoate and salicyl-2-benzoate

Saligenin (1.24 g, 0.01 moles) was dissolved in dichloromethane (50 mL) and the mixture was cooled to 0°C. Triethylamine (1.44 mL, 0.01 moles) and benzoyl chloride (1.17 mL, 0.01 moles) were added sequentially and the resultant mixture warmed to room temperature and stirred for 24 hours. The reaction mixture was quenched with ammonium chloride solution (100 mL). The organic layer was washed with water (100 mL), separated and dried over magnesium sulphate. Removal of dichloromethane in vacuo returned a brown oil (0.872 g). Flash chromatography was carried out using hexane:ethyl acetate mixtures. An initial elution with 100% hexane (150 mL) was followed by 6 further elutions (150 mL) increasing the ethyl acetate composition by 5% each time until a concentration of 30% ethyl acetate was achieved. A total of 65 fractions (10mL) were collected. Salicyl-7-benzoate (0.305g) eluted in fractions 48-58. Additional products from the reaction included salicyl-2,7-dibenzoate (0.561g) eluting in fractions 38-46 and salicyl-2-benzoate (0.230g) eluting in fractions 60-65. Products were characterised by ^{1}H-NMR at 600 MHz in both $CDCl_3$ and D_2O:CD_3OD (4:1). After structural confirmation by ^{1}H-NMR the products were immediately dissolved in DMSO to preserve stability, to generate 100 mg/mL stock solutions for use in substrate testing assays and kinetic analysis.

Salicyl-7-benzoate: ^{1}H-NMR (600 MHz, $CDCl_3$) δ: 8.074 (2H, d, *J* = 7.81 Hz, H-10 and H-14), 7.586 (1H, t, *J* = 7.45 Hz, H-12), 7.449 (2H t, *J* = 7.7 Hz, H-11 and H-13), 7.372 (1H, dd, *J* = 1.2 Hz and *J* = 7.45 Hz, H-6), 7.300 (1H, m, H-4), 6.981 (1H, d, *J* = 8.17 Hz, H-3), 6.938 (1H, t, *J* = 7.36 Hz, H-5), 5.388 (2H, s, 7-H_2). ^{1}H-NMR (600 MHz, CD_3OD:D_2O (8:2) δ: 8.074 (2H, d, *J* = 7.81 Hz, H-10 and H-14), 7.586 (1H, t, *J* = 7.45 Hz, H-12), 7.449 (2H t, *J* = 7.7 Hz, H-11 and H-13), 7.372 (1H, dd, *J* = 1.2 Hz and *J* = 7.45 Hz, H-6), 7.300 (1H, m, H-4), 6.981 (1H, d, *J* = 8.17 Hz, H-3), 6.938 (1H, t, *J* = 7.36 Hz, H-5), 5.388 (2H, s, 7-H_2).

Salicyl-2-benzoate: ^{1}H-NMR (600 MHz, $CDCl_3$) δ: 8.229 (2H, d, *J* = 7.45 Hz, H-3' and H-7'), 7.674 (1H, t, *J* = 7.46 Hz, H-5'), 7.544 (2H t, *J* = 7.81 Hz, H-4' and H-6'), 7.564 (1H, dd, *J* = 1.82 Hz and *J* = 6.90 Hz, H-6), 7.400 (1H, m, H-4), 7.330 (1H, t, *J* = 7.45 Hz, H-5), 7.202 (1H, d, *J* = 7.99 Hz, H-3), 4.660 (2H, s, 7-H_2). ^{1}H-NMR (600 MHz, CD_3OD:D_2O (8:2) δ: 8.230 (2H, dd, *J* = 1.25 Hz and 8.34 Hz, H-3' and H-7'), 7.781 (1H, tt, *J* = 1.42 Hz and 7.85 Hz, H-5'), 7.630 (2H t, *J* = 7.74 Hz, H-4' and H-6'), 7.574 (1H, dd, *J* = 1.31 Hz and *J* = 7.63 Hz, H-6), 7.463 (1H, td, *J* = 1.85 Hz and 7.74 Hz, H-4), 7.419 (1H, td, *J* = 1.09 Hz and 7.63 Hz, H-5), 7.253 (1H, dd, *J* = 1.31 Hz and 7.85 Hz, H-3), 4.620 (2H, s, 7-H_2).

Salicyl-2,7-dibenzoate: ^{1}H-NMR (600 MHz, $CDCl_3$) δ: 8.161 (2H, dd, *J* = 1.09 Hz and 7.99 Hz, H-3' and H-7'), 7.957 (2H, dd, *J* = 1.09 Hz and 7.99 Hz, H-10 and H-14), 7.611 (1H, tt, *J* = 1.27 Hz and 7.45 Hz, H-5'), 7.576 (1H, dd, *J* = 1.27 Hz and 7.45 Hz, H-6), 7.518 (1H tt, *J* = 1.45 Hz and 7.45 Hz, H-12), 7.472 (2H, t, *J* = 7.81 Hz, H-4' and H-6'), 7.446 (1H, td, *J* = 1.63 Hz and 7.81 Hz, H-4), 7.378 (2H, t, *J* = 7.81 Hz, H-11 and H-13), 7.320 (1H, td, *J* = 0.91 Hz and 7.45 Hz, H-5), 7.277 (1H, d, *J* = 8.36 Hz, H-3), 5.407 (2H, s, 7-H_2).

Isolation of salicyl-7-benzoate glucoside, benzylsalicylate glucoside and 6'-HCH salicin

Freeze-dried, pre-milled leaves from the side branches of *Populus trichocarpa* 'ManMillan' (75 mg) were extracted in H_2O:MeOH (4:1) at 50 °C (water bath) for 10 minutes. After extraction, the solution was cooled and centrifuged (13,000 rpm). The supernatant was removed to a clean vial for purification of target peaks by reversed phase HPLC. 8 injections (100 µL each) were introduced into an Agilent 1100 HPLC system equipped with a quaternary pump and diode array detector. Separation was achieved using an Ascentis C18 column (5µm x 250 mm, Supelco, UK). A gradient of water and acetonitrile, each containing 0.1% formic acid, was used starting at 5% acetonitrile and increasing to 34% acetonitrile over 70 minutes using a constant flow rate of 1 mL/min. Peaks were identified using a wavelength of 254 nm and fractions from multiple runs were combined and evaporated using a Speedvac concentrator (Genevac, Suffolk, UK).

Benzylsalicylate glucoside: ^{1}H-NMR (600 MHz, D_2O:CD_3OD) δ: 7.816 (dd, *J* = 1.75 Hz and 7.78 Hz, H-10), 7.615 (m, H-12), 7.543 (d, *J* = 8.60 Hz, H-2 and H-6), 7.481 (t, *J* = 7.23 Hz, H-3 and H-5), 7.452 (m, H-4), 7.317 (d, *J* = 8.11 Hz, H-13), 7.217 (td, *J* = 0.88 Hz and 7.78 Hz, H-11), 5.401 (s, H_2-7), 5.126 (d, *J* = 7.78 Hz, H-1'), 3.878 (dd, *J* = 2.25 Hz and 12.44 Hz, H-6'), 3.698 (dd, *J* = 5.81 Hz and 12.44 Hz, H-6'), 3.562 (m, H-3' and H-5'), 3.485 (dd, *J* = 7.78 Hz and 9.43 Hz, H-2'), 3.441 (t, *J* = 9.43 Hz, H-4'), ^{13}C-NMR (100 MHz, D_2O:CD_3OD) δ: 63.72 (C-6'), 69.87 (C-7), 71.92 (C-4'), 75.84 (C-2'), 78.91 (C-3', C-5'), 103.07 (C-1'), 119.37 (C-13), 123.89 (C-9), 125.90 (C-11), 131.45 (C-2, C-6), 131.71 (C-3, C-4, C-5), 133.79 (C-10), 137.06 (C-12), 138.69 (C-1), 158.17 (C-14), 170.19 (C-8).

Salicyl-7-benzoate glucoside (Desoxysalireposide): ^{1}H-NMR (600 MHz, D_2O:CD_3OD) δ: 8.022 (dd, *J* = 1.21 Hz and 8.33 Hz, H-10 and H-14), 7.685 (tt, *J* = 1.21 Hz and 7.56 Hz, H-12), 7.537 (dd, *J* = 7.56 Hz and 8.33 Hz, H-11 and H-13), 7.514 (dd, *J* = 1.75 Hz and 7.56 Hz, H-6), 7.440 (td, *J* = 1.53 Hz and 8.01 Hz, H-4), 7.252 (d, *J* = 8.11 Hz, H-3), 7.181 (td, *J* =

0.77 Hz and 7.45 Hz, H-5), 5.602 (d, *J* = 12.28 Hz, H-7), 5.423 (d, *J* = 12.28 Hz, H-7), 5.141 (m, H-1'), 3.864 (dd, *J* = 2.25 Hz and 12.33 Hz, H-6'), 3.631 (dd, *J* = 5.86 Hz and 12.33 Hz, H-6'), 3.393 (t, *J* = 9.76 Hz, H-4'), 3.590 – 3.540 (m, H-2', H-3', H-5').

6'-HCH-salicin: ^{1}H-NMR (600 MHz, D_2O:CD_3OD) δ: 7.416 (dd, *J* = 1.53 Hz and 7.56 Hz, H-6), 7.395 (td, *J* = 1.64 Hz and 8.00 Hz, H-5), 7.172 (td, *J* = 0.66Hz and 7.34 Hz, H-4), 7.150 (d, *J* = 8.44 Hz, H-3), 6.206 (dt, *J* = 3.95 Hz and 9.76 Hz H-4"), 5.685 (dt, *J* = 1.75 Hz and 9.76 Hz, H-3"), 5.120 (m, H-1'), 4.752 (d, *J* = 12.93 Hz H-7), 4.611 (d, *J* = 12.93 Hz, H-7), 4.583 (dd, *J* = 2.19 Hz and 12.17 Hz, H-6'), 4.310 (dd, *J* = 6.52 Hz and 12.11 Hz, H-6'), 3.801 (ddd, *J* = 2.08 Hz, 6.50 Hz and 9.87 Hz, H-5'), 3.570-3.620 (m, H-2' and H-3'), 3.425 (m, H-4'), 2.830 (m, H-6"), 2.616-2.474 (m, H-6" and H_2-5"). ^{13}C-NMR (100 MHz, D_2O:CD_3OD) δ: 28.81 (C-5"), 38.39 (C-6"), 62.06 (C-7), 67.85 (C-6'), 72.32 (C-4'), 75.62 (C-2'), 76.65 (C-5'), 78.35 (C-3'), 81.80 (C-2"), 103.73 (C-1'), 118.19 (C-3), 124.76 (C-1), 126.27 (C-4), 129.31 (C-3"), 132.47 (C-5), 132.60 (C-6), 135.72 (C-4"), 157.39 (C-2), 172.76 (C-1"), 209.28 (C-7").

Enzyme assays

Glucosyl transferase activity assays were carried out as described by Lim *et al* (2002) with some modifications. A typical assay consisted of 5 μg of recombinant protein, 50 mM sodium phosphate pH 7.5, 1 mM ß-mercaptoethanol, 5 mM UDP-glucose and 1 mM phenolic substrate in a 200μL reaction volume. The reaction mix was incubated at 30°C for 30 min. The reaction was terminated by the addition of 20 μl of trichloroacetic acid (240 mg/ml) and centrifuged at 13,000 rpm for 3 min. The mix was then split into 2 equal aliquots of 100μL each. The liquid from one aliquot was evaporated using a Speedvac concentrator (Genevac, Suffolk, UK) and re-dissolved in ^{1}H NMR solvent comprising of 4:1 of D_2O:CD_3OD containing 0.01% d_4-trimethylsilylpropanoic acid (d_4-TSP). The other aliquot was mixed with 4:1 of H_2O:CH_3OH and was stored at -20°C for further analysis as necessary. Quantitative ^{1}H-NMR was carried out against the d_4-TSP internal standard (0.01% w/v) and characteristic signals

from the anomeric hydrogen of the newly formed glucoside (isorhamnetin glucoside: δ5.0674, *J* = 7.6 Hz; salicylaldehyde glucoside: δ5.2539, *J* = 7.7 Hz; 2-ethylphenol: δ5.0538, *J* = 7.6 Hz; saligenin glucoside (salicin): δ5.1022, *J* = 7.5 Hz; salicyl-7-benzoate glucoside: δ5.1479, *J* = 7.4 Hz; benzylsalicylate glucoside: δ5.1287, *J* = 7.8 Hz; phenol glucoside: δ5.1066, *J* = 7.6 Hz; 2-hydroxycinnamic acid glucoside: δ5.0905, *J* = 7.9 Hz; catechol glucoside: δ5.0430, *J* = 7.5 Hz).

The kinetic parameters of the substrates were determined at pH 7.5 by varying the substrate concentration between 0.05 mM and 2 mM and samples were collected at specific intervals and then subjected to quantitative ^{1}H NMR spectroscopy as described above. For salicyl-7-benzoate, the amount of degradation to salicyl alcohol and benzoate, in the assay, was quantified and for each sample (timepoint) the concentration of the available intact substrate was adjusted (via subtraction of the degraded portion) for the kinetic calculations. Data was then analyzed by non-linear regression of hyperbolic fits in MyCurveFit (https://mycurvefit.com/) and K_m and V_{max} were calculated by fitting substrate concentration versus reaction rate into the Michealis-Menten equation $Y= V_{max}X(K_m + X)^{-1}$. The kinetic parameters computed are mean of at least 3 independent experiments.

RNA interference

To generate RNAi constructs, two different ~300bp gene fragments from Potri.016G014500 were amplified from hybrid poplar (INRA 717-1B4 clone) using Phusion™ Hot start High-Fidelity DNA Polymerase (New England Biolabs) according to the manufacturer's instructions using *att*B-coupled PCR primers (Table S1). Purified *att*B-PCR products were cloned into the pDONR™221 vector (Life Technologies) to generate an entry clone via a BP recombination reaction. To generate an expression clone, gene fragments flanked by *att*L1 and *att*L2 borders were transferred via a LR clonase reaction to the Gateway® destination vector pK7GWIWG2(II) (Karimi *et al*., 2007) for expression of a self-complementary hairpin to induce RNA interference (RNAi). Purified plasmids were sequenced in order to check for

errors and to confirm correct assembly. The two constructs generated, pK7GWIWG2(II)-16G014500-RNAi-1 and pK7GWIWG2(II)-16G014500-RNAi-2, were transformed into *Agrobacterium tumefaciens* using the freeze and thaw method (Höfgen and Willmitzer, 1988).

Both constructs were transformed separately into hybrid poplar (*Populus tremula* × *P. alba*, clone INRA 717 1-B4) by *Agrobacterium*-mediated transformation (strain GV3101) as previously described by Gallardo *et al*., (1999) but with the following modifications to the published protocol: transformed calli were selected on callus induction medium (CIM) supplemented with 1 $mg \cdot L^{-1}$ 2,4-dichlorophenoxyacetic acid (2-4D) and 0.02 $mg \cdot L^{-1}$ thidiazuron (TDZ) and the appropriate antibiotics (50 $mg \cdot L^{-1}$ kanamycin, 200 $mg \cdot L^{-1}$ timentine, and 200 $mg \cdot L^{-1}$ cefotaxime). Shoots were regenerated from calli by transfer to shoot induction medium (SIM) containing 0.05 $mg \cdot L^{-1}$ 1- naphthaleneacetic acid (NAA) and 0.004 $mg \cdot L^{-1}$ TDZ. Regenerated shoots were transferred to 717 1-B4 stock media; half strength macro-nutrients Murashige and Skoog medium (1962) (MS Mod 1B, Duchefa Biochemie B.V), 20 $g \cdot L^{-1}$ sucrose, 0.5 $mg \cdot L^{-1}$ indol-3-acetic Acid (IAA) supplemented with 50 $mg \cdot L^{-1}$ kanamycin. *In vitro* cultures were maintained in a controlled growth room at 24 ± 1°C and low light conditions; 35 $\mu mol\ m^{-2}\ s^{-1}$ intensity in long-day (16 h light/8 h dark) photoperiod.

To confirm presence of the transgene, leaf tissue from *in vitro* propagated INRA 717 1-B4 shoots on individual transgenic events was collected for genomic DNA extraction using the DNeasy Plant Mini Kit (Qiagen). Concentration and quality was determined using a Nanodrop ND1000 spectrophotometer, with 5-15 ng of genomic DNA used in PCR to verify the presence of the *nptII* marker and hairpin RNAi fragments using PCR Master Mix (2X) (Thermofisher). Primer sequences are provided in Supplementary Table S1. The following cycling conditions were used: 94°C for 1 min, 35 cycles of 94°C for 40 s, 58°C for 1 min, and 72°C for 1 min, followed by a 10 min final extension step at 72°C. Amplification was verified by agarose gel electrophoresis and transformants selected for further characterisation.

Quantitative RT-PCR analysis

Samples comprising both leaf and stem tissue from the 1st to 3rd leaf from the shoot apical meristem were collected from 6-weeks-old plants grown *in vitro*, directly frozen in liquid nitrogen and ground to powder. RNA extraction was performed as described by Chang *et al*., (1993). Samples were treated with Ambion® DNA-*free*™ DNase kit and total cDNA was synthesised using a High Capacity cDNA Reverse Transcription Kit (Life Technologies) according to the manufacturer's instructions.

Primers for qRT-PCR were designed to obtain 75–150 bp products (Table S1). Real-time quantitative PCR was performed using 7500 Fast Real-Time PCR system (Applied Biosystems) using Fast SYBR® Green Master Mix (Applied Biosystems) according to the manufacturer's protocol. PCR cycling conditions were: 20 s at 95 °C; 40 cycles of 3 s 95 °C, 30 s 60 followed by a melt curve: 15 s at 95 °C, 1 min at 60 °C, 0.11 °C/s increase to 95 °C; 15 s 60 °C. Relative transcript levels were calculated by the comparative threshold cycle method. The poplar *TIP4*-like gene (Potri.009G093200) and *CDC2* (Potri.004G133500) were used as reference genes (Pettengill *et al*., 2012).

For qRT-PCR data analysis, the efficiencies and the threshold cycle (*Ct*) for all reactions per gene were estimated using the LinRegPCR package (Ramakers *et al.*, 2003) and between-run variation was removed by factor correction (Ruijter *et al.*, 2015). Normalised quantity (NRQ) data for the targets were calculated and a $\log_2(1/NRQ)$ transformation was used for statistical analysis according to Rieu and Powers (2009).

Metabolomic analysis of RNAi plants

RNAi lines were harvested at 6 weeks old. Intact whole plantlets were cut with scissors 1cm from the base of stem above the growth media and immediately immersed into 50 mL polypropylene centrifuge tubes (Grenier Bio-one Ltd, UK) pre-filled with liquid nitrogen. For

the constructs, plantlets within a magenta (typically 2-3) were pooled to form a biological sample. For control plants, generally shoots were regenerated *in vitro* from untransformed 717 1-B4 explants following the same process employed in the transgenics but without kanamycin selection. Plants were then micro-propagated and grown in magenta vessels in the same conditions as the transgenics. Individual six-week-old control plants grown *in vitro* in parallel with transgenics were used for metabolomics analysis. Empty vector transformed plants were also generated for metabolomic comparison to control plants. Plant tissue was milled in a mortar and pestle using liquid nitrogen. Tissues were stored at -80 °C until analysis. A minimum of 3 biological samples were generated per transgenic line for metabolomic analysis.

i) LC-MS metabolomic analysis of RNAi lines

Aliquots of pre-milled fresh-frozen tissue (50 mg) were weighed into a 2mL Eppendorf tube. Extraction was carried out using a mixture of water:methanol (4:1, 1mL) at 50 °C according to the procedures previously described in Ward *et al*. (2003). UHPLC-MS spectra were collected on an LTQ-Orbitrap Elite mass spectrometer coupled to a Dionex Ultimate 3000 RS UHPLC system equipped with a photodiode array detector (Thermo Fisher Scientific, Germany). Chromatographic separation of a 10 µL injection aliquot was achieved at 35 °C using a Hypersil GOLD reversed-phase column (1.9 µm, 30 × 2.1 mm i.d. Thermo Fisher Scientific, Germany). A gradient comprising mixtures of water and acetonitrile (0% to 75%), each containing 0.1 % formic acid was utilised and separation was carried out over a total run time of 40 minutes using a flow rate of 0.3 mL/min. Mass spectra were collected in negative ion mode and acquisition parameters have been described previously (Ward *et al*., 2020). MSMS spectra were recorded automatically using data dependant acquisition on the top 4 peaks. Higher-energy collisional dissociation (HCD) was used with a collision energy of 65 eV. Data was collected and inspected using Xcalibur v. 2.2 (Thermo Fisher Scientific, Germany). For statistical analysis data was processed in Compound Discoverer (v. 3.1., Thermo Fisher Scientific). The untargeted metabolomics workflow template was used to

align and process the data. Compounds were detected using a minimum intensity of 250,000. Identification was carried out by comparison to an in-house spectral library of known metabolites arising from data collection of authentic standards and isolated metabolites run under identical conditions. Annotation was carried out on the basis of comparison of accurate mass, predicted molecular formula, retention time and MSMS fragmentation patterns. Features arising from isotope and adduct peaks were removed from the data table. Known metabolites were extracted from the data table for further statistical analysis.

Hierarchical clustering analysis was carried out in Spotfire Analyst (v. 7.11.1, Tibco) using UPGMA clustering (unweighted pair group method with arithmetic mean) and the correlation distance measure. Data was mean-centred. Principal Component Analysis was carried out in Simca-P (v. 16.0.1, Sartorius Stedim Data Analytics AB).

ii) NMR analysis of RNAi plants

NMR spectra were acquired on a Bruker Avance Neo 600 MHz NMR spectrometer (Bruker Biospin, Germany), operating at 600.05 MHz for ^{1}H and 150.9 MHz for ^{13}C NMR spectra, using a 5mm TCI cryoprobe. For quantitation of key metabolites, NMR samples were prepared according to the methods described above, using a D_2O:CD_3OD (4:1, 1mL) extraction solvent containing an internal standard (d_4-3-(trimethylsilyl)propionic-2,2,3,3 acid, 0.01 % w/v). ^{1}H-NMR spectra were acquired at 300 °K using a Bruker water suppression pulse sequence (zgpr) with a 5s relaxation delay to suppress the residual water signal. Spectra consisted of 32,763 data points over a spectral width of 7142 Hz. For metabolomic data collection and subsequent quantitation of key metabolites, spectra consisted of data from 16 scans. For quantitation of substrate turnover and kinetic assays a total of 64 scans were collected. Spectra were automatically transformed using an exponential window and a line broadening of 0.5 Hz. Baseline correction, phasing and spectral referencing (to d_4-TSP) were carried out automatically by the instrument software (TOPSPIN v. 4.07, Bruker,

Germany). ^{1}H NMR spectra were automatically reduced in AMIX (v3.0, Bruker, Coventry, UK) to ASCII files with integrated regions of 0.01 ppm width. General metabolite quantitation was carried out using the known concentration of internal standard (d_4-TSP, 0.01% w/v) and characteristic chemical shift regions for each metabolite according to a library of authentic standards run under identical conditions. Statistical analysis of quantified metabolites was carried out by ANOVA in Spotfire v. 7.11.1 (TIBCO).

Results

Co-expression analysis leading to candidate UGT genes in Salix purpurea

Using the co-expression functionality in the Phytozyme 12.1 plant genomics database, mining of the available *Populus trichocarpa* data (version 3.0) against the expression profiles of the two acyltransferases, previously implicated in salicinoid biosynthesis (Chedgy *et al* 2015), identified several poplar UGTs with Pearson correlation coefficients, r > 0.85. Correlation against *PtACT47*, [PtSABT (benzoyl-CoA: salicyl alcohol O-benzoyltransferase); Potri.013G074500], revealed a single UGT - Potri.006G171100 (r = 0.883) whilst correlation against *PtACT49* [PtBEBT (benzoyl-CoA: benzyl alcohol O-benzoyltransferase); Potri.019G043600] revealed two further highly correlated UGTs – Potri.016G014500 (r = 0.922) and Potri.006G171200 (r = 0.917) in addition to Potri.006G171100 (r = 0.822) that also correlated to *PtACT47*. The latter two adjacent *PtUGT* genes on chromosome 6 are duplicates having 100% identity and have been previously designated as PtUGT78M1 isozymes (Veljanovski and Constabel, 2013; Yoshida *et al* 2015).

To further investigate the potential of the implicated poplar UGTs in the biosynthesis of salicinoids we next examined the *Salix purpurea* genome (ver. 1.0) for close protein homologues. With a threshold set at 80% similarity, of the three UGTs, only one, Potri.016G014500, has close homologues in the *Salix purpurea* genome. These are SapurV1A.0398s0160.1 (94.4% similarity) and SapurV1A.1999s0070.1 (94.2% similarity),

two genes that are apparently duplicates (99.79% identical to each other) residing on willow chromosomes 7 and 17 respectively. The other poplar candidates - Potri.006G171100 and the adjacent duplicate gene (Potri.006G171200), have no close homologues in the *S. purpurea* genome. The highest similarity (76.9%) for both of these genes was to a UGT (SapurV1A.0271s0260.1) residing on willow chromosome 18. However, this willow UGT is the orthologue of Potri.018G096000.1 (93.3% similarity), a gene that did not emerge from the acyltransferase coexpression analysis.

Thus, although the correlative expression analysis against the published poplar acyltransferases suggest that PtUGT78M1 may be a strong candidate as it correlated with *both* PtSABT and PtBEBT, the lack of close orthologues in *S. purpurea*, which shares the salicortin pathway, appears to rule out this enzyme. As the UGT encoded by Potri.016G014500 had very close homologues in *Salix.* (SapurV1A.0398s0160 and SapurV1A.1999s0070) these were selected from willow for functional analysis and were named SpUGT71L2 and SpUGT71L3 respectively through the UGT nomenclature committee (MacKenzie *et al*, 1997). The amino-acid sequences deduced from the sequences of these genes are shown in Figure S1 aligned with that of the poplar homologue. These proteins possessed the characteristic 44-amino-acid conserved plant secondary product glycosyltransferase (PSPG) consensus motif in their C-terminal region as shown (Figure S1).

Cloning and bacterial expression of UGTs

SpUGT71L2 and SpUGT71L3 were cloned from *Salix purpurea* 'Uralensis' (NWC844) and are single exon genes of 1443 bp length encoding proteins of 480 aa. The coding sequence of each enzyme was cloned and expressed in *E. coli* with an N-terminal 6xHis tag for affinity purification. Induction of target proteins of 55.1 kDa were confirmed by SDS PAGE (Figure S2) and the enzymes were purified by a 2-step metal affinity chromatography. About 4 mg of

pure protein was obtained from a 1L culture and aliquots were flash frozen. We performed a preliminary screen for potential substrates using UDP-glucose as the sugar donor. For both SpUGT71L2 and SpUGT71L3 salicyl alcohol, catechol, salicylaldehyde, 2-hydroxycinnamic acid and salicyl-7-benzoate (see later), were converted to their corresponding glucosides, identified by ^{1}H-NMR and LC-MS (Table S2), confirming, as expected, that the near-identical enzymes have the same functionality. The fairly efficient conversion of catechol to its glucoside was particularly interesting since Fellenberg *et al* (2020) reported that catechol was not a substrate for the poplar homologue, PtUGT71L1. The product, catechol glucoside, was confirmed by LCMS (Figure S3) and generated a peak at 13.23 min and an $[M-H]^-$ ion at *m/z* 271.08191, corresponding to a formula of $C_{12}H_{15}O_7$. MSMS showed a fragment at *m/z* 109.02982 corresponding to loss of the glucoside moiety.

Salicyl-7-benzoate is the preferred substrate for UGT71L2

To further analyse SpUGT71Ls, a full characterisation and enzyme kinetic analysis was carried out on SpUGT71L2 only, using ^{1}H-NMR to quantify substrate and product levels. A total of 22 different phenolic substrates were tested (Table 1). The majority of substrates were commercially available but salicyl-7-benzoate, was synthesised. This was achieved by coupling of salicyl alcohol to benzoyl chloride, and chromatographic separation from its co-products, the regio-isomer salicyl-2-benzoate (Poisson *et al*., 2007) and salicyl-2,7-dibenzoate (Figure 2). Characterisation of salicyl-7-benzoate by ^{1}H-NMR agreed with previously reported literature values (Belyanin *et al*., 2012). Attempts to synthesise salicyl-7-salicylate led to an identifiable (by NMR) crude product, but increased instability relative to the benzoate prevented purification for use in enzyme studies.

The extended substrate survey indicated that SpUGT71L2 preferred to glucosylate phenolic substrates with substituents at the *ortho* position relative to the reacting hydroxyl group. The enzyme was tolerant of a number of *ortho*-substituents e.g. alkyl (2-ethylphenol), hydroxyl

(catechol), hydroxymethyl (salicyl alcohol), aldehyde (salicylaldehyde) and ester (benzylsalicylate) but had a strong preference for salicyl-7-benzoate. For salicyl alcohol (saligenin) the glycosylation was specific to the phenolic hydroxyl group (C-2). Notably, no turnover was observed for salicylic acid, 2-hydroxyphenylacetic acid and homogentisic acid, although the longer 2-hydroxycinnamic acid was a substrate. Phenols with alternate substitution patterns, e.g. 3-hydroxy cinnamic acid were not substrates, nor were cyclohexanols, or flavonoids as represented by quercetin, luteolin and isorhamnetin.

Prior to full kinetic analysis of the enzyme with the preferred substrate, salicyl-7-benzoate, it became necessary to consider the stability of this ester in the assay conditions. This compound was found to be stable in aprotic solvents, but we observed decomposition to salicyl alcohol and benzoic acid in aqueous solvents, particularly at higher pH. The glucoside of salicyl-7-benzoate was stable, as the anchimeric assistance of the free phenolate on the hydrolysis of the ester in salicyl-7-benzoate was now absent. However, use of a ^{1}H-NMR based enzyme assay allowed accurate quantitation of substrate, glucoside product and substrate decomposition products. We were able to assess the functionality of SpUGT71L2 over a pH range of 3.6-10 (Figure 3). Although more stable at lower pHs, salicyl-7-benzoate turnover was much lower in acidic conditions (Figure 3). After some experimentation, using ^{1}H-NMR to follow enzyme activity and spontaneous substrate hydrolysis, the optimum buffer and working pH range, was determined to be between 7.4 and 7.9. Amino acid sequence analysis of SpUGT71L2 in publicly available servers revealed no signal peptide for subcellular localization or transmembrane helices for membrane anchoring, thereby indicating that this protein was likely located in the cytosol *in planta*. Thus, in order to mimic the plant physiological state, we subsequently carried out all the enzyme assays at pH 7.5, adjusting the substrate concentration data to account for substrate losses at each time-point during the assay.

We performed full kinetic analysis on salicyl-7-benzoate and three alternate naturally occurring substrates and the Michaelis-Menten kinetic values are summarised in Table 2.

The catalytic efficiency of SpUGT71L2 was highest for salicyl-7-benzoate at 24.84 $mM^{-1}s^{-1}$, which was 7.5 fold higher than salicyl alcohol (3.33 $mM^{-1}s^{-1}$), 2 fold higher than catechol (11.99 $mM^{-1}s^{-1}$) and 5.5 fold higher than 2-hydroxycinnamic acid (4.54 $mM^{-1}s^{-1}$). Moreover, UGT71L2 had an approximately 5 fold higher k_{cat} values for salicyl-7-benzoate in comparison to the other substrates under study. SpUGT71L2 had a stronger binding affinity for catechol with a K_m of 0.4 mM, which was 3 fold lower than that of salicyl-7-benzoate (1.22 mM). A higher k_{cat} coupled with a higher catalytic efficiency indicates that UGT71L2 is more efficient in glycosylating salicyl-7-benzoate into its corresponding glucoside. Catechol is naturally occurring in *Salix sp.*, possibly as a product of decomposition of salicortin and its analogues (Ruuhola *et al* 2003) and therefore *in planta*, it has the potential to be an effective competitive inhibitor of UGT71L2 catalysed formation of salicyl-7-benzoate glucoside.

Salicyl-7-Benzoate glucoside is present in poplar and willow

As a prelude to generating knockdowns of UGT71L1 to test the role of this enzyme *in planta*, we examined the secondary metabolites of *P. trichocarpa* 'ManMillan' for the presence of salicyl-7-benzoate glucoside. Inspection of LC-MS data from leaf extracts of *P. trichocarpa* for metabolites with a molecular weight of 390 indicated two small peaks at 28.3 min and 28.67 min (Figure S4). MS spectra in both cases showed an ion relating to the formate adduct of compounds with molecular formula $C_{20}H_{22}O_8$. MS-MS of *m/z* 435.1295 for the peak appearing at 28.3 min generated fragments at *m/z* 227.0720 ($C_{14}H_{11}O_3$) and *m/z* 137.0250 ($C_7H_5O_3$). In contrast, MS-MS of *m/z* 435.1293 for the peak appearing at 28.67 min generated a fragment at *m/z* 121.0301 ($C_7H_5O_2$). Both peaks were isolated *via* repeated injection into an HPLC system and were fully characterised by 1D and 2D NMR. The peak appearing at 28.30 min corresponded to benzylsalicylate glucoside (Figure 2), a compound previously isolated from *Sarcandra glabra,* an evergreen shrub found in southern China (Wu *et al*., 2012). The peak at 28.67 min corresponded to salicyl-7-benzoate glucoside (Figure 2),

more commonly called desoxysalireposide (Belyanin et al., 2012), and this represents the first report of this compound in plants. The identification was confirmed by comparison to an authentic synthetic standard (Belyanin et al., 2012). The ratio of benzylsalicylate glucoside to salicyl-7-benzoate glucoside was found to be 2:1 in tissue from the growing tip and also in young leaves from side branches of *P. trichocarpa*. Both compounds were also detected in juvenile tissue of the hybrid *P.alba* × *P.tremula* clone 'INRA 717 1-B4', a line widely established for transformation, where the ratio was 1:1 in young leaves and 3.5:1 in the growing tip. Salicyl-7-benzoate glucoside could also be detected, at trace levels, in LC-MS data from aqueous methanol extracts of willow.

RNAi knockdown of UGT71L1 in transgenic poplar

To probe directly the role of UGT71L genes *in vivo*, we designed two RNAi constructs for PtUGT71L1 (Figure 4), each targeting a different region of the coding sequence, and generated stable transgenic lines in the *P.alba* × *P.tremula* 'INRA 717 1-B4' background. We confirmed the presence of the transgene in 24 independent lines through PCR amplification of the *nptII* marker and hairpin RNAi fragments. Six-week old transgenic plantlets grown *in vitro,* were analysed for PtUGT71L1 gene expression by qPCR (Figure 5). Subsequently the effects of decreased PtUGT71L1 on salicinoid and other key secondary metabolite composition were analysed by LC-MS based metabolite profiling of the same plantlets. Key discriminatory metabolites, and major salicinoids were also quantified in ^{1}H-NMR metabolomic profiles. Generally, control plantlets were generated by tissue culture of the parental line (INRA 717 1-B4) and were grown in parallel to the transgenic lines. Empty vector transformed plantlets were also generated and showed no appreciable difference to these untransformed control plantlets.

10 RNAi lines showing much decreased expression of UGT71L1 (16G-1-5, 16G-1-10, 16G-1-17, 16G-1-18, 16G-1-21, 16G-2-4, 16G-2-10, 16G-2-22, 16G-2-32, 16G-2-34), comprising

5 different insertion events from each construct, were selected and profiled by LC-MS and compared to data from wild type (INRA 717 1-B4) plantlets grown at the same time and harvested at the same timepoint. Extracts were made in aqueous methanol and samples were analysed in ESI negative ion mode following separation using reversed phase chromatography. A representative total ion chromatogram of one line (16G-1-5) vs wild type (INRA 717 1-B4) is shown in Figure S5. Tremulacin, was the major salicinoid present in INRA 717 1-B4 plantlets which also contained significant levels of 6'-HCH-tremulacin and salicortin. Initial inspection of the data from the RNAi lines indicated a significant reduction in levels of all these major salicinoids. Metabolite profiles of INRA 717 1-B4 poplar plantlets were shown to be qualitatively the same as larger greenhouse plants grown on soil (Figure S6). Furthermore, selected RNAi lines were taken through to greenhouse plants and the significant reduction in major salicinoids was carried through in full size plants (representative data of 16G-1-5, shown in Figure S6).

For in depth analysis, the LC-MS data of the RNAi plantlets was further processed with Compound Discoverer™ software against an in-house library of, NMR authenticated, phenolic glycoside ESI-MS spectra and retention times, to generate an aligned table of "compound annotated" data. Known metabolites were analysed using hierarchical clustering analysis (Figure 6). All RNAi knockdown lines clustered separately from INRA 717 1-B4 wild type samples, whilst within the RNAi knockdown cluster, segregation was observed between lines arising from the two different constructs, with lines arising from construct 2 displaying a greater effect. Metabolites that were reduced in 16G-1 and 16G-2 lines included 6'-HCH-tremulacin, 6'-HCH-salicin, 6'-HCH-tremuloidin, salicortin, tremulacin, salicin, catechol, glucogallin, feruloyl quinate, benzoyl glucuronide and isomers of grandidentatin, coumaroyl salicin and chlorogenic acid. Populoside and its isomer were also reduced in the majority of lines. Detectable reductions in salicyl-7-benzoate glycoside (and the isomer benzylsalicylate glycoside) in the transgenic lines were recorded (Figure 6), even though the levels in the controls are low relative to other salicinoids. It is of note that salicyl-7-benzoate could not be

detected in either control or RNAi lines. Similarly, catechol glycoside was detected at low levels, in both control and RNAi lines, but appeared unchanged by the transgenesis.

Additional metabolite changes, possibly as a consequence of off-target effects, were able to discriminate between lines generated from construct 1 and those generated from construct 2. 16G-2 lines were found to contain increased levels (relative to INRA 717 1-B4 and 16G-1 lines) of tremuloidin, phenylethyl diglycosides, grandidentin and an isomer of cinnamoylsalicin. The same 16G-2 lines also contained lower levels (relative to INRA 717 1-B4 and 16G-1 lines) of salireposide, 6'-HCH-salicortin, benzyl-HCC-glucoside (Evans *et al*., 2004), salicin-7-sulfate (Noleto-Dias *et al*., 2018), diosmetin glucoside, chlorogenic acid, coumaroylquinate, and coumaroylsalicin. Average peak areas (Table S3) for 16G-1, 16G-2 and INRA 717 1-B4 samples were generated from the LC-MS data table and were in broad agreement with the hierarchical clustering analysis. The quantified LC-MS data table was also analysed by Principal Component Analysis (Figure S7). The analysis confirmed the separation between RNAi lines and control and also indicated differences arising from the use of different constructs. The associated loadings plot confirmed results observed via the hierarchical clustering analysis. Identical extracts were also analysed by ^{1}H-NMR, a quantitatively more robust technique compared to LC-MS, and key discriminatory salicinoids for each RNAi knockout line were quantified (Figure 7 and Table S4) relative to a known concentration of internal standard (0.01% w/v d_4-TSP). Calculated metabolite concentrations again indicated a substantial reduction in major salicinoid levels (tremulacin 18-40% of WT, salicortin 28-54% of WT and salicin (30-55% of WT) across the ten selected 16G-1 and 16G-2 lines.

6'-HCH-substituted salicinoids were also prevalent in INRA 717 1-B4 wild-type and were found to be reduced in the transgenic lines (Figure 7). Although there have been previous reports of the characterisation of 6'-HCH-salicortin (Feistel *et al*., 2015) and 6'-HCH-tremulacin (Keefover-Ring *et al*., 2014) we report here for the first time the identification of 6'-HCH-salicin (Figure 2), an isomer of salicortin, and 6'-HCH-tremuloidin, an isomer of

tremulacin. Both of these compounds are slightly more polar than their respective isomers and clear peaks could be observed for them in the LC-MS chromatogram of INRA 717 1-B4 poplar (Figure S5). 6'-HCH-salicin eluted at 20.7 min and its MS spectrum showed an ion at m/z 423.1290 corresponding to a compound with molecular formula $C_{20}H_{24}O_{10}$. The peak was isolated by preparative HPLC and characterised by ^{1}H and ^{13}C-NMR. The arrangement of the functional groups was confirmed by ^{1}H-^{13}C HMBC analysis which showed correlations from the ester carbonyl carbon (δ173.2) to the H_2-6' moiety of the glucose (δ4.62 and 4.61). In addition, a correlation between the H-1' anomeric hydrogen of the glucose (δ5.12) and C-1 of the salicyl moiety (δ157.4) further confirmed that the glucose moiety was the central part of the structure. 6'-HCH-tremuloidin eluted at 29.48 min and showed an ion at m/z 527.1544 corresponding to a molecule with molecular formula $C_{27}H_{28}O_{11}$. Again, the MS-MS fragmentation was identical to its more abundant isomer tremulacin. This molecule has been putatively assigned to 6'-HCH-tremuloidin on the basis of its MS behaviour, its polarity and proximity to its isomeric peak in line with the observed properties of 6'-HCH-salicin vis-à-vis salicortin. It was not quantified in the NMR analysis as levels were too low for successful isolation and authentication by NMR. However, three 6′-HCH salicinoids were quantified in the transgenic lines and were found to be reduced by similar amounts (Figure 7 and Table S4) to the more well-known compounds.

Discussion

The combined bioinformatic analysis of *P.trichocarpa* and *S. purpurea* genomics resources for genes co-expressed with the published poplar acyltransferases (Chedgy *et al*, 2015) clearly identified UGT71L genes as likely candidates for the biosynthesis of the glucosyl moiety in the common salicinoids. The *E. coli* produced willow isozymes were active against

a number of *ortho*-substituted phenols but showed a clear preference for salicyl-7-benzoate. Thus, in agreement with Fellenberg *et al* (2020), UGT71L1/2/3 emerged as strong candidates for participation in the proposed plant pathways from benzyl- or salicylbenzoate through to salicortin/tremulacin (Babst *et a*l 2010). The substantial knockdown of tremulacin and related compounds in the transgenic RNAi poplar hybrid reinforces the notion that UGT71L1 is a key player in this pathway. The product of UGT71L2 enzyme, salicyl-7-benzoate glucoside, was demonstrated to be present (as a minor salicinoid) in rapidly growing tissue of poplar. The co-occurrence of its isomer, benzylsalicylate glucoside, is also of interest as the *in vitro* studies of the willow enzyme indicated that benzylsalicylate was also a substrate for SpUGT71L2, albeit with lower affinity. The results do not rule out the possibility that other endogenous substrates, with more highly functionalised lower rings, are involved. However, the stability of such *ortho*-hydroxybenzoates is not expected to be high as demonstrated by salicyl-7-salicylate which was synthesised as a candidate, but was too unstable to isolate or test with the *E.coli* produced enzyme.

In general, our *in vitro* results with the willow enzyme agree with those reported from the poplar enzyme (Fellenberg *et al* 2020). However, we observed that catechol was also a good substrate *in vitro*, and thus as catechol occurs in willow extracts, a role for SpUGT71L2s in metabolism of this compound *in vivo* seemed possible.

RNAi knockdown, in poplar, of the UGT would be expected to cause a build-up of substrates and a reduction in glycosylated salicinoids. Salicyl-7-benzoate could not be detected in either RNAi or control lines. The instability of salicyl-7-benzoate at physiological pH, may be an explanation of why build-up of substrate was not observed, especially in the RNAi knockdown lines. It would also indicate that production of salicyl-7-benzoate via acyltransferases and processing by PtUGT71L1 is tightly coupled *in vivo*. Similarly, for catechol no build-up of aglycone was observed in the RNAi lines. Instead a reduction in catechol was observed in the knockdowns.

This result, coupled with the observation of only trace amounts of catechol glycoside in both control and RNAi poplar lines, indicates that glycosylation of catechol by PtUGT71L1 was probably not a major process *in vivo* and suggests that the decrease in catechol observed in poplar RNAi knockdowns was more likely related to lower levels of salicortin, tremulacin and 6′-HCH-salinoids which can give rise to this metabolite *via* hydrolytic fission of HCH groups (Ruuhola *et al* 2003) that most probably occurs on extraction and sample handling.

Our results with poplar PtUGT71L1 RNAi knockdown lines are in agreement with those of Fellenberg *et al* (2020) obtained using gene-editing, where complete silencing of salicortin and tremulacin formation was observed in transgenic hairy roots. It remains to be seen if gene-deletion of PtUGT71L1 also completely eliminates salicortin and tremulacin from leaf and stem tissue of poplar plants. Because RNAi did not result in complete knockout of tremulacin and analogues, we cannot rule out a contribution from other UGTs in the production of salicinoids. However, indications are that PtUGT78M1, a candidate enzyme (Veljanovski and Constabel, 2013; Yoshida *et al* 2015) that shows similar substrate specificities to PtUGT71L1 *in vitro* (Fellenberg *et al* 2020) is not involved *in planta*, as a UGT78M1 homologue is not present in willow. Gene editing of UGT78M1 in poplar hairy roots was not successful and thus its function has not been determined *in planta* (Fellenberg *et al* 2020). UGT71L1/2/3 are the first genes/enzymes of the salicinoid pathway to be functionally characterised both *in vitro* and *in planta*. They can be placed early in a putative metabolic grid as shown in Figure 8. The onward steps leading from the UGT71L product, salicyl-7-benzoate glucoside, to the main salicinoids (salicortin and tremulacin) in poplar hybrid INRA 717 1-B4 are intriguing. The indicated oxidation of benzoate to HCH may involve chemically reactive transient intermediates formed by dearomatisation of pendant benzoate groups. A further unknown aspect in this metabolic grid in poplar is the provenance of the 6′-HCH derivatives. These may arise from a similar oxidative dearomatisation of glucosyl-6′-benzoate substituents, although no evidence for the presence of metabolites such as 6′-benzoyltremulacin has been found (Keefover-Ring *et al* 2014). However, the

possibility that HCH groups are added to the salicyl 7- and glucose 6′ positions via acyl transfer is depicted in Figure 8. It should also be considered that HCH groups may migrate inter- and/or intra-molecularly *via* trans-esterification reactions. Furthermore, the contribution of non-enzyme catalysed reactions to the production of some salicinoids are also unknown. In particular, the hydrolysis of HCH groups, and any intermediates leading to them, can be responsible for a significant proportion of salicin (and tremuloidin) production. However, the direct synthesis of some salicin from salicyl alcohol, *via* UGT71L or other UGTs still remains a possibility. Our results do not define the proportion of salicin (and tremuloidin) that arise from the decomposition of salicortin (tremulacin) versus direct synthesis from saligenin, as there are inconsistences in the metabolomic data, particularly for tremuloidin which is of low abundance relative to salicin. At this stage it is also difficult to judge the relative contribution of the acyltransferases PtBEBT and PtSABT to the pathway (Figure 8) and the possibility of conversion of benzylbenzoate to salicyl-7-benzoate by hydroxylation remains.

The metabolomics study of the RNAi lines revealed significant changes in other phenylalanine-derived metabolites. Some of these e.g. populoside, salireposide and granditentatin and derivatives can be attributed to branches of the benzoate pathway, leading to salicinoids, and thus feedback perturbations in levels may be expected. Others e.g. chlorogenic acid, quinic acid, glucogallin and feruloylquinate, are much further removed from salicinoids, suggesting that 'off target' effects may result in a more generalised feedback inhibition at a number of points in the complex network of phenylalanine synthesis from shikimate and forward metabolism into phenylpropanoids including lignin. Indeed, it has already been shown (Morse *et al* 2007) that there is significant cross-talk between salicinoid and quinate pathways in poplar. In this work creation of a sink *via* overexpression of a bacterial salicylate hydroxylase did not significantly reduce salicinoids but resulted in significant repression of caffeoyl-, feruloyl- and coumaroyl-quinates. Thus, in poplar, where salicinoids are abundant, regulatory control must involve significant dynamic links from the benzoate branch into several parts of the shikimate-phenylalanine-lignin pathway.

Although many questions concerning the biosynthesis of the salicinoids are yet to be answered, in this work we have confirmed a key role for UGT71L genes in the pathway leading to salicortin and tremulacin in both willow and poplar, and a potential for salicyl-7-benzoate and its (more stable) glycoside to be intermediates leading to these major salicinoids. However, until the steps leading to the HCH ring have been determined, the exact placement of the UGT reaction in the sequence, and definition of any further endogenous substrates (e.g. salicyl-7-HCH) for this enzyme, cannot yet be finalised, as indicated in Figure 8.

Data availability

Salix purpurea UGT sequences are available from NCBI Genbank. Accession numbers SpUGT71L2: MT525101; SpUGT71L3: MT525102.

Acknowledgements

Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council (BBSRC) of the UK. The authors acknowledge Prof Fernando Gallardo and Dr Lara Jimenez-Bermudez, Molecular Biology and Biochemistry Department and Central Research Services (SCAI) from University of Malaga for kindly providing *Populus tremula* x *P. alba.* INRA clone 717 1-B4 plants and for technical advice, respectively. We also thank Elena Stepanova, National Research Tomsk Polytechnic University, for the generous gift of a standard of benzylsalicylate glucoside and Murray Grant, University of Warwick, for providing the pProEXHTa plasmid. The work described in this study was supported by successive Institute Strategic Programme grants: Cropping Carbon [BBS/E/C/00005199] and Tailoring Plant Metabolism [BB/E/C/000I0410], funded by the BBSRC.

Author Contributions

J.L.W. and M.H.B. were responsible for the conception and design of the study; S.K., S.C-M., C.H. F de J, A.R., S.H, C.L. and J.L.W. conducted the experiments; C.L., G.R. and J.L.W. undertook the spectroscopic data analysis and structure determination; S.K., J.L.W. and M.H.B prepared the manuscript.

Conflicts of interest

The authors declare no conflict of interest

References

Babst BA, Harding SA, Tsai C-J. 2010. Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in *Populus.* Journal of Chemical Ecology 36, 286-297.

Belyanin ML, Stepanova EV, Ogorodnikov VD. 2012. First total synthesis of natural acyl derivatives of some phenol glycosides of the family *Salicaceae*. Carbohydrate Research 363, 66-72.

Boatright J, Negre, F Chen X, *et al*. 2004. Understanding *in vivo* benzenoid metabolism in Petunia petal tissue. Plant Physiology 135, 1993-2011.

Boeckler GA, Gershenzon J, Unsicker SB. 2011. Phenolic glycosides of the Salicaceae and their role as anti-herbivore defences. Phytochemistry 72**,**1497-1509.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reports 11, 113–116.

Chedgy RJ, Kollner TG, Constabel CP. 2015. Functional characterisation of two acyltransferases from *Populus trichocarpa* capable of synthesising benzylbenzoate and salicylbenzoate, potential intermediates in salicinoid phenolic glycoside biosynthesis. Phytochemistry 113, 149-159.

Desborough MJR, Keeling DM. 2017. The aspirin story – from willow to wonder drug. British Journal of Haematology 177, 674-683.

Evans TP, Clausen TP, Reichardt PB, *et al.* 2004. Structurally intriguing glucosides from Alaskan Little tree willow (*Salix arbusculoides*). Journal of Natural Products 58, 1897-1900.

Feistel F, Paetz C, Lorenz S, *et al*. 2015. The absolute configuration of salicortin, HCH-salicortin and tremulacin from *Populus trichocarpa* × *deltoides* Beaupré. Molecules 20, 5566-5573.

Fellenberg, C, Corea, O, Yan, L-H, *et al*. 2020. Discovery of salicylbenzoate UDP-glycosyltansferase, a central enzyme in polar salicinoid phenolic glycoside biosynthesis. Plant Journal 102, 99-115

Gallardo F, Fu J, Canton FR, *et al*. 1999. Expression of a conifer glutamine synthetase gene in transgenic poplar. Planta 210, 19–26.

Höfgen R, Willmitzer L. 1988. Storage of competent cells for *Agrobacterium* transformation. Nucleic Acids Research 16, 9877.

Hsu F-L, Nonaka GI, Nishioka I. 1985. Tannins and related compounds. 32. Acylated flavanols and procyanidins from *Salix sieboldiana*. Phytochemistry 24, 2082-92.

Karimi M, Depicker A, Hilson P. 2007. Recombinational cloning with plant Gateway vectors. Plant Physiology 145,1144-1154.

Keefover-Ring K, Ahnlund M, Abreu IN. *et al*. 2014. No evidence of geographical structure of salicinoid chemotypes within *Populus tremula*. PLoS One **9**, e107189.

Lim E, Doucet CJ, Elias L, *et al*. 2002. The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxysalicylic acid and other benzoates. Journal of Biological Chemistry 277, 586-592.

Mackenzie PI, Owens IS, Burchell B, *et al*. 1997. The UDP-glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics 7, 255-69.

Morse AM, Tschaplinski TJ, Dervinis C, *et al*. 2007. Salicylate and catechol levels are maintained in nahG transgenic poplar. Phytochemistry 68, 2043-2052

Murashige T, Skoog F. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiolgia Plantarum 15, 473-497.

Noleto-Dias C, Ward JL, Bellisai A, *et al*. 2018. Salicin-7-sulfate: A new salicinoid from willow and implications for herbal medicine. Fitoterapia 127, 166-172.

Pettengill EA, Parmentier-Line C, Coleman GD. 2012. Evaluation of qPCR reference genes in two genotypes of *Populus* for use in photoperiod and low-temperature studies, BMC Research Notes **5**, 366.

Poisson T, Dalla V, Papamicaël C, *et al*. 2007. DMAP-organocatalyzed *O*-silyl-*O*-(or *C*-)-benzoyl interconversions by means of benzoyl fluoride. Synlett 381-386.

Ramakers C, Ruijter JM, Deprez RH, *et al*. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neuroscience Letters 339, 62–66.

Rieu I, Powers SJ. 2009. Real-time quantitative RT-PCR: design, calculations, and statistics. Plant Cell **21**, 1031–3.

Ruijter JM, Ruiz Villalba A, Hellemans J, *et al*. 2015. Removal of between-run variation in a multi-plate qPCR experiment. Biomolecular Detection and Quantification 5, 10–14.

Ruuhola T, Julkunen-Tiitto R, Vainiotalo P. 2003. *In vitro* degradation of willow salicylates. Journal of Chemical Ecology **29**, 1083-1097.

Tsai CJ, Guo W, Babst B, *et al*. 2011. Salicylate metabolism in *Populus*. *BMC Proceedings* **5** (Suppl 7):19.

Veljanovski V, Constabel CP. 2013. Molecular cloning and biochemical characterisation of two UDP-glycosyltransferases from poplar. Phytochemistry **91**, 148-57.

Ward JL, Harris C, Lewis J, *et al*. 2003. Assessment of 1H-NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. Phytochemistry ***62***, 949-957.

Ward JL, Wu Y, Harflett C, *et al*. 2020. Miyabeacin: A new cyclodimer presents a potential role for willow in cancer therapy. Nature Scientific Reports. doi.org/10.1038/s41598-020-63349-1.

Wu H, Xiaoru H, Xiaopo Z, *et al*. 2012. Benzyl 2-β-glucopyranosyloxybenzoate, a new phenolic acid glycoside from *Sarcandra glabra*. Molecules 17, 5212-5218.

Yoshida K, Ma DW, Constabel CP. 2015. The MYB182 protein down-regulates proanathocyanidin and anthocyanin biosynthesis in poplar by repressing both structural and regulatory flavonoid genes. Plant Physiology **167**, 693-710.

Zenk MH. 1967. Pathways of salicyl alcohol and salicin formation in *Salix purpurea* L. Phytochemistry 6**,** 245-252.

Table 1. Substrate specificity of SpUGT71L2

Substrate	% turnover[a]
Salicyl-7-benzoate	77 ± 9
Salicyl-2-benzoate	0
Salicyl alcohol (2-hydroxybenzyl alcohol; saligenin)	11 ± 6
Benzyl alcohol	0
3-hydroxybenzyl alcohol	0
4-Hydroxybenzyl alcohol	0
Cinnamyl alcohol	0
Salicylaldehyde	17 ± 10
Salicylic acid	0
Benzyl salicylate	5 ± 0.1
Phenol	5 ± 1.5
Catechol	26 ± 9
2-Ethylphenol	19 ± 7
2-Hydroxycinnamic acid	15 ± 1
3-Hydroxycinnamic acid	0
2-Hydroxyphenyl acetic acid	0
Homogentisic acid	0
cis/trans 1,2-Dihydroxycyclohexane	0
2-Hydroxycylcohexane carboxylic acid	0
Quercetin	0
Luteolin	0
Isorhamnetin	0

[a] UGT71L2 assay consisted of 5 μg of recombinant protein, 50 mM Na phosphate pH 7.5, 1 mM ß-mercaptoethanol, 5 mM UDP-glucose and 1 mM phenolic substrate. The reaction mix was incubated at 30°C for 30 min. Turnover was quantified by ^{1}H-NMR relative to a known concentration of internal standard (d_4-TSP, 0.01% w/v). Typically, three replicate experiments were performed, and values represent mean ± standard deviation.

Table 2. Kinetic data[a] for SpUGT71L2

Substrate	V_{max} (mM min^{-1})	K_m (mM)	Turnover number kcat (s^{-1})	Catalytic efficiency k_{cat}/K_m ($mM^{-1}s^{-1}$)
Saligenin	0.298±0.02	1.91±0.49	5.16±0.08	3.33±0.46
Salicyl-7-benzoate	0.15±0.04	1.22±0.61	27.74±7.99	24.84±6.55
Catechol	0.06±0.01	0.40±0.01	4.82±0.74	11.99±2.24
2-hydroxy cinnamic acid	0.144±0.03	1.35±0.24	5.32±1.17	4.54±2.02

[a] Kinetics experiments were conducted at pH 7.5 with substrate concentrations between 0.05 mM and 2 mM. Reaction products were quantified by ^{1}H NMR spectroscopy. Data was analyzed by non-linear regression of hyperbolic fits and K_m and V_{max} were calculated by fitting substrate concentration versus reaction rate into the Michealis-Menten equation $Y= V_{max}X(K_m + X)^{-1}$. The kinetic parameters computed are mean ± standard deviation of at least 3 independent experiments.

Figure Legends

Figure 1. Structures of key metabolites

Figure 2. Structures of newly synthesised and isolated metabolites. Endogenous metabolites are indicated in bold type.

Figure 3. pH profile of SpUGT71L2 activity against salicyl-7-benzoate. The optimal pH was determined by assay of the conversion of salicyl-7-benzoate to the glucoside using different buffering systems to cover the range, as follows:- citrate buffer (pH 3.6-5.1); sodium phosphate (pH 5.6-7.9) and glycine-sodium hydroxide buffer (pH 8.6-10.6). Data represents mean values from 3 independent experiments. Error bars are standard error of the mean.

Figure 4. RNAi construct design. (a) Map of pK7GWIWG2(II)-16G014500-RNAi construct. LB: left T-DNA border; RB: right T-DNA border; p35S: CaMV35S promoter; attB1 and attB2: recombination sites for LB reaction; 16G014500 sense/16G014500 antisense: segment of the coding sequence to produce self-complementary hairpin; CmR: chloramphenicol resistance; T35S: 35S terminator; pNOS: nopaline synthase promoter; *nptII*: kanamycin resistant marker; tNOS: nopaline synthase terminator. (b) Schematic representation of poplar *16G014500* gene. 16G014500-RNAi-1 / 2: segments of the coding sequence cloned into pK7GWIWG2(II) vector to produce constructs pK7GWIWG2(II)-16G014500-RNAi 1 and 2 used in this study. Black arrows indicate primers used for PCR screening (A) and cloning (B).

Figure 5. UGT71L1 gene expression in plantlets of RNAi knockdown lines. To test significance one-way ANOVA was used followed by Dunnett's post-test for multiple comparisons against the WT INRA 717 1-B4 control; (*) $P < 0.05$ (**) $P < 0.01$(***) $P < 0.001$ and n.s., not significant ($P \geq 0.05$). Graphs represents NRQ averages ± SEM from three biological replicates.

Figure 6. Hierarchical clustering analysis of LC-MS metabolomics data (negative ion mode), following extraction (4:1 H_2O:CH_3OH) of poplar tissue from wild type (4 biological replicates) and transgenic lines (5 each of 16G-1 and 16G-2, data shown is average of 3 biological replicates. Analysis was carried out in Spotfire Analyst using UPGMA clustering (unweighted pair group method with arithmetic mean) and the correlation distance measure. Data was mean-centred. Colour scale: red: max value; blue: min value.

Figure 7. Quantified salicinoids (mg/g fresh weight) from ^{1}H-NMR analysis of transgenic and wild type poplar following extraction in 4:1 D_2O:CD_3OD. To test significance one-way ANOVA was used followed by Dunnett's post-test for multiple comparisons against the WT INRA 717 1-B4 control; (*) $P < 0.05$ (**) $P < 0.01$(***) $P < 0.001$ and n.s., not significant ($P \geq 0.05$). Graphs represent average ± standard deviation from 3 biological replicates.

Figure 8. Proposed biosynthetic network relating to UGT71L1. Solid arrows relate to confirmed transformations. Dashed arrows related to postulated reactions suggested by metabolomic analysis of RNAi knockdown lines. Compounds depicted as bold type represent the major focus of discussion.

Figure 1

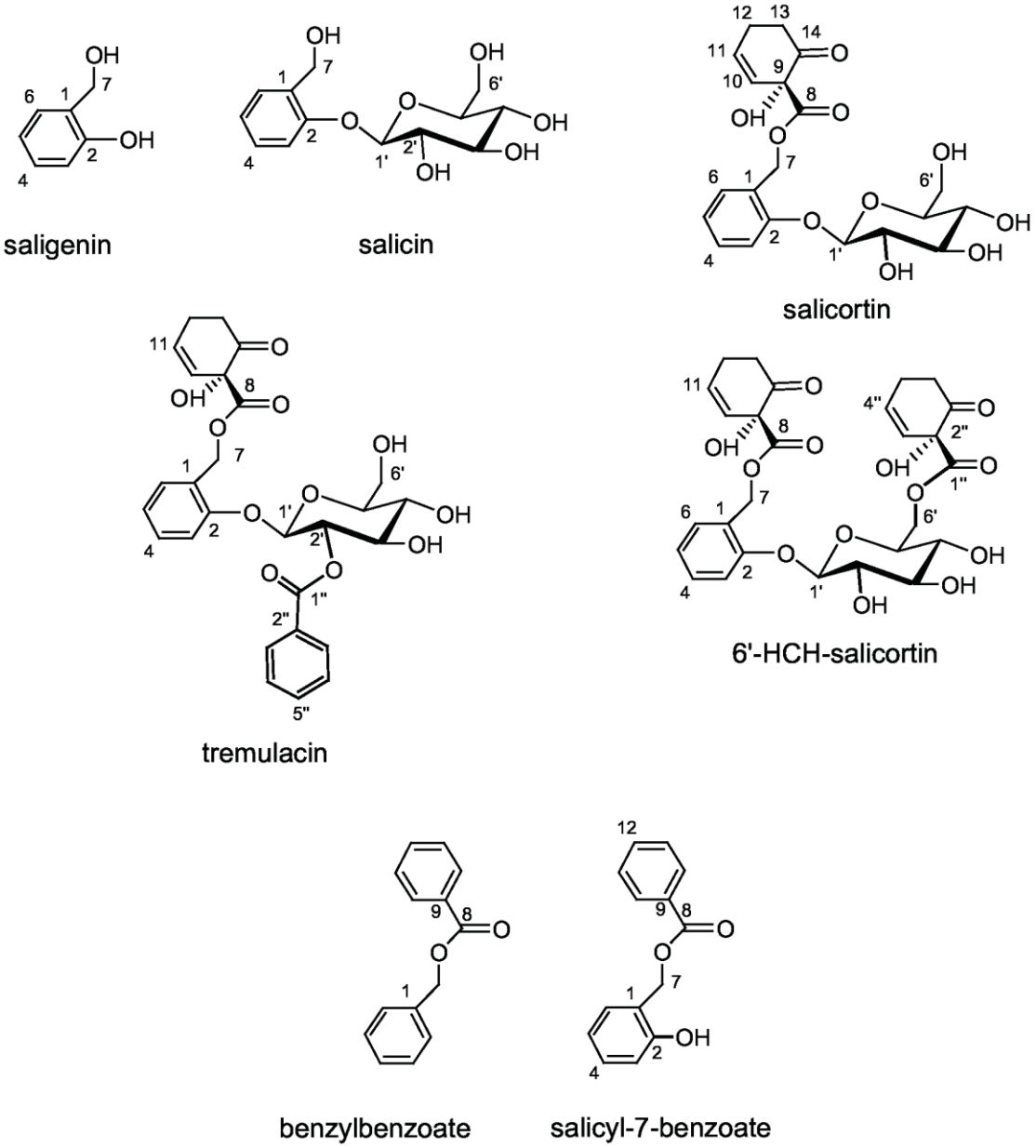
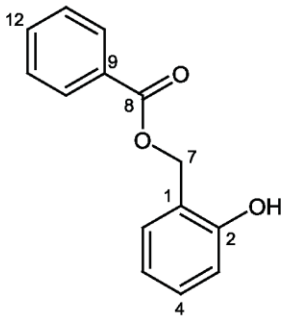
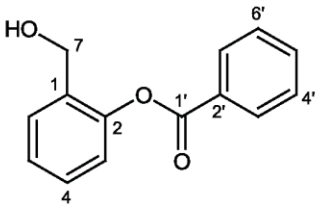


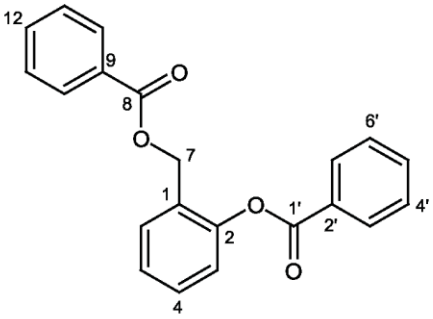
Figure 2



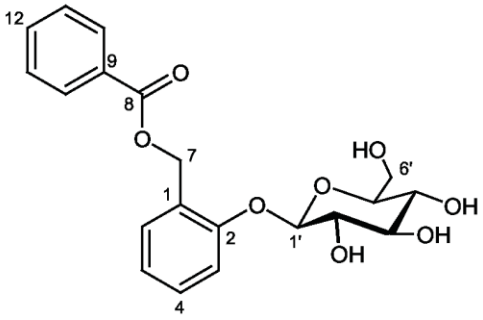
Salicyl-7-benzoate



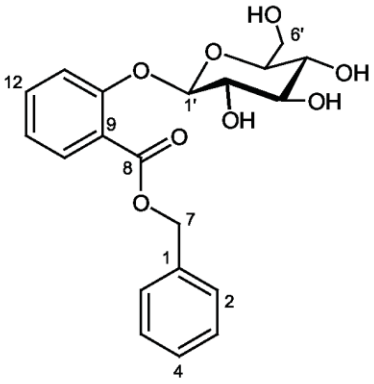
Salicyl-2-benzoate



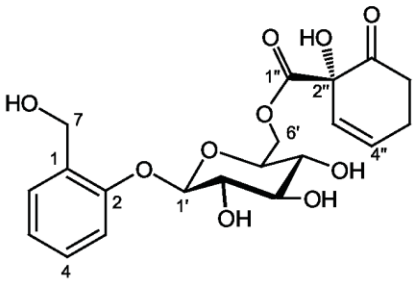
Salicyl-2,7-dibenzoate



Salicyl-7-benzoate glucoside (Desoxysalireposide)



Benzylsalicylate glucoside



6'-HCH-salicin

Figure 3

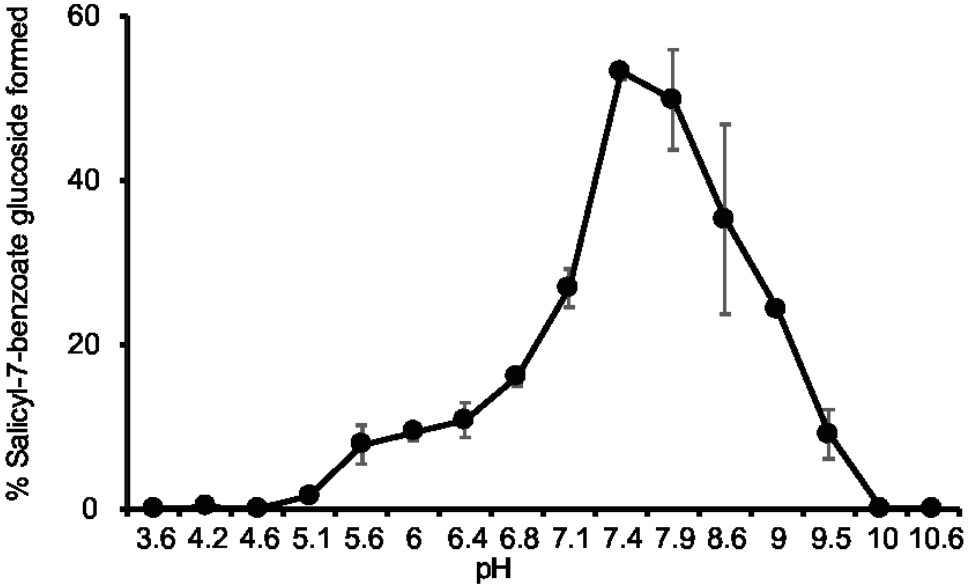


Figure 4

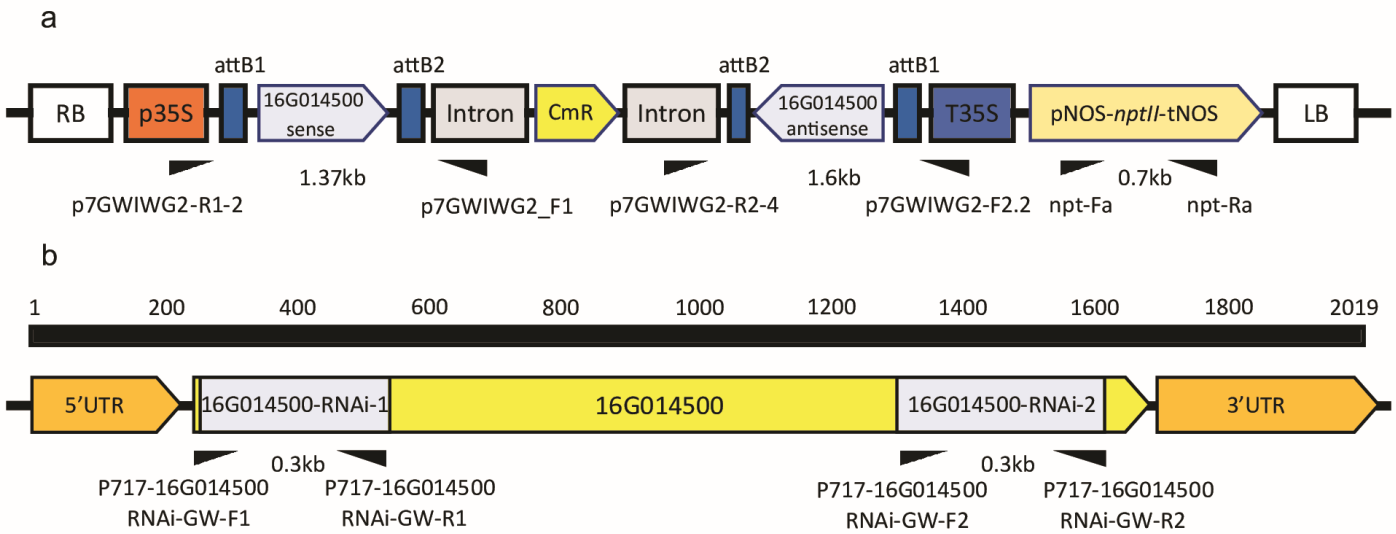
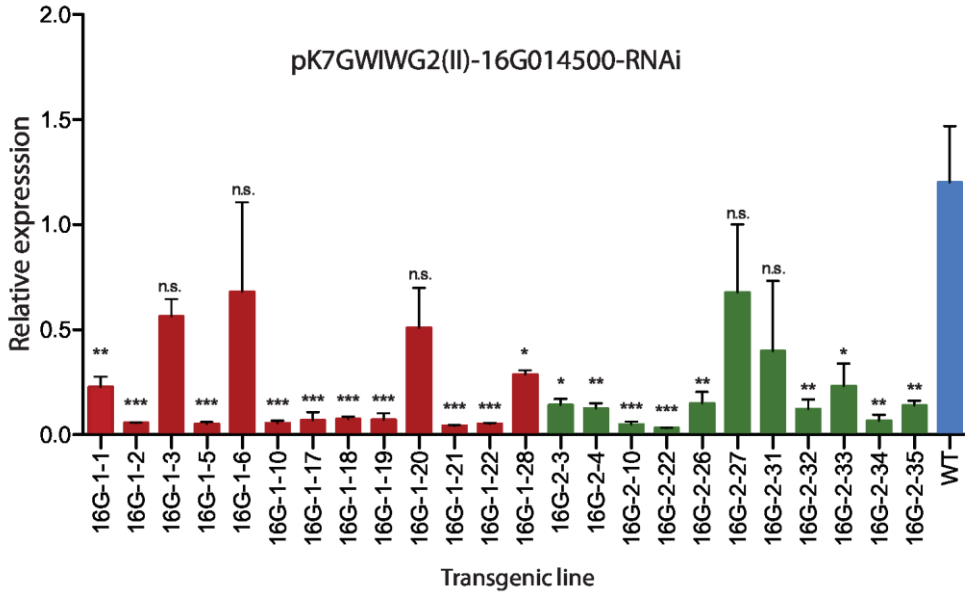


Figure 5



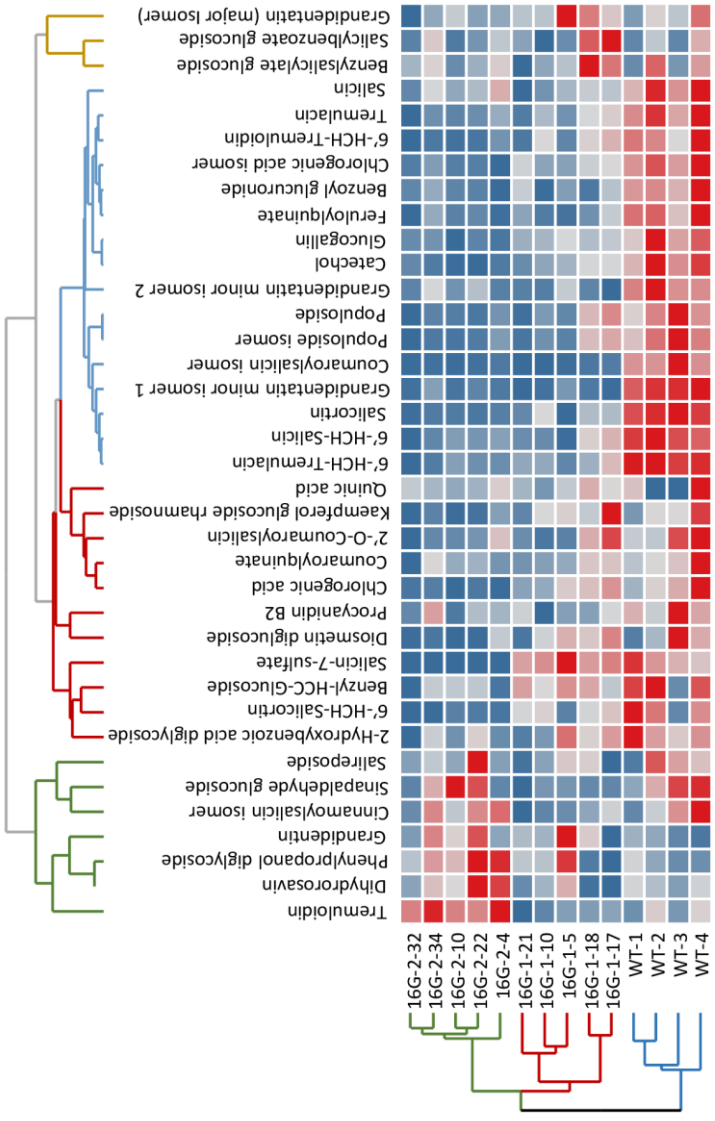


Figure 6

Figure 7

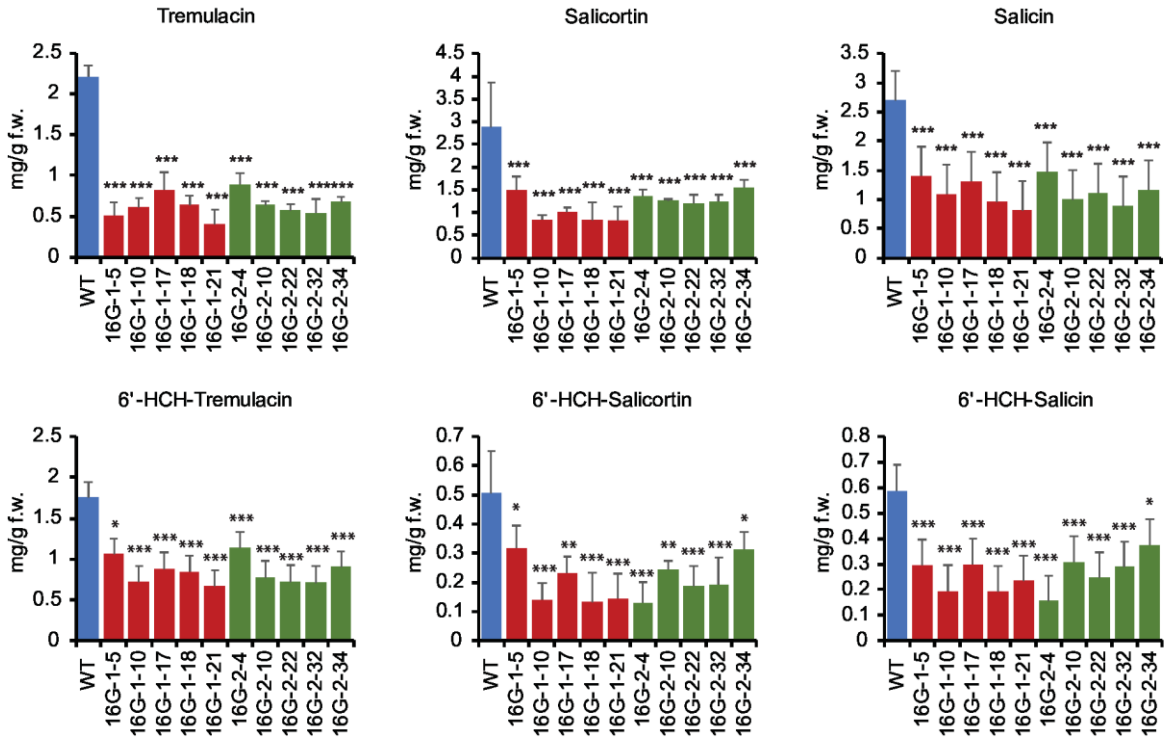


Figure 8