the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) doi: 10.1111/tpj.14615

Discovery of salicyl benzoate UDP-glycosyltransferase, a central enzyme in poplar salicinoid phenolic glycoside biosynthesis

Christin Fellenberg[1], Oliver Corea[1], Lok-Hang Yan[2], Finn Archinuk[1], Eerik-Mikael Piirtola[1,3], Harley Gordon[1], Michael Reichelt[4], Wolfgang Brandt[5], Jeremy Wulff[2], Jürgen Ehlting[1] and C. Peter Constabel[1,*]

[1]*Centre for Forest Biology and Department of Biology, University of Victoria, Victoria, British Columbia, Canada,*
[2]*Department of Chemistry, University of Victoria, Victoria, British Columbia, Canada,*
[3]*Department of Chemistry, University of Turku, Turku, Finland,*
[4]*Department of Biochemistry, Max-Planck Institute for Chemical Ecology, Jena, Germany, and*
[5]*Department of Bioorganic Chemistry, Leibniz Institute for Plant Biochemistry, Halle, Germany*

Received 4 July 2019; revised 26 September 2019; accepted 28 October 2019.
*For correspondence (e-mail cpc@uvic.ca).

SUMMARY

The salicinoids are anti-herbivore phenolic glycosides unique to the Salicaceae (Populus and Salix). They consist of a salicyl alcohol glucoside core, which is usually further acylated with benzoic, cinnamic or phenolic acids. While salicinoid structures are well known, their biosynthesis remains enigmatic. Recently, two enzymes from poplar, salicyl alcohol benzoyl transferase and benzyl alcohol benzoyl transferase, were shown to catalyze the production of salicyl benzoate, a predicted potential intermediate in salicinoid biosynthesis. Here, we used transcriptomics and co-expression analysis with these two genes to identify two UDP-glucose-dependent glycosyltransferases (UGT71L1 and UGT78M1) as candidate enzymes in this pathway. Both recombinant enzymes accepted only salicyl benzoate, salicylaldehyde and 2-hydroxycinnamic acid as glucose acceptors. Knocking out the UGT71L1 gene by CRISPR/Cas9 in poplar hairy root cultures led to the complete loss of salicortin, tremulacin and tremuloidin, and a partial reduction of salicin content. This demonstrated that UGT71L1 is required for synthesis of the major salicinoids, and suggested that an additional route can lead to salicin. CRISPR/Cas9 knockouts for UGT78M1 were not successful, and its *in vivo* role thus remains to be determined. Although it has a similar substrate preference and predicted structure as UGT71L1, it appears not to contribute to the synthesis of salicortin, tremulacin and tremuloidin, at least in roots. The demonstration of UGT71L1 as an enzyme of salicinoid biosynthesis will open up new avenues for the elucidation of this pathway.

Keywords: CRISPR/Cas9, phenolic glycosides, salicortin, tremulacin, salicin, Salicaceae.

INTRODUCTION

The poplars (Populus) and willows (Salix) are widespread, temperate trees that are known for their diverse phenolic secondary metabolites. These compounds include the proanthocyanidins (condensed tannins), hydroxycinnamate and phenolic esters, and the salicyl alcohol-based phenolic glycosides called salicinoids (Figure 1). The salicinoids are exclusive to the Salicaceae family, and more than 30 structures have been identified (Boeckler *et al.*, 2011; Keefover-Ring *et al.*, 2014). Salicinoids are characterized by a salicyl alcohol glucoside core, which is esterified with diverse acyl groups such as benzoic, cinnamic or *p*-coumaric acid. Many salicinoids also contain a 1-hydroxy-6-oxocyclohex-2-enoyl (HCH) moiety. The profile of salicinoids can vary extensively with species, but typically one or two abundant and several minor salicinoids are found (Boeckler *et al.*, 2011). For example, in *Populus tremuloides* and *Populus tremula* (trembling aspens), tremulacin and salicortin are the dominant compounds (Lindroth and Hwang, 1996), while *Populus nigra* accumulates primarily homaloside D and salicortin (Boeckler *et al.*, 2013). Overall, salicortin is the most widely distributed salicinoid (Boeckler *et al.*, 2011).

The function of salicinoids in defense against herbivory by both insects and vertebrates is well established (Lindroth and Hwang, 1996; Boeckler *et al.*, 2011). High salicinoid content deters browsing by mammals, for example the porcupine (*Erethizon dorsatum*), which climbs *P. tremuloides* to feed on the leaves (Diner *et al.*, 2009). For

© 2019 The Authors
The Plant Journal © 2019 John Wiley & Sons Ltd

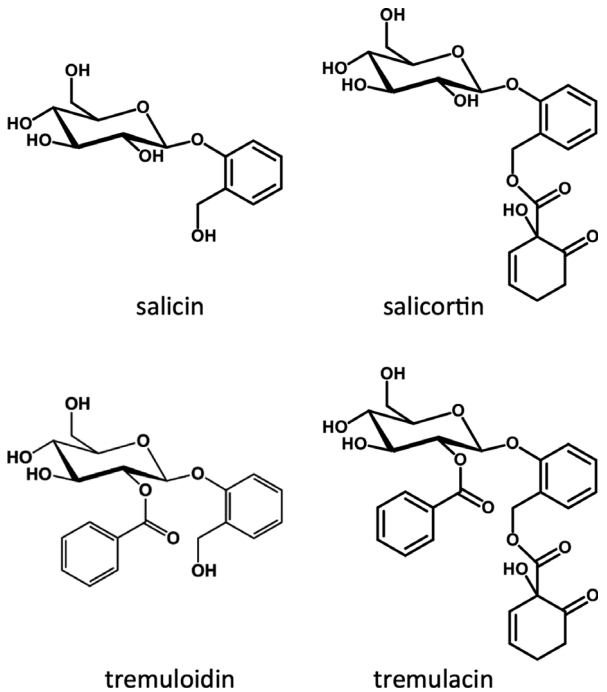


Figure 1. Salicyl alcohol-based phenolic glycosides (salicinoids) commonly found in Populus species.

tree-feeding lepidopterans such as forest tent caterpillar and gypsy moth, trees high in salicinoids are consistently linked to reduced growth and performance (Hwang and Lindroth, 1997). Negative effects on growth and development of diverse herbivorous insects has been demonstrated directly by coating leaves with purified tremulacin (Kelly and Curry, 1991; Hemming and Lindroth, 1995). In particular, the HCH moiety is important for salicinoid toxicity against lepidopterans (Lindroth and Hwang, 1996); its mechanism of toxicity is believed to be driven by fragmentative generation of *ortho*-quinone methide (Clausen *et al.*, 1990). By contrast, specialist insects such as the poplar leaf beetles in the genus Chrysomela are not deterred by salicinoids, but in fact use these for their own defense. The larvae sequester salicin into defensive glands, where it is enzymatically hydrolyzed and oxidized to salicylaldehyde. This is released by the larvae if attacked by predators (Pasteels *et al.*, 1983). In humans, the salicinoids may have beneficial effects: salicortin was identified in a screen of anti-adipogenic compounds in *P. balsamifera*, extracts of which are used in traditional medicine by the Cree First Nation in northern Canada (Martineau *et al.*, 2010). It is also notable that poplar and willow bark is used in traditional medicine and is rich in salicylic acid, the basis for the analgesic aspirin.

Total salicinoid content can reach 30% of the DW in *P. tremuloides* leaves (Donaldson *et al.*, 2006), although 8–16% is typical. Interestingly, there is extensive variation in salicinoid profiles and concentrations among different genotypes and natural populations (Hemming and Lindroth, 1995; Donaldson *et al.*, 2006). In *P. tremula,* Keefover-Ring *et al.* (2014) demonstrated broad variation in salicinoid content (1–17% DW) in natural populations, as well as strong heritability of salicinoid profiles. Furthermore, developmental patterns of salicinoid content between juvenile and mature foliage can vary dramatically, depending on the Populus species (Rehill *et al.*, 2006). Thus, the synthesis of these phenolics is under strong genetic and developmental control. They are much less responsive to environmental signals and are thus considered as constitutive secondary metabolites.

Despite the extensive knowledge of salicinoid structures and distribution, little is known about their biosynthesis. Zenk (1967) carried out C-14 labeling experiments in Salix, and demonstrated the incorporation of salicylaldehyde into salicin, the simplest salicinoid. Likewise, Babst *et al.* (2010) showed that salicyl alcohol is incorporated into salicin, but not salicortin; conversely, benzyl alcohol, benzaldehyde and benzoic acid are all incorporated with varying efficiencies into salicortin, but not salicin. Subsequent conversions remain unclear, however, and no enzymes have been identified to date. These earlier experiments suggest that salicortin and higher salicinoids are synthesized by a pathway that is distinct from salicin biosynthesis. Based on Babst *et al.* (2010), one possible biosynthetic route to salicortin is via benzyl-CoA and benzyl alcohol, and in particular their ester benzyl benzoate (Babst *et al.*, 2010). Building on this hypothesis, we used expression patterns and phylogenetic analysis to identify two BAHD acyltransferases from poplar, and showed that they can generate esters from benzyl-CoA and alcohol acceptors. Biochemical characterization determined that one of these acyltransferases is the expected benzoyl-CoA: benzyl alcohol O-benzoyltransferase (BEBT), which synthesizes benzyl benzoate from benzyl-CoA and benzyl alcohol, whereas the second enzyme is a benzoyl-CoA: salicyl alcohol O-benzoyltransferase (SABT), which forms salicyl benzoate using salicyl alcohol as co-substrate (Chedgy *et al.*, 2015). Both these enzymes are promising candidates and have activities consistent with a possible salicinoid pathway. If salicyl benzoate is an intermediate in the biosynthesis of salicortin and other higher salicinoids, glucosylation would be a potential next step in the pathway. Benzyl benzoate could first be hydroxylated to generate salicyl benzoate, and then glucosylated. In both cases, to further validate a route to salicortin via these esters, a salicyl benzoate-specific glucosyltransferase would have to be identified.

The UDP-glucose-dependent glycosyltransferases (UGTs) are widespread enzymes in plants with many roles in secondary plant metabolism, hormones and detoxification of other small molecules (Bowles *et al.*, 2006). The key unifying feature of this group of enzymes is the conserved

© 2019 The Authors

C-terminal plant secondary product glycosyltransferase (PSPG) motif that constitutes the UDP-sugar-binding site. UGTs comprise the largest family (GT family 1) within the glycosyltransferase superfamily and are encoded by a large number of genes in plant genomes. Caputi *et al.* (2012) list 107 UGT genes in *Arabidopsis thaliana*, 182 genes in soya and over 200 genes in apple. In Populus, we previously identified 171 UGT genes (Veljanovski and Constabel, 2013). The UGTs are systematically named based on protein sequence similarity, with the numbered UGT families defined as sharing more than 40% protein sequence identity. The plant UGTs fall within UGT families UGT71 to UGT100, according to the CAZy nomenclature system (Mackenzie *et al.*, 1997; Yonekura-Sakakibara and Hanada, 2011). In poplar, few UGTs with known functions are known; to date only UGT78L1 and UGT84A17, which glycosylate flavonoids and hydroxycinnamic acids, respectively, have been characterized (Veljanovski and Constabel, 2013; Babst *et al.*, 2014).

Phylogenetic analysis of characterized UGT genes suggests that many of these cluster with other UGTs with similar sugar acceptors; for example, many flavonoid-specific UGTs fall into the same clade (Gachon *et al.*, 2005). However, in other cases, UGTs with similar primary sequences were shown to have very distinct substrate preferences (Modolo *et al.*, 2009; Osmani *et al.*, 2009). Conversely, the same acceptor molecules can be substrates for UGT enzymes from phylogenetically distant clades. Thus, it has proven difficult to predict substrates or identify specific candidate genes based on primary sequence data alone, and functional analysis of individual genes is required (Osmani *et al.*, 2009; Caputi *et al.*, 2012).

Salicinoid synthesis is developmentally regulated, and the pathway does not appear to be stress-inducible. Therefore, the identification of candidate genes by differential screening based on inducible expression has not been successful. However, with large transcriptomic datasets, co-expression analysis is becoming a powerful approach. Here, we took advantage of a large poplar transcriptome dataset to identify two candidate salicinoid UGTs. We biochemically characterize these UGTs and demonstrate that they catalyze the glucosylation of salicyl benzoate. We further demonstrate a role in salicinoid biosynthesis using CRISPR/Cas9-engineered knockout hairy root cultures for one of the new UGTs. Our data confirm that UGT71L1 is a key enzyme in salicinoid biosynthesis, providing a road map for further elucidation of the pathway.

RESULTS

Identification of glycosyltransferases involved in salicinoid biosynthesis

To identify candidate genes involved in the salicinoid pathway, we used a large leaf RNA-seq transcriptome dataset from nearly 200 natural accessions of *Populus trichocarpa* to screen for genes co-expressed with SABT and BEBT. Both analyses identified partially overlapping sets of co-expressed genes containing a number of genes with enzyme annotations related to functions in secondary metabolism (Tables 1 and S1). The SABT co-expression data included 3-deoxy-D-arabino-heptulosonic acid 7-phosphate (DAHP) synthase, cytochromes P450, epimerase/dehydratases with undefined functions, and two UGT enzymes (Table 1). We focused on the UGTs, based on our hypothesis that salicyl benzoate is glucosylated as an early step in salicinoid biosynthesis. One of the co-expressed UGT genes encoded UGT78M1, a UDP-glucose-dependent glycosyltransferase that we had previously investigated, but for which no natural substrates were found (Veljanovski and Constabel, 2013). The second co-expressed UGT gene encoded UGT71L1; this gene had been identified earlier as a candidate gene for the salicinoid pathway but not characterized further (Tsai *et al.*, 2011).

We first investigated expression patterns of the two UGTs in different plant tissues in comparison with SABT using the Popgenie *P. tremula* expression database (http://popgenie.org). Data were extracted and plotted as normalized expression in several vegetative and reproductive organs (Figure 2). The highest UGT expression was observed in young organs (buds, flowers, leaves, roots), while expression was lowest in expanded flowers, mature leaves and petioles, and seeds. UGT71L1 showed the highest relative expression, but the expression pattern of both UGTs and SABT followed the same pattern. This supported our co-expression data and assumption that these genes are involved in the same biosynthetic pathway.

For additional insight into potential functions of UGT71L1 and UGT78M1, we aligned their protein

Table 1 Top 10 genes most highly co-expressed with poplar benzoyl-CoA:SABT

Gene ID	PCC[a]	Gene annotation (PFAM)
Potri.013G074500	1	Transferase family (SABT)
Potri.009G117000	0.7265	Alpha/beta hydrolase fold
Potri.018G051300	0.7191	Cytochrome P450
Potri.004G082000	0.7100	AMP-binding enzyme
Potri.016G014500	0.7099	UDP-glycosyltransferase (UGT71L1)[b]
Potri.014G019700	0.7081	NAD-dependent epimerase/dehydratase
Potri.002G120700	0.7074	NAD-dependent epimerase/dehydratase
Potri.003G066400	0.7058	Cytochrome P450
Potri.009G104600	0.7026	Alpha/beta hydrolase fold
Potri.006G171100	0.6925	UDP-glycosyltransferase (UGT78M1)[b]
Potri.002G099200	0.6865	Class-II DAHP synthetase

SABT, salicyl O-benzoyltransferase.
[a]Pearson correlation coefficient.
[b]Candidate genes characterized in this study.

© 2019 The Authors

sequences with similar previously characterized UGTs and carried out a phylogenetic analysis (Figure 3). We included a set of Arabidopsis UGTs that had been systematically screened for activity with a variety of hydroxylated benzoic acids, including salicylic acid (Lim *et al.*, 2002). Many previously studied UGTs are active with flavonoids, and UGT78M1 grouped with a clade that consists of known flavonoid UGTs. Despite this similarity, this enzyme does not accept common flavonoid substrates that we had tested (Veljanovski and Constabel, 2013). However, it was active with 2,4,5-trichlorophenol, suggesting it might accept other phenolics as substrates.

By contrast, UGT71L1 grouped within a clade of enzymatically diverse UGTs that included a characterized UGT from *Medicago truncatula,* UGT71G1 (Figure 3). The Medicago enzyme was found to have broad substrate preferences and showed activity with the triterpene saponin medicagenic acid, as well as several isoflavones and quercetin (Achnine *et al.*, 2005). Furthermore, the Arabidopsis UGTs in this cluster, AtUGT71C1 and AtUGT71C4, glucosylate benzoic acid derivatives at the 3-OH position although their *in vivo* substrates are not known (Lim *et al.*, 2002). Thus, the localization of PtUGT71L1 within this cluster would be consistent with a role in glycosylation of salicyl benzoate or other salicyl alcohol-derived compounds.

Biochemical characterization of co-expressed poplar UGT enzymes

To functionally characterize UGT71L1 and UGT78M1, we cloned their respective cDNAs into a his-tagged protein expression vector for expression in *Escherichia coli* (Veljanovski and Constabel, 2013). The recombinant proteins were solubilized from lysed cells and purified using nickel affinity resin. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis confirmed that the proteins had the expected sizes, 53 kDa for UGT71L1 and 51 kDa forUGT78M1 (Figure S1). The UGT71L1 preparation contained additional polypeptides; however, more stringent washes could not be applied as they led to loss of the enzyme. Western blot analysis with a his-tag specific antibody confirmed that we had purified full-length proteins in both cases (Figure S1). Preliminary enzyme assays using UDP-glucose as a sugar donor and the widely accepted substrate 2,4,5-trichlorophenol as the sugar acceptor indicated that both proteins had been obtained in active form. We subsequently tested this activity for temperature stability and at multiple pHs (Figure S2). Both enzymes had very similar temperature profiles, with maximal activity at 40°C and 43°C for UGT78M1 and UGT71L1, respectively. They also had identical pH preference profiles, with greatest activity at pH 7.5 (Figure S3).

We next tested the enzymes with salicyl benzoate, the product of the SABT enzyme and potential intermediate for salicinoid biosynthesis (Chedgy *et al.*, 2015). Salicyl benzoate is not commercially available, and was synthesized from salicyl alcohol and benzoic acid as described in Experimental Procedures. Both enzymes efficiently converted salicyl benzoate to a product that eluted with a shorter retention time in reversed-phase high-performance liquid chromatography (HPLC), as expected for a glycosylated compound (Figure 4a). Boiled enzyme controls did not show any product formation detectable by HPLC-UV. Liquid chromatography-mass spectrometry (LC-MS) analysis of the salicyl benzoate reaction product identified a peak of m/z 413 in positive mode and m/z 435 in negative mode (Figure 4b). These corresponded to the Na^+ and formate adducts of salicyl benzoate glucoside, which has the expected mass of 390 Da. We additionally tested 29 different phenolic substrates including benzoate and salicylate derivatives, hydroxycinnamic acids and different flavonoids (Table S2). Remarkably, both enzymes showed an almost identical substrate profile, with only salicyl

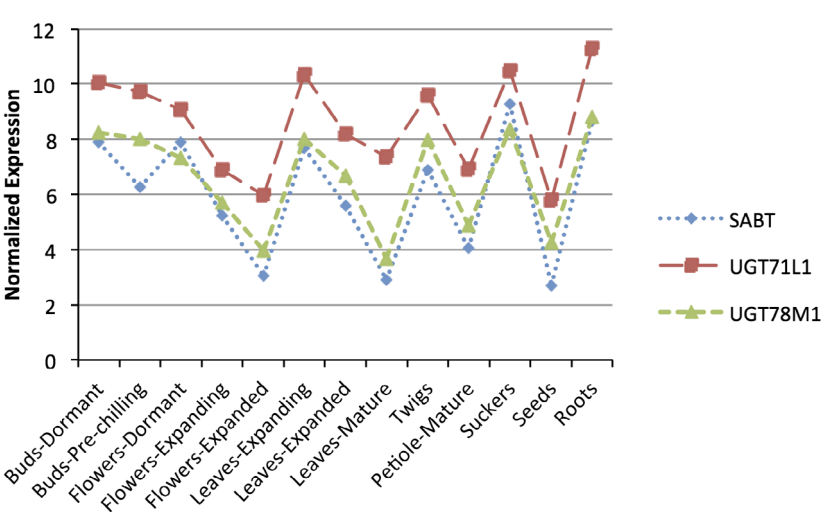


Figure 2. Co-expression of UGT71L1 and UGT78M1 with the salicyl alcohol O-benzoyltransferase (SABT) gene in developmental stages and organs of *Populus tremula*.
Expression data were obtained from the PopGenie RNA-seq expression database (*P. tremula* exAtlas; http://popgenie.org).

© 2019 The Authors

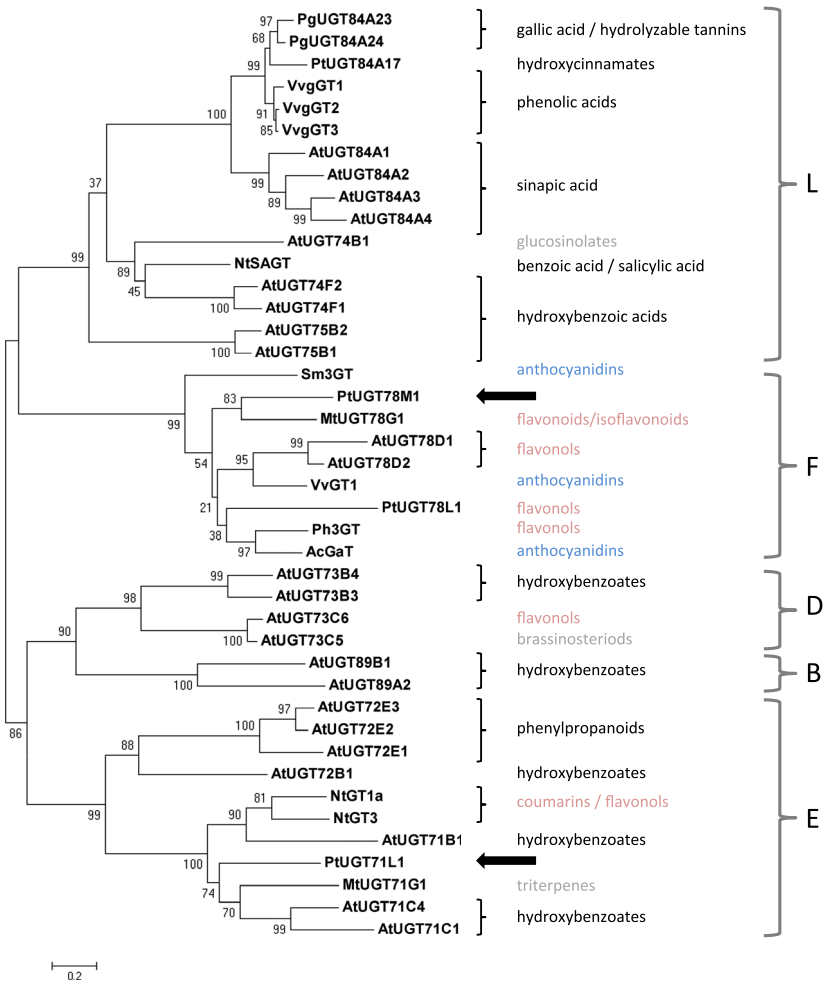


Figure 3. Maximum-likelihood phylogenetic tree of poplar UGT71L1 and UGT78M1 together with other characterized UGTs.
Bootstrap support (1000) for the branches is shown. UGT groups are as defined by Ross et al. (2001).

aldehyde and 2-hydroxycinnamic acid (*o*-coumaric acid) showing readily detectable products (Figure 4c and e). LC-MS analysis again confirmed the product peaks as the respective glucosides (Figure 4d and f). Enzyme activity with salicyl alcohol was detected, but only at extremely low rates, which necessitated an overnight incubation of the assay; this substrate was therefore not included in further work. Salicylic acid was not a substrate for either enzyme, and no activity was detectable using other benzoates and hydroxycinnamate derivatives as glucose acceptors (Table S2). Neither UGT showed activity with an alternate sugar donor, UDP-galactose.

Detailed analysis of the UGT activity with salicyl benzoate was carried out in a buffer at pH 5.1, as we determined that this compound was not stable in solutions above pH 5.5. For other substrates, the assays were conducted at the enzyme optimum of pH 7.5. Therefore, we carried out assays at both pHs for comparative purposes. Relative conversion rates for different substrates were obtained by setting the specific activities for salicyl benzoate for each enzyme to 100% (Table 2). Both UGTs showed much lower activity with the other accepted substrates, salicylaldehyde and 2-hydroxycinnamic acid, at both tested pHs. Generally, enzyme activities were reduced at lower pH. These experiments indicated that salicyl benzoate is the preferred substrate for both UGTs, and that both enzymes are salicyl benzoate-UDP glucose glycosyltransferases.

We determined kinetic parameters of both enzymes for the three substrates, at both pHs where possible. This enabled comparisons among all substrates at pH 5.1, while also providing data closer to the enzyme's pH optimum. For UGT71L1, the K_m determined (29.3 μM at pH 5.1) for salicyl benzoate was 30-fold lower than that for salicylaldehyde (at pH 7.5) and 20-fold lower than for 2-hydroxycinnamic acid at pH 5.1 (Table 3); together with a lower V_{max}, this results in at least two orders of magnitude higher catalytic efficiency (V_{max}/K_m) for salicyl benzoate compared with the other substrates. This established that salicyl benzoate was the preferred substrate of this enzyme. Similarly, we determined a K_m of 39.1 μM for salicyl benzoate with UGT78M1, which was

© 2019 The Authors

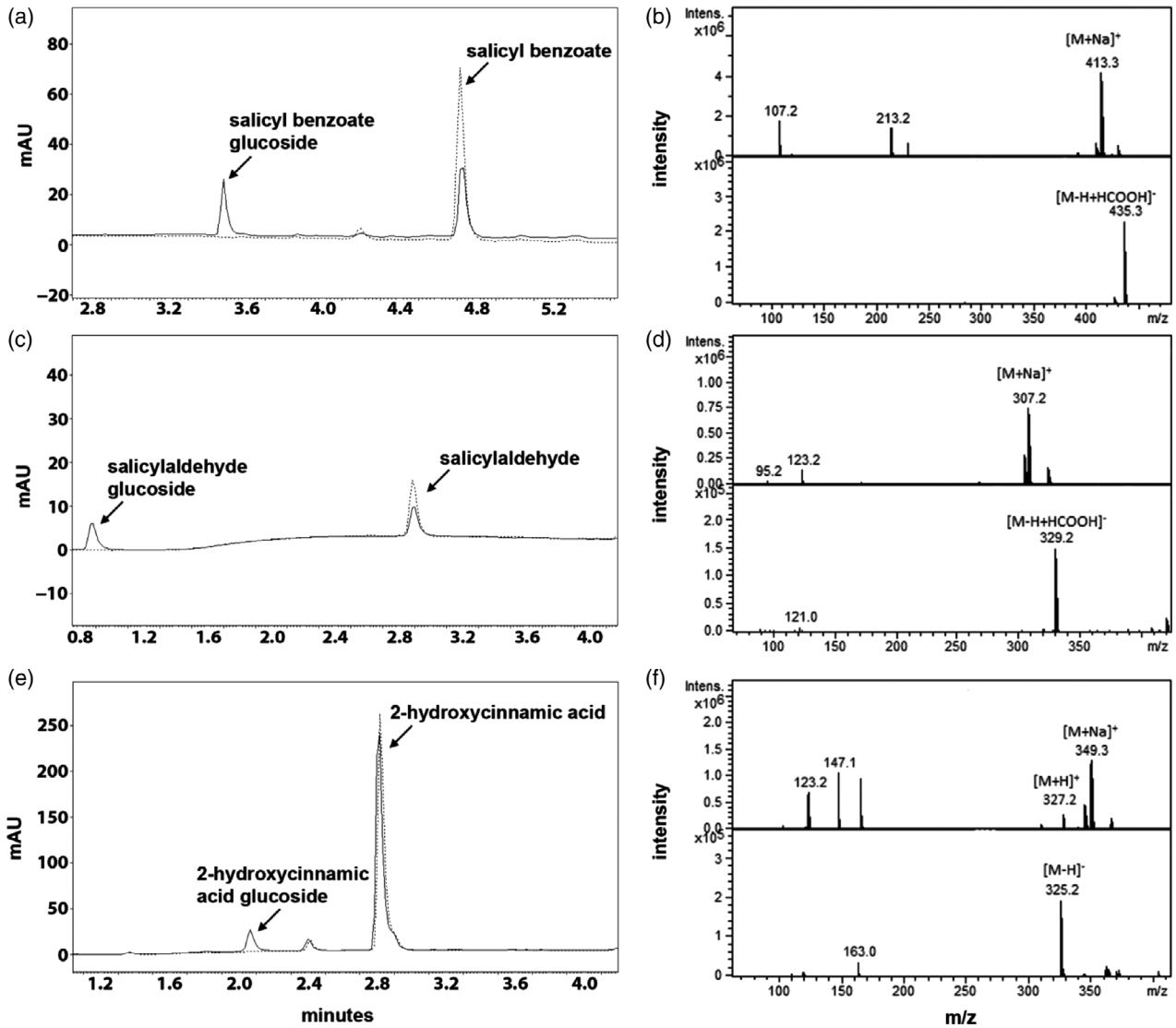


Figure 4. Characterization of UGT enzyme assay product formation by high-performance liquid chromatography (HPLC) and liquid chromatography-mass spectrometry (LC-MS) using (a, b) salicyl benzoate, (c, d) salicylaldehyde, and (e, f) 2-hydroxycinnamic acid as glucose acceptors.
Left panels shows HPLC chromatograms (280 nm) of enzyme assays performed with native enzyme (solid line) in comparison to boiled enzyme controls (dotted line). Products were further characterized by LC-MS (right panels). Expected product masses and their aglycones were verified in both positive and negative ionization mode shown in the upper and lower halves of each panel, respectively.

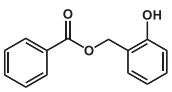
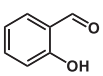
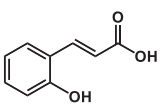
also lower than for the other substrates tested. For this UGT, however, 2-hydroxycinnamic acid gave a K_m of 70.9 μM at pH 5.1, suggesting a relatively greater affinity for this metabolite compared with UGT71L1. The higher V_{max} values obtained for UGT78M1 likely reflect the higher purity of this protein preparation and thus a higher UGT enzyme concentration in the assay compared with the UGT71L1 preparation, which contained a number of additional co-purified proteins (Figure S1). However, for both enzymes, these K_m values were comparable to those determined for secondary metabolite-specific UGTs in our previous work with UGT78M1 (He *et al.*, 2006; Veljanovski and Constabel, 2013).

Molecular modeling of UGT structures

The similar substrate preferences of the UGTs, despite their low primary sequence identity (27.9%), prompted us to investigate their respective protein structures. Several plant UGT structures have been solved, which enabled us to model the new poplar UGTs. Optimal templates were identified and homology models created for UGT78M1 based on Arabidopsis UGT72B1 (Brazier-Hicks *et al.*, 2007) and UGT71L1 based on Medicago UGT78G1 (He *et al.*, 2006). Despite using different templates for modeling, the tertiary structure models of UGT78M1 and UGT71L1 were almost identical (Figure 5). The strong conservation of

© 2019 The Authors

Table 2 Glycosyltransferase activity of UGT71L and UGT78M1 towards different phenolic substrates at two pHs

		Conversion rate (%)[a]			
		UGT71L1		UGT78M1	
Substrate	Structure	pH 7.5	pH 5.1	pH 7.5	pH 5.1
Salicyl benzoate		nd	100	nd	100
Salicylaldehyde		6.25	nd	0.47	0.20
2-hydroxycinnamic acid		4.47	0.56	0.15	0.35

[a]Each recombinant UGT was incubated with 50 μM phenolic substrate and 5 mM of the sugar donor. Relative conversion rates were obtained by setting the specific activities for salicyl benzoate for each enzyme to 100% (UGT71L1 602.5 pkat mg^{-1} and UGT78M1 31 419.5 pkat mg^{-1}). Activities were determined in triplicate and nd means no activity detectable or too low to determine kinetic properties.

tertiary structure for UGTs has been noted previously (Osmani *et al.*, 2009). As seen for other UGTs, in the models UDP-glucose and phenolic glucose acceptors fit into the cleft that is formed between the N- and C-terminal portions of the polypeptide (Figure 5a and b). In both structures, the co-substrate UDP-glucose is held in position by key amino acid residues within the PSPG motif, but these are different in each of the structures. In UGT71L1, the sugar moiety interacts with Glu378 and Glu394 (Figure 5c); in UGT78M1, this function is carried out by Asp356 and Asp372 (Figure 5d). The catalytic dyad is conserved in both UGTs, as His23/ Asp127 in UGT71L1 and His21/ Asp118 in UGT78M1.

The optimal glucose acceptor, salicyl benzoate, can be stabilized in the active site of both enzymes, which allows the transfer of the glucose facilitated by proton abstraction of its 2-hydroxyl group by the catalytic dyad. Hydrophobic interactions of key residues with the benzene ring enhance the affinity and stabilize positioning of salicyl benzoate. Val18 appears to be critical for this interaction in UGT78M1, whereas Pro90 is available for this function in UGT71L1 (Figure 5c and d). These hydrophobic interactions are missing if salicylaldehyde or 2-hydroxycinnamate are modeled, resulting in non-productive alternative dockings and lower affinity (Figures S3 and S4). This is consistent with the lower conversion rates that we observed in enzyme assays for these substrates.

Generation and analysis of UGT CRISPR mutants in poplar hairy roots

In order to test the roles of these UGT activities in salicinoid synthesis *in vivo*, we employed a genome editing strategy using CRISPR/Cas9, an approach recently adapted to poplar (Zhou *et al.*, 2015; Elorriaga *et al.*, 2018). The interspecific poplar hybrid most commonly transformed (*P. tremula* × *Populus alba*, clone *7171-B4*) is highly heterozygous; this makes selection of suitable gRNAs challenging, as alleles of target sequences contain many species-specific single nucleotide polymorphisms (SNPs). To facilitate gRNA selection in this poplar hybrid, we made use of a database of potential SNP-free gRNAs for *P. tremula* × *P. alba* created for this purpose by Zhou *et al.* (2015).

Table 3 Michaelis–Menten kinetic parameters of recombinant UGT71L1 and UGT78M1 with selected phenolic substrates

	UGT71L1			UGT78M1		
	K_m [μM]	V_{max} [pkat mg^{-1}]	V_{max}/K_m	K_m [μM]	V_{max} [pkat mg^{-1}]	V_{max}/K_m
Salicyl benzoate pH 5.1	29.3 ± 2.2	954.4 ± 27.2	32.57	39.1 ± 12.0	59 410.0 ± 9080.0	1519.44
Salicylaldehyde pH7.5	957.7 ± 70.8	467.9 ± 12.1	0.49	3141.7 ± 281.4	10 262.2 ± 467.4	3.27
Salicylaldehyde pH5.1	nd[a]	nd	–	4005.9 ± 322.0	4729.8 ± 210.8	1.18
2-Hydroxycinnamic acid pH 7.5	2465.1 ± 288.4	917.8 ± 50.0	0.37	368.8 ± 29.3	264.3 ± 6.4	0.72
2-Hydroxycinnamic acid pH 5.1	620.8 ± 42.6	54.6 ± 1.2	0.09	70.9 ± 2.7	263.5 ± 2.6	3.72

[a]nd, no activity detected or too low to determine kinetic properties.

© 2019 The Authors

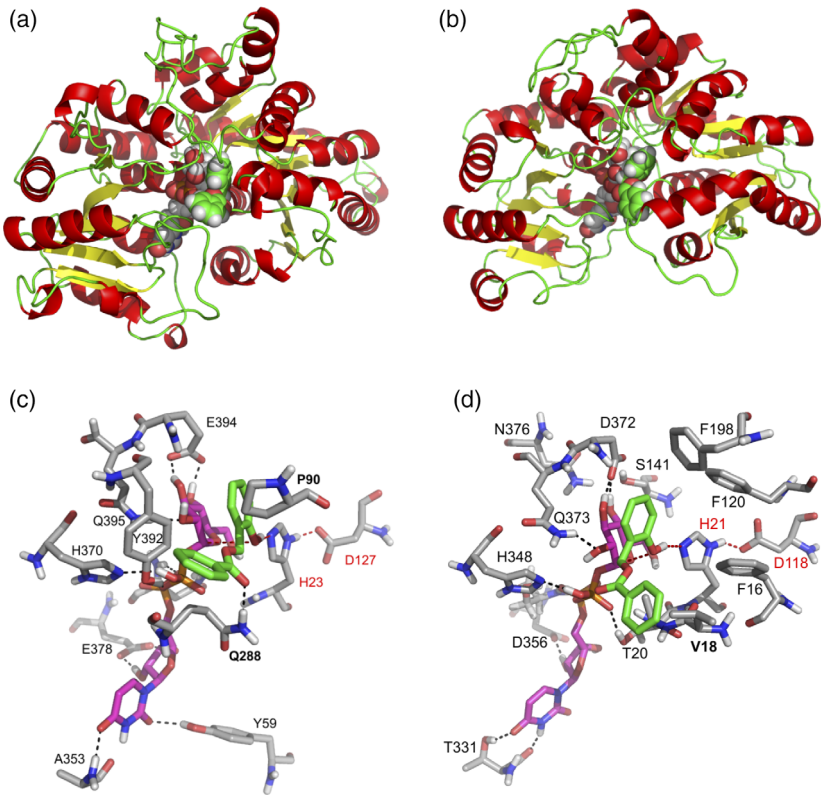


Figure 5. Tertiary structure of (a) UGT71L1 and (b) UGT78M1 bound with UDP-glucose (magenta carbon atoms) and salicyl-benzoate (green carbon atoms).
(c) Detail of the active sites of UGT71L1 and UGT78M1 with bound UGT and salicyl-benzoate are shown in (c) and (d). Red dotted lines indicate the reaction coordinates and black dotted lines important hydrogen bonds for recognition of the substrates. Bold labeled amino acid residues are essential for the binding of salicyl-benzoate and red labeled ones represent the catalytic diad.

We selected unique gRNA targets for both UGTs near the 5′-end of the coding sequence and with the lowest potential for off-target matches. These target gRNA sequences were cloned into vectors containing Cas9 and other CRISPR components using the P201G plant transformation vector created by Jacobs *et al.* (2015). The vector also contains eGFP as a selective marker, facilitating selection of hairy roots containing the CRISPR construct. Transgenic hairy roots expressing CRISPR constructs were generated using previously established methods (Yoshida *et al.*, 2015). Positive transformants were selected for GFP expression, and then further subcultured in order to obtain sufficient tissue for phytochemical analysis.

Multiple transformed hairy root lines were screened for altered salicinoid content using UHPLC-UV-MS. *Populus tremula* × *P. alba* roots contain primarily salicortin and tremulacin, which were quantified using UV detection at 270 nm. Salicin and tremuloidin are present at 10-fold lower concentrations, and we quantified these less abundant compounds by MS detection (see Experimental Procedures and Table S3). Overall, the salicinoid content of the hairy root extracts showed substantial variability; this likely reflects the uneven growth of hairy roots, which makes it challenging to obtain tissue samples of similar age and size across all samples. Despite this, in approximately 40% of the confirmed positive hairy root lines expressing the UGT71L1 gRNA construct, we observed the complete loss of tremulacin, tremuloidin and salicortin. In affected lines salicin content was reduced, but not eliminated (Figure 6). By contrast, none of the UGT78M1 gRNA hairy root lines we created showed significant differences in salicinoid content relative to the empty vector control lines.

To confirm that the phytochemical phenotype of the UGT71L1 CRISPR hairy root cultures observed was due to specific mutations in the UGT71L1 gene, we amplified and sequenced the gRNA-targeted genome sequences for a subset of these hairy root lines. Polymerase chain reaction (PCR) products were sequenced directly to identify mutations in both alleles. Where sequence analysis was ambiguous, PCR products were cloned and multiple colonies sequenced. Because both parental alleles contain distinct allele-specific SNPs in the gRNA target region, these SNPs could be used to confirm that both the *P. tremula* and *P. alba* alleles had been amplified and that mutations are present in both alleles (Figure 7a). All transgenic lines showing a loss of salicinoids had biallelic mutations in the UGT71L1 gene. As expected from the CRISPR/Cas9 system, we found short deletions and insertions in both alleles to be the most prevalent mutations (Figure 7a). The majority of the identified indels were located three nucleotides upstream of the PAM within the seed sequence of the gRNA. To confirm specificity of the gRNAs, we checked for off-target effects in sequences in the *P. tremula* × *P. alba* 717I-B4 genome sequences with similarity to UGT71L1.

© 2019 The Authors

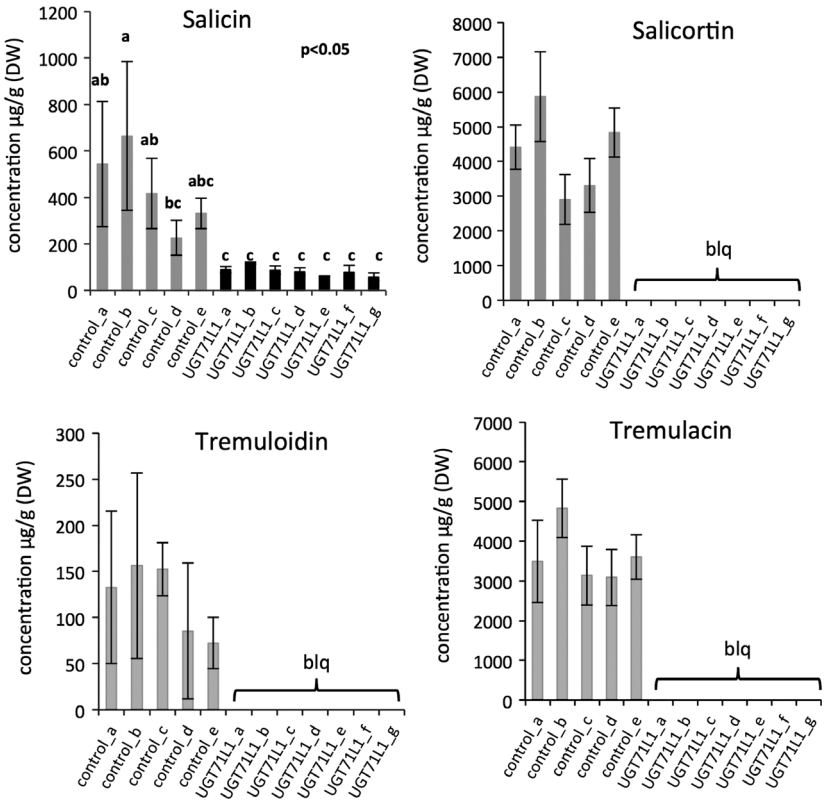


Figure 6. Salicinoid content in UGT71L1 CRISPR knockout and control poplar hairy root cultures. Salicortin and tremulacin are quantified based on the UV response at 270 nm, and salicin and tremuloidin were quantified based on the molecular ions (M-H). Bars show means ± SD of at least three biological replicates per line. Data for salicin were analyzed by one-way ANOVA followed by Tukey's honest significant difference (HSD) *post hoc* test in R (https://www.r-project.org). Different letters indicate statistically significant differences at 95% confidence. blq, below limit of quantification.

The five most similar potential off-target sequences were identified using the Pta717 variant data base (http://aspendb.uga.edu/s717), and also amplified and sequenced. No mutations in any of these potential targets were detected (Figure 7b). Together, these data indicated that biallelic mutations in UGT71L1 were only present in the affected hairy root lines. Therefore, the loss of salicinoids was directly caused by the loss of UGT71L1.

By contrast, sequencing of gRNA target amplicons from the UGT78M1 hairy root culture uncovered no mutations or changes to genomic sequences from wild-type. This result matched the phytochemical analysis, which had shown no change in salicinoid content in UGT78M1 transgenic hairy root cultures. Closer inspection of the target sequences and comparison with an updated *P. tremula* × *P. alba 717-1B4* genome database (version 2) indicated that our targeted gRNA sequence (based on version 1 of the genome) contained a mismatch, which is located three nucleotides upstream of the PAM sequence and present in both the *P. tremula* and *P. alba* UGT78M1 alleles. The *P. alba* allele also contained two additional mismatches in this critical region. These differences explain why a mutation could not be generated with our construct. Although we were unable to produce UGT78M1 mutant hairy roots, the results confirmed the specificity of the CRISPR system, as a single gRNA mismatch was sufficient to prevent CRISPR/Cas9 from generating mutations.

DISCUSSION

The salicinoid phenolic glycosides are ecologically important phenolics, yet their biosynthesis has been difficult to elucidate. Here, we describe an enzyme of salicinoid biosynthesis, salicyl benzoate: UDP-glucose glycosyltransferase. At least two genes encode this activity, UGT71L1 and UGT78M1, and both recombinant enzymes showed strong preference for salicyl benzoate. CRISPR/Cas9-induced mutations in UGT71L1 caused a complete loss of salicortin, tremulacin and tremuloidin in hairy root cultures, but only led to reduction of salicin content. The implication of this enzyme in salicinoid biosynthesis, together with our previous characterization of an enzyme that generates salicyl benzoate (Chedgy *et al.*, 2015), provide the outlines of a potential biosynthetic route to the salicinoids.

Our CRISPR/Cas9 genome editing experiments demonstrated that UGT71L1 is absolutely required for salicinoid accumulation, at least in roots. We are confident that this phenotype is due to the specific mutations in UGT71L1 observed, as we detected no off-target mutations when we sequenced the five most similar loci to the gRNA in all transgenic lines showing a loss of salicinoids. Furthermore, in experiments with UGTM78L1, a single mismatch in the gRNA to the target was sufficient to prevent mutations from being generated, further demonstrating the

© 2019 The Authors

(a)

Line	Allele	Sequence	Indel	Frequency
WT	tremula	GGTGTACCCCATGTTTACAC----------TGTGGGGCC	0	3/6
	alba	GGTGTACCCCATGTTTACAC----------TGTGGGACC	0	3/6
UGT71L1_a	tremula	GGTGTACCCCATGTTT----------TGTGGGGCC	–4	3/10
	alba	GGTGTACCCCATGTTTACACA----------TGTGGGACC	+1	7/10
UGT71L1_d	tremula	GGTGTACCCCATGTTTAC----------TGTGGGGCC	–2	2/8
	alba	GGTGTACCCCATGTTTACACT----------TGTGGGACC	+1	6/8
UGT71L1_e	tremula	GGTGTACCCCATGTTTA----------GTGGGGCC	–4	3/10
	alba	GGTGTACCCCATGTTTACA----------TGTGGGACC	–1	7/10
UGT71L1_f	tremula	GGTGTACCCCATGTTTACAC----------GTGGGGCC	–1	6/11
	alba	GGTGTACCCCATGTTTACACG----------TGTGGGACC	+1	5/11
UGT71L1_g	tremula	GGTGTACCCCATGTT----------GGCC	–10	3/6
	alba	GGTGTACCCCATGTTTACACTGAA...GAGATGGGTTTTTATGGGACC	+37	3/6

(b)

Site	Allele	Off target site sequence	Poplar gene ID	Mismatches	Events	Mutations
A	tremula	agACCCaAaGTTTACACTcT**GGG**	Potri.001G119600	5	6	0
	alba	agACCCaAaGTTTACACTtT**AGG**				
B	tremula	GTAggCCATGTTTACAaTaT**AGG**	Potri.001G383500	4	6	0
	alba	GTAggCCATGTTTACAaTaT**AGG**				
C	tremula	GTcCAgCATcTTTACtCTGT**GGG**	Potri.004G120700	4/5	6	0
	alba	GTcCAgCATcTTTACttTGT**GGG**				
D	tremula	acACCCCcaGTTTACACaGT**GGG**	Potri.016G015700	5	6	0
	alba	acACCCCcaGTTTACACaGT**GGG**				
E	tremula	GTgggCCATGTTTACAaTGT**TGG**	Potri.011G103700	4	6	0
	alba	GTgggCCATGTTTACAaTGT**TGG**				

Figure 7. Characterization of CRISPR/Cas9 mediated gene editing of UGT71L1 in transgenic poplar hairy roots.
(a) Mutations identified at the UGT71L1 locus of several independent transgenic lines. Underlined sequences indicate gRNA target region and bold letters highlight PAM sequence. Yellow highlights indicate the single nucleotide polymorphism (SNP) used to assign the alleles of the *Populus tremula* × *Populus alba* (clone 717I-B4) hybrid.
(b) Analysis of potential off-target sites for the UGT71L1 gRNA. All selected off-target sites have four or five mismatches indicated by lower case (two or fewer mismatches within 12 nt upstream of PAM). PAM sequence is indicated in bold letters.

precision of this approach in our system. Likewise, Elorriaga *et al.* (2018) detected no off-target mutations for several gRNAs in over 50 transgenic events in the same *P. tremula* × *P. alba* hybrid used here. Overall, in our experiments approximately 40% of all transgenic events led to mutations, corroborating previous demonstrations of the power of CRISPR for testing gene function in poplar (Elorriaga *et al.*, 2018).

Because UGT71L1 was shown to be necessary for salicinoid biosynthesis, and because the recombinant enzyme has a clear preference for salicyl benzoate compared with other phenolic compounds, our work suggests that salicyl benzoate and its glucoside are likely intermediates in salicinoid biosynthesis. We thus propose that the larger salicinoids are synthesized from UGT71L1-derived salicyl benzoate glucoside. This is then converted to salicortin by enzymes that ultimately oxidize the benzoyl moiety and generate the HCH group (Figure 8). Subsequently, other acyl transferases could add the additional benzoyl group to generate tremulacin. Although in many secondary pathways the addition of glucose occurs late in biosynthesis, in some pathways glycosylation is an early step, occurring prior to further oxidation of other components. This is the case for the seco-iridoid pathway, where dihydroxyloganetic acid is first glycosylated, and the resulting glucoside, loganic acid, is the substrate for two sequential P450-catalyzed enzymatic steps to secologanin (Miettinen *et al.*, 2014). Salicinoid biosynthesis may therefore involve a similar sequence of reactions. An early glycosylation step in the pathway would be consistent with the predicted instability of a free HCH moiety (Clausen *et al.*, 1990).

The proposed involvement of salicyl benzoate in salicinoid synthesis builds on a model of the pathway proposed by Babst *et al.* (2010), who first considered benzyl benzoate

© 2019 The Authors

as a potential intermediate. Salicyl benzoate could theoretically be produced via either the SABT or BEBT enzymes. SABT can generate salicyl benzoate directly by transfer of a benzyl group from benzoyl-CoA to salicyl alcohol (Chedgy *et al.*, 2015). However, the biosynthesis of most salicinoids by this route would not be consistent with previous *in vivo*-labeling experiments by Zenk (1967) and Babst *et al.* (2010), which indicate that salicyl alcohol is efficiently incorporated into salicin, but not into salicortin and larger salicinoids. An alternate route could proceed via BEBT to first generate benzyl benzoate. This could then be 2-hydroxylated to produce salicyl benzoate (Figure 8). In this scenario, free salicyl alcohol would not be involved in salicortin synthesis. However, in this case the role of SABT in salicinoid biosynthesis is not immediately apparent.

It is interesting to note that knocking out only the UGT71L1 gene was sufficient to prevent the synthesis of tremulacin, salicortin and tremuloidin, at least in root tissue. This makes the role of the second glycosyltransferase, UGT78M1, unclear, as it was not required for the synthesis of these salicinoids in roots. This apparent lack of redundancy and overlap in *in vivo* functions with UGT71L1, despite very similar *in vitro* activities, was surprising; it could suggest that UGT78M1 may be specialized for another aspect of phenolic or salicinoid biosynthesis, or that this gene is required for salicinoid synthesis in other tissues or cells. We found no indications of organ-specific patterns of expression in our *in silico* gene expression analysis (Figure 2). However, we did not analyze specific cell types, and the tissue-level localization of the

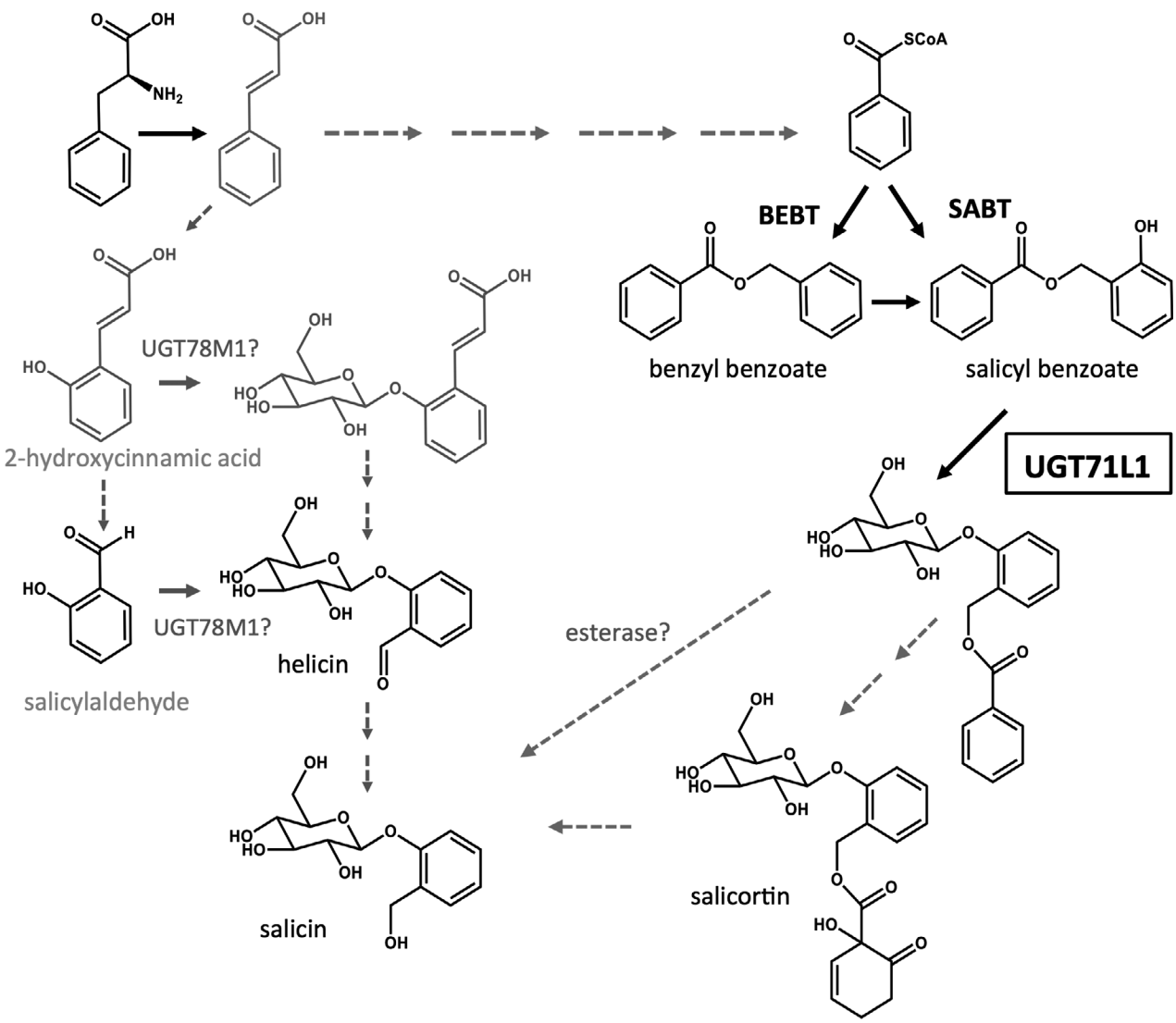


Figure 8. Position of UGT71L1 in proposed pathway leading to salicortin and salicin.
The function of UGT78M1 has not been demonstrated *in vivo* and is hypothesized only. Gray colors indicate predicted intermediates and reactions; black indicates demonstrated enzymes or reaction products.

© 2019 The Authors

salicinoids is not known. Defining the role of UGT78M1 will require the generation of knockouts for this gene as well.

Nevertheless, our data demonstrate that salicin biosynthesis is at least partially distinct from that of the other salicinoids. Tremulacin and salicortin were completely eliminated in the UGT71L1 knockout hairy roots, but salicin was only partially reduced in comparison to most of the control lines. This leaves open the possibility that UGT78M1 is important in the biosynthetic pathway for salicin, perhaps taking alternative substrates. While salicyl benzoate is by far the preferred substrate, the biological significance of the observed activity with salicylaldehyde or 2-hydroxycinnamic acid cannot be excluded. A parallel pathway to salicin could thus potentially proceed via UGT78M1-catalyzed glycosylation of these or other unknown intermediates (Figure 8). However, while speculative, a separate pathway would be consistent with the *in vivo* labeling work by Zenk (1967) and Babst *et al.* (2010), which showed that labeled salicylaldehyde is incorporated into salicin, but not into salicortin. Babst *et al.* (2010) also showed that benzoic acid and benzaldehyde are incorporated into salicortin but not salicin, again suggesting a separate pathway for these compounds. Therefore, regardless of the *in vivo* function of UGT78M1, all available data are consistent with the idea that there are distinct biosynthetic routes to salicin and salicortin (Figure 8).

On the other hand, our data also suggest a partial connection between the salicin and salicortin biosynthetic pathways, as eliminating UGT71L1 in hairy roots partially reduced salicin content in hairy roots. One explanation for this reduction is feedback inhibition from common pathway intermediates, or other metabolic connections. Alternatively, it is possible that some salicin is generated directly from salicortin or salicyl benzoate glucoside, via an esterase that removes the benzoate or the benzoate-derived HCH group (Figure 8). The susceptibility of salicortin to non-specific esterase activities that converts it to salicin has been shown (Julkunen-Tiitto and Meier, 1992). Conversion of salicortin to salicin would define a UGT71L1-dependent route for salicin, in addition to the UGT71L1-independent route mentioned above. Redundancy in benzenoid pathways and their organization as a metabolic grid rather than a strictly linear pathway has been documented for petunia (Widhalm and Dudareva, 2015). Whether this holds for salicinoid biosynthesis remains to be established.

The highly similar substrate preferences of both UGT71L1 and UGT78M1 are remarkable, given they only share 28% amino acid identity and cluster in distinct areas of the UGT phylogeny. UGT71L1 groups with UGTs that use benzoates as sugar acceptors, though typically with less specificity than we observed for UGT71L1 (Lim *et al.*, 2002). By contrast, UGT78M1 is situated within a clade that contains mostly flavonoid 3-O-glucosyltransferases (Figure 3). Thus, we had previously investigated UGT78M1 as a potential flavonoid UGT important for proanthocyanidin biosynthesis (Veljanovski and Constabel, 2013). However, UGT78M1 does not accept catechin or epicatechin, or any major flavonoids and simple phenolics tested, and its *in vitro* activity was unknown until the current work. This apparent convergent evolution of two phylogenetically distant UGT enzymes towards very similar substrate preferences is intriguing. Our modeling data suggest that within each polypeptide, different amino acid residues were recruited to stabilize salicyl benzoate in the enzyme binding pocket, while the catalytic dyad responsible for glucose transfer is similar in both enzymes. Further interpretation of this convergent evolution will require determination of the *in vivo* role of UGT78M1.

To summarize, this work identified two poplar UGT enzymes, both with an *in vitro* preference for salicyl benzoate as a glucose acceptor. UGT71L1 is required for the biosynthesis of salicortin, tremulacin and tremuloidin, but the specific physiological role of UGT78M1 remains unresolved. The implication of salicyl benzoate as a potential intermediate and the identification of one essential enzyme in salicinoid biosynthesis should accelerate progress on the identification and characterization of additional genes and enzymes.

EXPERIMENTAL PROCEDURES

Co-expression and in silico analysis of gene expression

Co-expression data were derived from an RNA-seq dataset of developing leaves from 191 natural accessions of *P. trichocarpa* grown at the University of British Columbia, Vancouver, BC, Canada (McKown *et al.*, 2019). Sequence data are available from NCBI BioProject PRJNA300564. The first uncurled leaves were collected from each accession from a clone bank consisting of coppiced trees. At least two replicate leaves per accession were used for RNA extraction and Illumina RNA-seq library production. Samples were sequenced on an Illumina HiSeq2000 at the Michael Smith Genome Science Centre (Vancouver, Canada) with 75-base paired end reads to a depth of 30 Gb per sample. Cleaned sequence reads from a total of 389 datasets were mapped to the *P. trichocarpa* reference genome (v. 3) using Tophat, and RNA abundance levels of 41 355 primary gene transcripts were determined using Cufflinks. Resulting fragments per kilobase of transcript per million mapped reads (FPKM) values were log-transformed [$\log_2(FPKM + 1)$] and mean-centered across samples for each gene. Based on these transcript abundance data and the target genes *PtSABT* (Potri.013G074500) and *PtBEBT* (Potri.019G043600), pairwise Pearson correlation coefficients (R) were calculated for all other genes in the dataset. Co-expressed genes were retained at an R-value threshold of 0.65.

In silico expression of individual candidate genes was investigated using tissue expression data within the Popgenie poplar database (exAtlas; http://popgenie.org). Data were extracted from *P. tremula* expression atlas (Sundell *et al.*, 2015) using the

© 2019 The Authors

identifiers Potra169534g27450 for UGT78M1, Potra002542g19139 for UGT71L1, and Potra003463g21747 for SABT.

Phylogenetic analysis

Protein sequences of *A. thaliana* and *P. trichocarpa* UGTs were retrieved from Phytozome 10, and other characterized UGT sequences were retrieved using the GenBank accessions noted in the respective publications. Sequence alignment was performed using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/). The alignment output was imported into MEGA5 (Tamura *et al.*, 2011) for phylogenetic tree reconstruction using the maximum-likelihood method. Bootstrap values were calculated from 1000 phylogenetic constructions and mapped back onto the original tree. The graphical representation of the phylogenetic tree was generated with FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Populus gene models were obtained from Phytozome (v3.0): PtUGT84A17 (Potri.009G095100), PtUGT78M1 (Potri.006G171100), PtUGT78L1 (Potri.009G133300), PtUGT71L1 (Potri.016G014500). Arabidopsis sequences were downloaded from TAIR10: AtUGT84A1 (At4g15480), AtUGT84A2 (At3g21560), AtUGT84A3 (At4g15490), AtUGT84A4 (At4g15500), AtUGT74B1 (At1g24100), AtUGT74F1 (At2g43840), AtUGT74F2 (At2g43820), AtUGT75B1 (At1g05560), AtUGT75B2 (At1g05530), AtUGT78D1 (At1g30530), AtUGT78D2 (At5g17050), AtUGT73B3 (At4g34131), AtUGT73B4 (At2g15490), AtUGT73C5 (At2g36800), AtUGT73C6 (At2g36790), AtUGT89A2 (At5g03490), AtUGT89B1 (At1g73880), AtUGT72B1 (At4g01070), AtUGT72E1 (At3g50740), AtUGT72E2 (At5g66690), AtUGT72E3 (At5g26310), AtUGT71B1 (At3g21750), AtUGT71C1 (At2g29750), AtUGT71C4 (At1g07250). GenBank accession numbers for the other sequences are: *Aralia cordata* AcGaT (BAD06514); *Medicago truncatula* MtUGT71G1 (AAW56092), MtUGT78G1 (ABI94025); *Nicotiana tabacum* NtGT1a (BAB60721), NtGT3 (BAB88934), NtSAGT (AF190634); *Punica granatum* PgUGT84A23 (KT159805), PgUGT84A24 (KT159807); *Petunia hybrida* Ph3GT (AAD55985); *Solanum melongena* Sm3GT (Q43641); *Vitis vinifera* VvGT1 (AAB81683), VvgGT1 (JN164679), VvgGT2 (JN164680), VvgGT3 (JN164681).

Heterologous expression and purification of UGTs

The full-length open reading frame of UGT71L1 was synthesized with codons optimized for expression in *E. coli* (Genescript) and cloned in frame into the BamHI/SalI site of the pQE30 vector (Qiagen). Cloning of UGT78M1 was described previously (Veljanovski and Constabel, 2013). Both UGTs were expressed in *E. coli* (M15rep4) induced for 4 h with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) at 30°C. The resulting recombinant proteins contained an N-terminal His-tag to enable purification by metal affinity chromatography. For recombinant protein expression, harvested cells were resuspended in 50 mM potassium phosphate (KPi) buffer (pH 7.5) and ultrasonicated to lyse cells. The crude extract was centrifuged for 5 min at 20 000 ***g***, and the supernatant containing soluble recombinant UGT was purified by Ni-NTA agarose (Qiagen) according to the manufacturer's instructions. Purity was examined by SDS–PAGE and Western blotting. Protein concentration was estimated using the Bradford assay with bovine serum albumin as a standard. Average yields of 1.5 mg UGT78M1 and 3 mg UGT71L1 per liter of *E. coli* culture were obtained.

SDS–PAGE and Western blotting

Recombinant UGTs were separated on 12% SDS–PAGE gels and transferred onto a 0.2-μm PVDF membrane (Bio-Rad). Membranes were blocked with 5% non-fat dry milk in Tween Tris-buffered saline (TTBS) for 2 h, and incubated with the anti-His monoclonal mouse antibody (1:3000; GE Healthcare Life Science) in blocking solution overnight. The membrane was incubated in TTBS for with goat anti-mouse horseradish peroxidase-coupled anti-IgG antibody in blocking solution (1:3000; Bio-Rad) for 1 h. His-tagged UGTs were visualized using the DAB substrate kit as described by the manufacturer (Thermo Scientific).

Recombinant UGT enzyme assays

Standard assays contained 5 μg purified recombinant UGT, 50 μM acceptor substrate and 5 mM UDP-sugar in 50 mM KPO_4 buffer (pH 7.5) in a reaction volume of 200 μl. Assays with salicyl benzoate were carried out in 50 mM acetate buffer (pH 5.1). Reactions were incubated for 30–120 min, depending on the substrate, at 30°C. The reactions were stopped with the addition of 20 μl 30% (v/v) TCA, precipitated proteins removed by centrifugation (5 min at 18 000 ***g***), and the supernatant analysed for product formation by HPLC. Enzyme concentration and incubation parameters were chosen so that the reaction velocity was linear during the incubation time for all kinetic measurements. Negative controls included running assays with boiled UGT enzyme preparations and a recombinant shikimate dehydrogenase preparation. For determination of the pH optimum, acetate buffer covered the range pH 3.6–5.6, phosphate buffer covered the range of pH 5.8–8.0, and glycine-NaOH buffer covered the range from pH 8.6 to pH 10.6. To determine temperature and pH optimum, the recombinant protein was pre-incubated for 10 min in buffer under the condition to be tested, and the reaction started by adding 50 μM 2,4,5-trichlorophenol and 5 mM UDP-glucose. Kinetic properties (K_m and V_{max} values) for each enzyme were determined at optimal pH (pH 7.5) at 30°C with saturating concentrations of UDP-glucose and varying concentrations of the sugar acceptor (5–5000 μM). Enzyme products were detected and quantified by reversed-phase-HPLC (System Gold, Beckman Coulter) using a Kinetex C18 column (50 × 4.6 mm, 5 μm; Phenomenex, Torrance, CA, USA) with 0.4% formic acid (solvent A) and 99.6% acetonitrile/0.4% formic acid (solvent B). Glucose conjugates were separated from their aglycones using a linear gradient going from 10% B in A to 80% B in A over 5.5 min at a flow rate of 2 ml min^{-1}. Total run time was 15 min.

Synthesis of salicyl benzoate

Salicyl alcohol was converted to the corresponding benzyl bromide by the method of Nieddu *et al.* (2008). Cyanuric chloride (2.45 g, 13.3 mmol) was added to 1.5 ml of DMF, and the mixture was stirred for 1.5 h at room temperature. A further 1.0 ml of DMF was added and the mixture was heated to 35°C for 1 h, resulting in the complete disappearance of cyanuric chloride by thin-layer chromatography (TLC). Distilled dichloromethane (20 ml) was then added, followed by NaBr (24.2 mmol). The resulting mixture was stirred at room temperature for 14 h, after which 2-hydroxybenzyl alcohol (salicyl alcohol; 1.5 g, 12.1 mmol) was added to afford a yellow suspension. This was stirred for a further 5 h at room temperature, and then filtered through a pad of Celite, rinsing with cold dichloromethane. The filtrate was washed with water (1 × 16 ml), 1 N HCl (2 × 8 ml) and brine (2 × 8 ml), before being dried over sodium sulfate at 0°C. Filtration and concentration *in vacuo* provided 2.64 g of an orange oil containing the desired intermediate 2-(bromomethyl)phenol along with residual DMF. This was used crude in the following step.

Coupling to benzoic acid was then carried out according to the method of Lee and Choi (1998). Cesium fluoride (7.6 g, 50 mmol) was added to a suspension of Celite 545 (5 g) in distilled water

© 2019 The Authors

(150 ml). The mixture was stirred for 20 min at room temperature, and then water was removed *in vacuo* at 60°C. The resulting white residue was triturated with acetonitrile (100 ml), filtered, and washed with additional acetonitrile (50 ml). Drying in a desiccator for 2 days afforded 12 g of CsF-Celite reagent for use in the coupling step. To effect coupling between salicyl bromide and benzoic acid, a flask containing the crude 2-(bromomethyl)phenol described above (2.64 g, ~12 mmol) was charged with benzoic acid (369 mg, 3.0 mmol), the CsF-Celite reagent (1.08 g) and acetonitrile (30 ml). The resulting yellow suspension was stirred at reflux for 6 h, then cooled to room temperature, filtered and concentrated *in vacuo*. Purification by flash column chromatography (SiO_2, eluting with 15:1 hexanes:EtOAc) afforded 340 mg of a mixture of benzoic acid and the desired benzoate product. This was dissolved in dichloromethane (70 ml) and washed with saturated $NaHCO_3$ (1 × 10 ml then 1 × 20 ml). The organic layer was dried over sodium sulfate, filtered and concentrated *in vacuo* to afford the desired salicyl benzoate as a yellow oil (132 mg, 19% yield).

In an alternative preparation, salicyl alcohol and benzoic acid were coupled directly through EDC coupling. To a solution of benzoic acid (49 mg, 0.40 mmol) and *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (85 mg, 0.44 mmol) was added 2-hydroxybenzyl alcohol (50 mg, 0.40 mmol). The resulting colorless solution was stirred for 14.5 h. The product solution was then washed with water and saturated $NaHCO_3$, then dried over sodium sulfate and concentrated *in vacuo*. Purification by flash column chromatography (SiO_2, eluting with a gradient of 20:1 to 10:1 hexanes:EtOAc) afforded 25 mg of the desired salicyl benzoate (27% yield).

Spectral details for the product from both preparations were consistent with those reported for the known salicyl benzoate (Lebrasseur *et al.*, 2007).

^{1}H-NMR (300 MHz, $CDCl_3$): δ 8.09–8.04 (m, 3H), 7.58 (tt, J = 7.4, 1.3 Hz, 1H), 7.47–7.40 (m, 2H), 7.36 (dd, J = 7.5, 1.7 Hz, 1H), 7.29 (ddd, J = 8.1, 7.4, 1.7 Hz, 1H), 6.97 (dd, J = 8.3, 1.0 Hz, 1H), 6.93 (td, J = 7.4, 1.2 Hz, 1H), 5.38 (s, 2H); ^{13}C-NMR (75 MHz, $CDCl_3$): δ 168.9, 155.9, 133.7, 132.3, 131.3, 130.1, 129.4, 128.6, 121.9, 120.6, 118.0, 63.9 ppm.

LC-MS identification of enzyme reactions products

For LC/MS analysis, methanolic extracts were separated on a Kinetic C18 column (100 × 4.6 mm, 2.6 μm; Phenomenex) and an Agilent 1100 series HPLC system (Agilent Technologies) following the elution gradient described previously (James *et al.*, 2017). The LC system was coupled to a Bruker Esquire 6000 ion-trap mass spectrometer (Bruker Daltonics, Bremen, Germany) and mass spectra obtained applying the method parameters as described (James *et al.*, 2017). The mass spectrometer was operated in alternating ionization mode in the range m/z 60 to 1000. Products of enzyme assays were identified in the full-scan spectra based on the [M + Na]+-adducts in positive ionization mode: salicylaldehyde-glucoside m/z 307, salicyl benzoate-glucoside m/z 413, 2-hydroxycinnamic acid-glucoside m/z 349, and based on the formate adducts [M + COOH]- in negative ionization mode for salicylaldehyde-glucoside m/z 329, and salicyl benzoate-glucoside m/z 435, and based on the [M-H]- for 2-hydroxycinnamic acid-glucoside m/z 325.

Molecular modeling

For the generation of protein homology models of UGT78M1 and UGT78L1, the program YASARA (Yet Another Scientific Artificial Reality Application, version 16.4.16; Krieger *et al.*, 2009; Krieger and Vriend, 2014, 2015) was used. For both proteins, 10 suitable template structures from the PDB were identified for homology modeling. Based on alternative alignments, YASARA created a total of 96 and 76 models for UGT78M1 and UGT78L1, respectively. All models were analyzed with PROCHECK (Laskowski *et al.*, 1993) for stereo-chemical quality and using PROSA II for native folding (Sippl, 1990, 1993). From these tests, the best model for UGT78M1 was based on the Arabidopsis UGT72B1 structure (Brazier-Hicks *et al.*, 2007; PDB code 2VCH), which was 39.3% identical and 58.6% similar (BLOSUM62) to UGT78M1 at the sequence level. For UGT78L1, the best model was based on Medicago MtUGT78G1 (Modolo *et al.*, 2009; PDB code 3HBJ), with a sequence identity of 52.4% and similarity of 69.7%. For these models, combined Prosa scores of −12.2 (UGT78M1) and −13.36 (UGT78L1) were obtained. For UGT78M1, PROCHECK analysis confirmed that 90.9% of amino acids are in most favored areas and 9.1% are in additionally allowed areas, while for UGT78L1 90.9% of amino acids are in most favored areas and 8.4% in related areas, with one outlier. In both models, the co-crystallized UDP was automatically taken from the template X-ray structures into the protein model. Subsequently, keeping the position of the UDP fixed, substrates were docked with the docking module of Molecular Operating Environment (MOE; https://www.chemcomp.com/). All docking poses were manually inspected for appropriate binding and position of the hydroxyl group to be glycosylated with respect to the catalytic diad.

Construction of the CRISPR/Cas9 expression vector and hairy root transformation

Potential target gRNA sequences for UGT71L1 and UGT78M1 in the *P. tremula* × *P. alba* hybrid (INRA 717-1B4) were obtained from the sPta 717 Variant Database (version 1.1) from AspenDB (http://aspendb.uga.edu/s717). Targets with the lowest off-target potential and near the 5′-end of the coding sequences were selected. CRISPR constructs were prepared using the vectors previously developed by Jacobs *et al.* (2015) and obtained from Addgene (Addgene.org), with minor modifications as described here. P201G was double-digested with SpeI and ApaI and purified by gel extraction. MtU6 and scaffold DNA was amplified from the pUC gRNA shuttle vector using the primers 35S*Spe*I_MtU6F/MtU6R and ScaffoldF/*Apa*I_ScaffoldR, creating 20 nt overhangs with backbone to enable DNA assembly. Linearized plasmid, purified MtU6, scaffold and the single-stranded sgRNA with 20 nt overhangs on both ends were combined and incubated for 1 h at 50°C using the NEBuilder DNA Assembly Kit (New England BioLabs). The reaction products (1 μl) were moved into *E. coli* XL1-Blue via electroporation and colonies were grown on LB Kan50. Purified plasmid was sequenced in order to confirm correct assembly. Constructs were electroporated into *Agrobacterium rhizogenes* strain ARqua1 (Quandt *et al.*, 1993), and positive transformants were identified by colony PCR with 35SSpeI_MtU6_F and ApaI_Scaffold_R (Table S3). Hairy roots were generated on *in vitro* cultured *P. tremula* × *P. alba* clone 7171-B4 as described previously (Yoshida *et al.*, 2015).

DNA extraction and mutation analysis

Approximately 100 mg of fresh frozen hairy root tissue was homogenized in 0.8 ml of CTAB buffer [100 mM Tris-HCl pH 8.0, 1.4 M NaCl, 20 mM EDTA, 2% (w/v) cetyltrimethyl ammonium bromide (CTAB), 1% (w/v) polyvinylpyrrolidone-40] in 2-ml cryotubes with four steel beads for 2 × 45 sec at 5000 rpm using a Precellys tissue homogenizer (Bertin Technologies, Montigny-le-Bretonneux, France). After incubation at 65°C for 30 min and centrifugation for 2 min at 12 000 ***g***, the supernatant was transferred to a new tube. The supernatant was vigorously extracted

© 2019 The Authors

with 0.8 ml of chloroform:isoamyl alcohol (96:4 v/v), and the tubes were incubated at room temperature (5 min). Following centrifugation, the aqueous phase was re-extracted with chloroform:isoamyl alcohol. Genomic DNA was precipitated from the aqueous phase with 600 μl of ice-cold isopropanol followed by centrifugation at 12 000 ***g*** and 4°C for 10 min. The resulting pellet of genomic DNA was washed with 0.5 ml of cold 70% ethanol, air-dried for 30 min and re-suspended in 50 μl of Tris-EDTA buffer.

For PCR amplification, 1–2 μl of the genomic DNA preparation was used as template with primers flanking the target site (Table S3). Amplicons were designed to include SNPs as markers for each allele, amplified from each transgenic line, and purified using EZ-10 Spin Column PCR Product Purification Kit (BioBasic, Markham, Canada) and Sanger-sequenced (Eurofins, Toronto, Canada). Resulting sequences were aligned with wild-type sequences with ClustalW in BioEdit and checked manually for mutations. If sequence files showed double peaks near the target site, they were further analyzed using the TIDE web tool (http://tide.nki.nl) to determine if mutations are biallelic or heterozygous (Brinkman *et al.*, 2014). If sequencing results suggested a biallelic or heterozygous mutation, the PCR product was cloned using pUCM-T Cloning Vector Kit (BioBasic) and moved into *E. coli* (XL1-Blue). Up to 10 colonies were sequenced to characterize both mutations. Potential off-target sites were identified by BLAST-searching the gRNA sequences against the *P. tremula* × *P. alba* clone 7171-B4 genome using the Probe Search function in the sPta 717 Variant Database. Hits were considered as potential off-target sites if they were within the coding region of a gene, contained the NGG PAM region, and had two or fewer mismatches within the core region of the gRNA (12 nt upstream of PAM). Verification that no off-target mutations occurred was carried out by PCR-amplifying and sequencing these regions.

Extraction and UPLC-MS analysis of salicinoids from hairy root cultures

Fifty micrograms of freeze-dried hairy roots was extracted in methanol as described (Ma *et al.*, 2018). The resulting extract was dried in a Speed-Vac and re-suspended in 50% methanol to a concentration of 80 mg tissue DW ml^{-1} for UPLC. The samples were filtered using a syringe filter (4 mm, 0.2 μm PTFE; Phenomenex), and 2.0 μl of each sample was analyzed using a Waters Acquity UPLC System coupled to an Acquity PDA eLambda Detector and an Acquity QDa single quadrupole mass spectrometer (Waters, Milford, MA, USA). Separation was carried out with an Acquity BEH C18 (2.1 mm × 50 mm, 1.7 μm) and a column manager to maintain the column at 30°C. The auto-sampler temperature was 10°C. A two-solvent gradient consisting of solvent A (ddH_2O with 0.1% formic acid) to solvent B (acetonitrile with 0.1% formic acid) at a flow rate of 0.5 ml min^{-1} was as follows: 0–0.1% B (0–0.5 min), 0.1–20% B (0.5–5 min), 20–50% B (5–8 min), 50–90% B (8–9 min), 90% B column wash (9–10 min), 90–0.1% B (10–10.1 min), and 0.1% B for column equilibration (10.1–11 min). The UV detector range was 210–820 nm, and UV data were processed using the 269–271 nm wavelength range. The MS conditions were as follows: capillary voltage 0.8 kV, probe temperature 600°C, source temperature 120°C, and cone gas (N_2) flow. MS analysis was performed in negative ionization mode. The MS full scan range was 50–1000 m/z. Compound-specific MS data were collected using SIR methods that were optimized for each investigated compound and its molecular ion. The SIR method targeted masses, and optimized cone voltages were as follows: salicin (285 m/z, 18 V), salicortin (423 m/z, 20 V), tremuloidin (389 m/z, 18 V), tremulacin (527 m/z, 18 V) and catechin (289 m/z, 16 V). UV and MS data were processed using TargetLynx (version 4.1). Standards for salicortin, tremulacin and tremuloidin were obtained from Dr Richard Lindroth, University of Wisconsin, and salicin was available from Sigma-Aldrich. Calibration curves for each target compound were generated using serial dilutions of salicin, salicortin, tremuloidin and tremulacin (1000–0.5 ng μl^{-1}). The molecular ion for salicin (m/z 285) showed a linear response up to 50 ng, tremulacin (m/z 527) up to 100 ng, and salicortin (m/z 423) and tremuloidin (m/z 389) up to 200 ng. A mix of catechin and tremulacin was used as a quality control sample to monitor the stability and ionization efficiency of the mass spectrometer throughout the analysis. UV signals at 270 nm gave a linear response over the entire concentration range up to 2 mg ml^{-1} (Table S2), and this was used to quantify salicortin and tremulacin given their high abundance.

ACKNOWLEDGEMENTS

The generous support of the Natural Science and Engineering Council of Canada (NSERC) in the form of Discovery Grants and the UVic CREATE Training Program 'Forests and Climate Change', as well as the Max-Planck Society is gratefully acknowledged. The authors thank Richard Lindroth for the salicinoid standards, Gerry Gourlay for help with statistics, and Brad Binges and the Centre for Forest Biology for greenhouse support.

AUTHOR CONTRIBUTIONS

CF planned research, carried out experiments, analyzed the data, and drafted the manuscript. OC, LHY, FA, EMP, HG, MR and WB participated in some experiments and analyzed data. JW and JE contributed to planning and supervising the research, and to editing the manuscript. CPC devised research strategy, supervised research, and co-wrote the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

RNAseq data related to this work are available from NCBI BioProject Accession PRJNA300564.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Purification of UGT71L1 and UGT78M1.

Figure S2. Temperature stability and pH optimum for both poplar UGTs.

Figure S3. Active site of UGT71L1 with bound UDP-glucose.

Figure S4. Active site of UGT78M1 with bound UDP-glucose.

Figure S5. CRISPR/Cas9 mediated gene editing of UGT78M1 in transgenic poplar hairy root cultures.

Table S1. Top 10 genes most highly co-expressed with BEBT.

Table S2. Substrates tested as glucose acceptors for UGT71L1 and UGT78M1.

Table S3. Calibration curves for all salicinoids analyzed via LC-MS.

Table S4. List of primers used in this study.

© 2019 The Authors

REFERENCES

Achnine, L., Huhman, D.V., Farag, M.A., Sumner, L.W., Blount, J.W. and Dixon, R.A. (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. *Plant J.* **41**, 875–887.

Babst, B.A., Harding, S.A. and Tsai, C.J. (2010) Biosynthesis of phenolic glycosides from phenylpropanoid and benzenoid precursors in *Populus*. *J. Chem. Ecol.* **36**, 286–297.

Babst, B.A., Chen, H.Y., Wang, H.Q., Payyavula, R.S., Thomas, T.P., Harding, S.A. and Tsai, C.J. (2014) Stress-responsive hydroxycinnamate glycosyltransferase modulates phenylpropanoid metabolism in *Populus*. *J. Exp. Bot.* **65**, 4191–4200.

Boeckler, G.A., Gershenzon, J. and Unsicker, S.B. (2011) Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry*, **72**, 1497–1509.

Boeckler, G.A., Gershenzon, J. and Unsicker, S.B. (2013) Gypsy moth caterpillar feeding has only a marginal impact on phenolic compounds in old-growth black poplar. *J. Chem. Ecol.* **39**, 1301–1312.

Bowles, D., Lim, E.K., Poppenberger, B. and Vaistij, F.E. (2006) Glycosyltransferases of lipophilic small molecules. *Annu. Rev. Plant Biol.* **57**, 567–597.

Brazier-Hicks, M., Offen, W.A., Gershater, M.C., Revett, T.J., Lim, E.K., Bowles, D.J., Davies, G.J. and Edwards, R. (2007) Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Natl Acad. Sci. USA*, **104**, 20 238–20 243.

Brinkman, E.K., Chen, T., Amendola, M. and van Steensel, B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168.

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S. and Martens, S. (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* **69**, 1030–1042.

Chedgy, R.J., Köllner, T.G. and Constabel, C.P. (2015) Functional characterization of two acyltransferases from *Populus trichocarpa* capable of synthesizing benzyl benzoate and salicyl benzoate, potential intermediates in salicinoid phenolic glycoside biosynthesis. *Phytochemistry*, **113**, 149–159.

Clausen, T.P., Keller, J.W. and Reichardt, P.B. (1990) Aglycone fragmentation accompanies ß-glucosidase catalyzed hydrolysis of salicortin, a naturally-ocurring phenol glycoside. *Tetrahedron Lett.* **31**, 4537–4538.

Diner, B., Berteaux, D., Fyles, J. and Lindroth, R.L. (2009) Behavioral archives link the chemistry and clonal structure of trembling aspen to the food choice of North American porcupine. *Oecologia*, **160**, 687–695.

Donaldson, J.R., Stevens, M.T., Barnhill, H.R. and Lindroth, R.L. (2006) Age-related shifts in leaf chemistry of clonal aspen (*Populus tremuloides*). *J. Chem. Ecol.* **32**, 1415–1429.

Elorriaga, E., Klocko, A.L., Ma, C. and Strauss, S.H. (2018) Variation in mutation spectra among CRISPR/Cas9 mutagenized poplars. *Front. Plant Sci.* **9**, 594.

Gachon, C.M.M., Langlois-Meurinne, M. and Saindrenan, P. (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends Plant Sci.* **10**, 542–549.

He, X.Z., Wang, X.Q. and Dixon, R.A. (2006) Mutational analysis of the Medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (Iso) flavonoid glycosylation. *J. Biol. Chem.* **281**, 34 441–34 447.

Hemming, J.D.C. and Lindroth, R.L. (1995) Intraspecific variation in aspen phytochemistry – effects on performance of gypsy moths and forest tent caterpillars. *Oecologia*, **103**, 79–88.

Hwang, S.Y. and Lindroth, R.L. (1997) Clonal variation in foliar chemistry of aspen: effects on gypsy moths and forest tent caterpillars. *Oecologia*, **111**, 99–108.

Jacobs, T.B., LaFayette, P.R., Schmitz, R.J. and Parrott, W.A. (2015) Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* **15**, 16.

James, A.M., Ma, D.W., Mellway, R. *et al.* (2017) Poplar MYB115 and MYB134 transcription factors regulate proanthocyanidin synthesis and structure. *Plant Physiol.* **174**, 154–171.

Julkunen-Tiitto, R. and Meier, B. (1992) The enzymatic decomposition of salicin and its derivatives obtained from Salicaceae species. *J. Nat. Prod.* **55**, 1204–1212.

Keefover-Ring, K., Ahnlund, M., Abreu, I.N., Jansson, S., Moritz, T. and Albrectsen, B.R. (2014) No evidence of geographical structure of salicinoid chemotypes within *Populus tremula*. *PLoS ONE*, **9**, e107189.

Kelly, M.T. and Curry, J.P. (1991) The influence of phenolic compounds on the suitability of three Salix species as hosts for the willow beetle *Phratora vulgatissima*. *Entomol. Exp. Appl.* **61**, 25–32.

Krieger, E. and Vriend, G. (2014) YASARA View-molecular graphics for all devices-from smartphones to workstations. *Bioinformatics*, **30**, 2981–2982.

Krieger, E. and Vriend, G. (2015) New ways to boost molecular dynamics simulations. *J. Comp. Chem.* **36**, 996–1007.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D. and Karplus, K. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Prot. Struct. Funct. Bioinf.* **77**, 114–122.

Laskowski, R.A., Macarthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK – A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.

Lebrasseur, N., Gagnepain, J., Ozanne-Beaudenon, A., Léger, J.-M. and Quideau, S. (2007) Efficient access to orthoquinols and their [4+2] cyclodimers via SIBX-mediated hydroxylative phenol dearomatization. *J. Appl. Chem.* **72**, 6280–6283.

Lee, J.C. and Choi, Y. (1998) An improved method for preparation of carboxylic esters using CsF-Celite/alkyl halide/CH_3CN combination. *Synth. Comm.* **28**, 2021–2026.

Lim, E.K., Doucet, C.J., Li, Y., Elias, L., Worrall, D., Spencer, S.P., Ross, J. and Bowles, D.J. (2002) The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.* **277**, 586–592.

Lindroth, R.L. and Hwang, S.-Y. (1996) Diversity, redundancy, and multiplicity in chemical defense systems of aspen. In *Phytochemical Diversity and Redundancy in Ecological Interactions* (Romeo, J.T., Saunders, J.A. and Barbosa, P. eds.). New York, NY: Plenum, pp. 25–56.

Ma, D.W., Reichelt, M., Yoshida, K., Gershenzon, J. and Constabel, C.P. (2018) Two R2R3-MYB proteins are broad repressors of flavonoid and phenylpropanoid metabolism in poplar. *Plant J.* **96**, 949–965.

Mackenzie, P.I., Owens, I.S., Burchell, B. *et al.* (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics*, **7**, 255–269.

Martineau, L.C., Herve, J., Muhamad, A., Saleem, A., Harris, C.S., Arnason, J.T. and Haddad, P.S. (2010) Anti-adipogenic activities of *Alnus incana* and *Populus balsamifera* bark extracts, Part I: sites and mechanisms of action. *Planta Med.* **76**, 1439–1446.

McKown, A.D., Klapste, J., Guy, R.D., Corea, O.R.A., Fritsche, S., Ehlting, J., El-Kassaby, Y.A. and Mansfield, S.D. (2019) A role for SPEECHLESS in the integration of leaf stomatal patterning with the growth vs. disease trade-off in poplar. *New Phytol.* **223**, 1888–1903.

Miettinen, K., Dong, L.M., Navrot, N. *et al.* (2014) The seco-iridoid pathway from *Catharanthus roseus*. *Nature Comm.* **5**, 3606.

Modolo, L.V., Li, L.N., Pan, H.Y., Blount, J.W., Dixon, R.A. and Wang, X.Q. (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *J. Mol. Biol.* **392**, 1292–1302.

Nieddu, G., De Luca, L. and Giacomelli, G. (2008) A chemoselective, easy bromination of (hydroxymethyl)phenols. *Synthesis*, 3937–3940.

Osmani, S.A., Bak, S. and Moller, B.L. (2009) Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry*, **70**, 325–347.

Pasteels, J.M., Rowell-Rahier, M., Braekman, J.C. and Dupont, A. (1983) Salicin from host plant as precursor of salicylaldehyde in defensive secretion of Chrysomeline larvae. *Physiol. Entomol.* **8**, 307–314.

Quandt, H.J., Pühler, A. and Broer, I. (1993) Transgenic root-nodules of *Vicia hirsuta* – a fast and efficient system for the study of gene-expression in indeterminate-type nodules. *Mol. Plant Microbe. Interact.* **6**, 699–706.

Rehill, B.J., Whitham, T.G., Martinsen, G.D., Schweitzer, J.A., Bailey, J.K. and Lindroth, R.L. (2006) Developmental trajectories in cottonwood phytochemistry. *J. Chem. Ecol.* **32**, 2269–2285.

© 2019 The Authors

Ross, J., Li, Y., Lim, E. and Bowles, D.J. (2001) Higher plant glycosyltransferases. *Genome Biol.* **2**(3004), 1.

Sippl, M.J. (1990) Calculation of conformational ensembles from potentials of mean force – an approach to the knowledge-based prediction of local structures in globular-proteins. *J. Mol. Biol.* **213**, 859–883.

Sippl, M.J. (1993) Recognition of errors in 3-dimensional structures of proteins. *Proteins Struct. Funct. Genet.* **17**, 355–362.

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y., Sjödin, A., Van de Peer, Y., Jansson, S., Hvidsten, T.R. and Street, N.R. (2015) The plant genome integrative explorer resource: PlantGenIE.org. *New Phytol.* **208**, 1149–1156.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molec. Biol. Evol.* **28**, 2731–2739.

Tsai, C.J., Guo, W., Babst, B. *et al.* (2011) Salicylate metabolism in Populus. *BMC Proc.* **5**(Suppl. 7: I9). https://doi.org/10.1186/1753-6561-5-S7-I9

Veljanovski, V. and Constabel, C.P. (2013) Molecular cloning and biochemical characterization of two UDP-glycosyltransferases from poplar. *Phytochemistry*, **91**, 148–157.

Widhalm, J.R. and Dudareva, N. (2015) A familiar ring to it: biosynthesis of plant benzoic acids. *Mol. Plant*, **8**, 83–97.

Yonekura-Sakakibara, K. and Hanada, K. (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J.* **66**, 182–193.

Yoshida, K., Ma, D.W. and Constabel, C.P. (2015) The MYB182 protein down-regulates proanthocyanidin and anthocyanin biosynthesis in poplar by repressing both structural and regulatory flavonoid genes. *Plant Physiol.* **167**, 693–710.

Zenk, M.H. (1967) Pathways of salicyl alcohol and salicin formation in *Salix purpurea* L. *Phytochemistry*, **6**, 245–252.

Zhou, X.H., Jacobs, T.B., Xue, L.J., Harding, S.A. and Tsai, C.J. (2015) Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and redundancy. *New Phytol.* **208**, 298–301.

© 2019 The Authors