Complete Biosynthesis of the Anti-Diabetic Plant Metabolite Montbretin A

Sandra Irmisch¹, Sharon Jancsik³, Macaire Man Saint Yuen¹, Lufiani L. Madilao¹ and Joerg Bohlmann¹,²,³∗

¹Michael Smith Laboratories, University of British Columbia, Vancouver BC
²Department of Botany, University of British Columbia, Vancouver BC
³Department of Forest and Conservation Sciences, University of British Columbia, Vancouver BC

Short title: Biosynthesis of Montbretin A

Corresponding author: Joerg Bohlmann, bohlmann@mail.ubc.ca

One sentence summary: Discovery of the final two genes and enzymes for the biosynthesis of an anti-diabetic compound from montbretia enables its metabolic engineering in heterologous hosts.

Author contributions: SI and JB conceived, designed and supervised the research. SI, SJ, LLM, carried out the experimental work. SI and MMSY analyzed the data. SI and JB interpreted the results and wrote the paper. All authors read, edited, and approved the final version of the manuscript.

Funding information: The research was supported with funds to JB from the GlycoNet Networks of Centres of Excellence (GlycoNet NCE) and the Natural Sciences and Engineering Research Council of Canada (NSERC, Discovery Grant). SI was supported by the Alexander-von-Humboldt foundation through a Feodor Lynen Research Fellowship. JB is a Distinguished University Scholar.
Abstract

Diabetes and obesity are affecting human health worldwide. Their occurrence is increasing with lifestyle choices, globalization of food systems, and economic development. The specialized plant metabolite montbretin A (MbA) is being developed as an anti-diabetes and anti-obesity treatment due to its potent and specific inhibition of the human pancreatic α-amylase. MbA is a complex acylated flavonol glycoside formed in small amounts in montbretia (*Crocosmia x crocosmiiflora*) corms during the early summer. The spatial and temporal patterns of MbA accumulation limit its supply for drug development and application. We are exploring MbA biosynthesis to enable metabolic engineering of this rare and valuable compound. Genes and enzymes for the first four steps of MbA biosynthesis, starting from the flavonol precursor myricetin, have recently been identified. Here, we describe the gene discovery and functional characterization of the final two enzymes of MbA biosynthesis. The UDP-glycosyltransferases, CcUGT4 and CcUGT5, catalyze consecutive reactions in the formation of the disaccharide moiety at the 4'-hydroxy position of the MbA flavonol core. CcUGT4 is a flavonol glycoside 4'-O-xylosyltransferase that acts on the second to last intermediate (MbA-XR₂) in the pathway. CcUGT5 is a flavonol glycoside 1,4-rhamnosyltransferase that converts the final intermediate (MbA-R²) to complete the MbA molecule. Both enzymes belong to the UGT family D-clade and are specific for flavonol glycosides and their respective sugar donors. This study concludes the discovery of the MbA biosynthetic pathway and provides the complete set of genes to engineer MbA biosynthesis. We demonstrate successful reconstruction of MbA biosynthesis in *Nicotiana benthamiana*. 
Introduction

Diabetes and obesity are major health challenges of the 21st century. Diabetes alone affects over 422 million people worldwide and is among the top ten leading causes of death with 1.6 million people killed in 2016 (https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death). The prevalent type 2 diabetes mellitus (T2D) is characterized by the body’s inefficient use of insulin, which leads to elevated blood glucose levels with detrimental effects on different organs and increased risk of dying prematurely. T2D, which used to occur mostly in adults, is increasingly affecting children and is linked to lack of physical activity, sugary diets, over-weight and obesity. Measures to control T2D and the pre-diabetic disease stage include healthy diets, physical activity as well as effective nutraceuticals and improved drugs. Drugs that reduce postprandial carbohydrate breakdown are critical for T2D intervention by controlling blood glucose levels. However, drugs that are currently in use can cause gastrointestinal side effects leading to patient non-compliance (Scheen, 1997; Scott and Spencer, 2000; Scheen, 2003).

The specialized plant metabolite montbretin A (MbA) was discovered as an improved treatment option for T2D and has successfully passed efficiency and toxicity studies in animals (Tarling et al., 2008; Yuen et al., 2016). MbA is a complex acylated flavonol glycoside, described as myricetin 3-O-(glucosyl-6'-O-caffeoyl)-1,2-β-D-glucosyl 1,2-α-L-rhamnoside 4'-O-α-L-rhamnosyl 1,4-β-D-xyloside (Figure 1A). MbA functions as a highly specific and efficient inhibitor of the human pancreatic α-amylase (HPA), the first enzyme in the starch degradation chain (Tarling et al., 2008; Williams et al., 2015). However, further development of MbA as a widely available pharmaceutical or nutraceutical is limited by lack of scalable production systems. The only known source of MbA are the corms of the flowering plant montbretia (Crocosmia x crocosmiiflora) (Asada et al., 1988; Tarling et al., 2008), where MbA is produced in small amounts during a short window of time during seasonal corm development in the early summer (Irmisch et al., 2018). Montbretia is native to southern and eastern Africa and is grown for horticultural purposes around the world. The corms serve as the below-ground storage organ and support vegetative propagation. The rare taxonomic occurrence, narrow spatial and temporal patterns of MbA accumulation in montbretia, together with lack of large-scale cultivation, render the plant inadequate for production of MbA as a pharmaceutical or nutraceutical. Due
to its complex chemical structure, large-scale chemical synthesis of MbA is also not feasible. We therefore proposed a synthetic biology approach for MbA production through metabolic engineering of a recombinant plant or microbial system. This approach requires knowledge of a complete set of genes and enzymes of MbA biosynthesis.

In recent work, we showed that MbA biosynthesis occurs almost exclusively in young developing corms of montbretia (Irmisch et al., 2018). The modular MbA biosynthesis involves the formation and assembly of seven building blocks: the flavonol core myricetin, two units of UDP-rhamnose (UDP-Rha), two units of UDP-glucose (UDP-Glc), UDP-xylose (UDP-Xyl) and caffeoyl-CoA (Irmisch et al., 2018). The six individual biosynthetic reactions of the MbA assembly pathway, and their specific sequence as they occur in montbretia corms, have also recently been described (Irmisch et al., 2018; Irmisch et al. 2019b). In addition, the genes and encoded enzymes for the first four of the six biosynthetic steps have been cloned and functionally characterized (Figure 1A) (Irmisch et al., 2018; Irmisch et al., 2019b). Specifically, two different UDP-sugar dependent glycosyltransferases (UGTs; CcUGT1 and CcUGT2), followed by a BAHD-acyltransferase (CcAT), and a third UGT (CcUGT3) are responsible for the step-wise conversion of the flavonol myricetin to myricetin 3-O-rhamnoside (MR), myricetin 3-O-glucosyl 1,2-rhamnoside (MRG), myricetin 3-O-(6'-O-caffeoyl)-glucosyl 1,2-rhamnoside (mini-MbA) to MbA-XR² (myricetin 3-O-(glucosyl-6'-O-caffeoyl)-1,2-glucosyl 1,2-rhamnoside) (Figure 1A).

Transcripts of genes encoding enzymes of MbA biosynthesis show distinct temporal and spatial expression with highest transcript levels in young developing corms in the early summer and low expression in old corms of the previous year, matching the dynamics of MbA accumulation (Irmisch et al., 2018; Irmisch et al., 2019a; Irmisch et al., 2019b). These patterns of differential gene expression have enabled the gene discovery of the first four steps of MbA assembly as well as montbretia genes affecting myricetin biosynthesis. Using CcUGT1, CcUGT2, CcUGT3 and CcAT1 or CcAT2, together with montbretia genes of myricetin biosynthesis, specifically a MYB-transcription factor (CcMYB4), flavonol synthase (CcFLS) and flavonol 3’5’-hydroxylase (CcCYP2), the formation of MbA-XR² could be reconstructed in Nicotiana benthamiana (Irmisch et al., 2019a; Irmisch et al., 2019b).
The formation of the disaccharide moiety at the 4'-hydroxy of myricetin defines the final two steps of MbA biosynthesis. Based on in vitro enzyme assays with corm protein extracts (Irmisch et al., 2019b), we proposed the final two reactions involve two UGT-mediated glycosylation reactions with UDP-Xyl and UDP-Rha, respectively (Figure 1A). Glycosylation of specialized metabolites in plants is typically catalyzed by UGTs of the GT-1 family, which use UDP-activated sugar donors and contain a highly conserved Plant Secondary Product Glycosylation (PSPG) motif (Gachon et al., 2005; Caputi et al., 2012; Lombard et al., 2014). UGTs constitute one of the largest gene families in plants, and this gene family can be divided into 18 different clades (clades A-R) (Ross et al., 2001; Caputi et al., 2012; Wilson and Tian, 2019). The UGT family contains members in several clades that catalyze the glycosylation of flavonoids at various hydroxyl groups (Caputi et al., 2012), as well as UGTs that catalyze glycosylation of the sugar moiety of glycosides (Richman et al., 2005; Frydman et al., 2013; Yonekura-Sakakibara et al., 2014). The vast majority of previously characterized UGTs use UDP-glucose, but some UGTs accepting UDP-galactose, UDP-rhamnose or UDP-xylose have also been reported (Jones et al., 2003; McCue et al., 2007; Shibuya et al., 2010; Yonekura-Sakakibara et al., 2012; Sayama et al., 2012; Itkin et al., 2013).

In our prior work, we identified a gene family of 159 different UGTs in the corm transcriptome of montbretia, including CcUGT1 (UGT77B2), CcUGT2 (UGT709G2) and CcUGT3 (UGT703E1) involved in the formation of MbA-XR$^2$ (Irmisch et al., 2018; Irmisch et al., 2019b). Here we further explored the montbretia UGT gene family for the discovery of the final two UGT enzymes that complete the MbA biosynthesis. We used transcriptome data that cover a time course of corm development for a co-expression guided gene discovery. Through cDNA cloning, enzyme characterization, and pathway reconstruction in N. benthamiana, we identified and functionally characterized CcUGT4 (UGT703H1) as the enzyme that catalyzes the xylosylation of MbA-XR$^2$ to myricetin 3-O-(glucosyl-6'-O-caffeoyl)-1,2-glucosyl 1,2-rhamnoside 4'-O-xyloside (MbA-R$^2$); and CcUGT5 (UGT729A2) as the enzyme that catalyzes the rhamnosylation of MbA-R$^2$ to MbA.

**Results**

**Identification of candidate UGTs by time-course co-expression analysis**
Previously characterized genes for MbA biosynthesis share a distinct temporal expression pattern during the development of young corms (yC). Transcript levels peak during the spring and early summer in developing yC followed by a decline towards autumn, and are low in old corms (oC) of the previous growing season (Irmisch et al., 2018; Irmisch et al., 2019a; Irmisch et al., 2019b). To identify the remaining UGTs required for the completion of MbA biosynthesis, we explored the RNA-Seq library prepared from yC (harvested June 10\textsuperscript{th} 2016) described in Irmisch et al. (2018) and a newly generated set of five RNA-Seq libraries developed from yC harvested at five different time points (June 27\textsuperscript{th}, July 22\textsuperscript{nd}, August 16\textsuperscript{th}, September 12\textsuperscript{th}, and October 6\textsuperscript{th} 2016). Transcriptomes for each of these six time points of yC development were constructed using RNA-Bloom (Nip et al., 2019), predicted peptides were combined and redundancies reduced yielding 40,565 non-redundant (NR) transcript contigs covering translated sequences with an average length of 331 amino acids (aa). Using reciprocal BLASTP searches, 190 UGTs (≥ 250 aa) were identified in the montbretia yC-time course transcriptome (Supplemental Data Set 1). UGT sequences were filtered for sequences with high counts per millions (cpm) at the June 10\textsuperscript{th} time point and at least 5-fold higher transcript abundance at June 10\textsuperscript{th} compared to October 6\textsuperscript{th}, which resulted in the identification of 19 candidate UGTs. These 19 UGTs included the three previously characterized UGTs involved in MbA biosynthesis, CcUGT1, CcUGT2 and CcUGT3 (Irmisch et al., 2018; Irmisch et al., 2019b). To identify UGTs with expression profiles similar to those of known MbA biosynthesis genes, we generated a heatmap of expression data comprising the 19 UGTs and the two characterized CcATs, CcAT1 and CcAT2, involved in MbA biosynthesis (Irmisch et al., 2018). Based on clustering with known MbA biosynthesis genes, we selected six full-length candidate UGT transcripts, UGT703G1 (contig E2.L.3032), UGT703F1 (contig E2.L.4789), UGT709R1 (contig E0.U.334646), UGT703H1 (contig E1.L.26519), UGT703E4 (contig E2.L.7030) and UGT729A2 (contig E0.L.130572) for further characterization (Figure 1B).

We validated these six UGT transcripts for presence and differential expression (DE) in the previously described contrasting transcriptomes that compared yC and oC gene expression (Irmisch et al., 2018). Three of these six candidates, UGT703G1, UGT703F1 and UGT709R1, were present as full-length sequences with 36.8-fold, 9.9-fold and 10.2-fold higher transcript abundance, respectively, in yC compared to oC. Two UGTs, UGT703E4 and UGT703E4, were
non-full length. UGT729A2 was missing a sequence for 84 aa at the predicted UGT C-terminus, and UGT703E4 was only detected as a short fragment encoding for 250 aa. UGT729A2 and UGT703E4, showed 90.7-fold and 2.8-fold higher transcript levels in yC compared to oC, respectively. Unexpectedly, we were initially not able to detect the sixth candidate, UGT703H1, in the yC/oC transcriptome dataset. However, upon closer inspection of the yC/oC transcriptome data, we identified a contig of 3,692 nucleotides (nt) length (DN68292_c0_g1_i1) which covered two separate non-overlapping open reading frames (ORFs) encoding UGTs. The presence of two ORFs on one contig may be due to in silico misassembly. The shorter of the two ORFs matched UGT703H1. We had initially missed this contig in our data analysis, which was trained to only select for the longest ORF on any given contig.

CcUGT4 (UGT703H1) and CcUGT5 (UGT729A2) catalyze the two final reactions of the MbA pathway

The cDNAs covering the full-length ORFs of UGT703G1, UGT703F1, UGT709R1, UGT703H1, UGT703E4 and UGT729A2 were amplified by PCR from yC cDNA template. In a phylogenetic analysis of family-1 UGTs (Irmisch et al., 2018), proteins encoded by UGT703G1, UGT703F1, UGT703E4, UGT703H1 and UGT729A2 clustered with UGT clade D, and UGT709R1 clustered with clade P (Figure 2, Supplemental Table S1). All six proteins possess a glutamine in the C-terminal position of the PSPG-motif. The amino acid in this position has been described to affect sugar donor specificity (Supplemental Figure S1A) (Kubo et al., 2004; Ono et al., 2010).

To assess candidate UGTs for glycosyltransferase activity, we expressed the six cDNAs individually in Escherichia coli, verified protein expression by immunoblot analysis (Supplemental Figure S1), and performed enzyme assays with the recombinant proteins using MbA-XR\(^2\) and UDP-Xyl (step 5 in MbA biosynthesis, see Figure 1A) or MbA-R\(^2\) and UDP-Rha (step 6 in MbA biosynthesis, see Figure 1A) as substrates. Product analysis was done by LC-UV/MS. The initial activity screen was performed with protein extracts of the E. coli expression strains without purification of recombinant UGTs. This screen revealed UGT activity for UGT703H1 and UGT729A2 (Supplemental Figure S1C and D).

When incubated with MbA-XR\(^2\) (peak 1) and UDP-Xyl, protein extracts containing UGT703H1 showed the formation of a single product (peak 2) with m/z
1081.5 identified as MbA-R\textsuperscript{2} based on matching retention time and fragmentation pattern with an authentic standard (Figure 3A and C; Supplemental Figure S2A). The fragmentation pattern of MbA-R\textsuperscript{2} shows the initial loss of Xyl (loss of 132, m/z 949), indicative for the attachment of Xyl to the 4'-hydroxy-position of the flavonol B-ring. Protein extract containing UGT729A2 converted MbA-R\textsuperscript{2} (peak 2) and UDP-Rha into a single product with m/z 1227.5 identified as MbA (peak 3) based on comparison with an authentic standard (Figure 3B and D; Supplemental Figure S2B). The fragmentation pattern of MbA shows the predominant initial loss of the rhamnosyl xyloside (minus 278, m/z 949) or the initial loss of the caffeoyl moiety (minus 162, m/z 1065), followed by the loss of the respective other to form m/z 787 (MRGG). We also detected a small m/z 1081.5 product peak in assays of protein extracts containing UGT703G1 with MbA-XR\textsuperscript{2} and UDP-Xyl, but its retention time did not match that of MbA-R\textsuperscript{2}. No activity was detected in any of the assays with protein extracts containing UGT703F1, UGT709R1, or UGT703E4 or in control assays with protein extract of \textit{E. coli} containing the empty vector (Figure 3, Supplemental Figure S1). Taken together, these activity screens identified UGT703H1 (CcUGT4) and UGT729A2 (CcUGT5) as the enzymes that catalyze the final two glycosylation steps in MbA biosynthesis.

For further characterization, CcUGT4 (UGT703H1) and CcUGT5 (UGT729A2) proteins were Ni-purified. We tested CcUGT4 for substrate specificity with UDP-Xyl as the sugar donor and different sugar acceptors, specifically the MbA pathway precursor and intermediates (myricetin, MR, MRG, mini-MbA, MbA-XR\textsuperscript{2}) as well as quercetin, kaempferol and caffeic acid (Supplemental Table S2). Among intermediates in MbA biosynthesis, in addition to MbA-XR\textsuperscript{2}, CcUGT4 was also active with MRG and mini-MbA but not with myricetin or MR. However, product formation with MRG and mini-MbA was only 0.6% and 5.2%, respectively, relative to product formation with MbA-XR\textsuperscript{2} as substrate (Supplemental Figure S3B and C). No activity was detected with any of the other acceptor substrates tested (Supplemental Table S2). CcUGT4 was specific for UDP-Xyl as the sugar donor and did not accept UDP-Rha or UDP-Glc when MbA-XR\textsuperscript{2} was used as the acceptor (Supplemental Figure S3A). We tested CcUGT5 for substrate specificity with UDP-Rha as the sugar donor and different acceptors, including myricetin, MR, MRG, mini-MbA, MbA-XR\textsuperscript{2}, MbA-R\textsuperscript{2}, MbA-CR\textsuperscript{2}, quercetin 4'-O-glucoside (spiraeoside), quercetin, kaempferol or caffeic acid (Supplemental Table S2). In addition to MbA-XR\textsuperscript{2}, CcUGT5 was active with
MbA-CR² (which is MbA-R² missing the caffeoyl moiety) as a substrate. Here, a single product peak m/z 1065 was observed and identified as MbA-C (MbA missing the caffeoyl moiety) (Supplemental Figure 4A and C). Quercetin 4'-O-glucoside (spireoside) did not serve as a substrate for CcUGT5 (Supplemental Table S2). In addition to UDP-Rha, CcUGT5 also accepted UDP-Xyl but not UDP-Glc as a sugar donor with MbA-R² as acceptor (Supplemental Figure S4B). However, product formation using UDP-Xyl and MbA-R² was only 0.7% relative to product formation with UDP-Rha, suggesting that UDP-Rha was the preferred sugar donor substrate for CcUGT5 (Supplemental Figure S4B and C).

**Transcript expression profiles of CcUGT4 and CcUGT5 support their role in MbA biosynthesis**

To validate and compare CcUGT4 and CcUGT5 transcript expression patterns in yC and oC over a time course of corm development, we measured transcript abundance using reverse transcription quantitative PCR (RT-qPCR) in RNA samples isolated from corms of the June 10th to October 6th 2016 time course (Figure 4). Matching the expression of other genes involved in MbA biosynthesis and matching the profiles of MbA accumulation (Irmisch et al., 2018; Irmisch et al., 2019a; Irmisch et al., 2019b), CcUGT4 and CcUGT5 transcript abundance was low and did not significantly change across all time points in oC, and was significantly higher in almost all of the yC samples compared to oC (Supplemental Table S3). Transcript abundance of CcUGT4 and CcUGT5 was highest in yC harvested in early June, with over 30-fold and 40-fold higher transcript levels, respectively, compared to oC of the same time point. Expression levels of both UGTs continuously dropped from June to October. CcUGT4 and CcUGT5 showed, respectively, 38-fold and 11-fold higher transcript abundance in yC harvested in early June compared to October.

**Reconstitution of MbA biosynthesis in N. benthamiana leaves**

We previously showed that the MbA pathway intermediates up to MbA-XR² could be produced in *N. benthamiana* by transient co-expression of the montbretia myricetin biosynthesis genes, CcMYB4, CcFLS and CcCYP2, together with genes for the first four steps of the MbA pathway, specifically CcUGT1, CcUGT2, CcAT1 and CcUGT3 (Irmisch et al., 2019a; Irmisch et al., 2019b). In addition, this system produced MbB-XR², which contains a coumaroyl moiety instead of the caffeoyl moiety. Here, we
extended the pathway reconstruction for the complete MbA biosynthesis in *N. benthamiana* by additional co-expression of *CcUGT4* and *CcUGT5*. *N. benthamiana* leaves were infiltrated with different combinations of *Agrobacterium tumefaciens* strains containing the above-mentioned genes as 3SSpro-gene constructs. Plants expressing myricetin and MbA-XR² biosynthesis genes served as controls (Irmisch et al., 2019b). Leaves were collected five days after infiltration, and MeOH/H₂O extracts were analyzed by LC-MS and LC/MS-QToF using authentic standards for MbA and intermediates in MbA biosynthesis for metabolite identification. In addition, reference compounds for MbB and intermediates in MbB biosynthesis were produced using MbA biosynthesis enzymes. This was possible as CcAT1 is active with both caffeoyl-CoA and coumaroyl-CoA (Irmisch et al., 2018) (Supplemental Figure S5).

In agreement with our previous work, leaves expressing myricetin and MbA-XR² biosynthesis genes produced MbA-XR² (m/z 949, peak 4 in Figure 5) and MbB-XR² (m/z 933, peak 2 in Figure 5) (Irmisch et al., 2019b) (for accurate masses see Supplemental Table S4). Peak areas corresponding to these products decreased when CcUGT4 alone or CcUGT4 in combination with CcUGT5 were co-expressed with genes for myricetin and MbA-XR² biosynthesis, indicating substrate conversion. Xylosylation of MbA-XR² or MbB-XR² through CcUGT4 would result in m/z 1081.5 or m/z 1065.5, respectively. A peak m/z 1081.5 (peak 10 in Figure 5) matching the retention time and fragmentation pattern of MbA-R² was detected in leaf extracts co-expressing CcUGT4 (Supplemental Figure 6C). Additionally, two m/z 1065.5 peaks were detected and peak 9 was identified as MbB-R² (Figure 5, Supplemental Figure 6A). Additional co-expression of CcUGT5 resulted in depletion of MbA-R² and MbB-R², and products of m/z 1227.5 (peak 16, Figure 5), identified as MbA, and m/z 1211.5 (peak 14, Figure 5), identified as MbB, were detected (Supplemental Figure 6 B and D). Interestingly, these samples also showed a m/z 1065.5 peak (peak 11, Figure 5) identified as MbA-C (MbA and MbB without the caffeoyl or coumaroyl moiety, Supplemental Figure 7), a possible degradation product of MbA or MbB. In all experiments, formation of peaks corresponding to MbA and MbB and intermediates in their biosynthesis was specific to the co-expression of the respective montbretia MbA biosynthesis enzymes. Other peaks of unknown identity were likely due to the modification of pathway intermediates through endogenous *N. benthamiana* enzymes and/or montbretia enzymes acting on *N. benthamiana* flavonoids. For each MbA biosynthesis step introduced into *N. benthamiana* we
quantified the respective MbA and MbB products based on an external MbA standard curve (Supplemental Table S5). When expressing myricetin and MbA-XR2 biosynthesis genes, transiently transformed leaves produced approximately 2.2 mg * g\(^{-1}\) (FW) MbB-XR2 and 42 μg * g\(^{-1}\) (FW) MbA-XR2. By expressing the complete MbA biosynthesis pathway N. benthamiana leaves produced approximately 0.7 mg * g\(^{-1}\) (FW) MbB and about 7 μg * g\(^{-1}\) (FW) MbA.

**Discussion**

The complex acylated flavonol glycoside MbA was reported in 2008 as a promising new anti-diabetes drug (Tarling et al., 2008). MbA acts as an efficient and specific inhibitor of the human pancreatic α-amylase (HPA, \(K_i = 8\text{nM}\)) and was proven in animal studies to reduce blood glucose levels without obvious adverse or toxic effects (Yuen et al., 2016). Phase-I clinical trials are under preparation. However, ultimately drug development and application for the millions of T2D and obesity patients, who could benefit, will be limited by short supply of MbA from its sole natural source, the montbretia corms (Irmisch et al., 2018). The goal of our work is to discover the complete set of genes and enzymes of the MbA biosynthetic pathway and to use these genes for metabolic engineering of MbA production in a heterologous host. Here we described the discovery and characterization of the final two UGTs in the MbA biosynthetic pathway, CcUGT4 (UGT703H1) and CcUGT5 (UGT729A2).

The CcUGT4 and CcUGT5 genes were identified by exploring temporal expression data and co-expression of MbA pathway genes. CcUGT4 catalyzes a 4’-hydroxy xylosylation, converting MbA-XR2 with UDP-Xyl into MbA-R2. The 1,4-rhamnosyl transferase CcUGT5 then converts MbA-R2 with UDP-Rha into the final MbA product. Like other MbA biosynthetic genes, both UGTs show a peak in transcript expression in the early summer, matching profiles of MbA accumulation (Irmisch et al., 2018) and supporting their biological function. Beyond *in vitro* enzyme characterization, we validated functions of CcUGT4 and CcUGT5 by successfully reconstructing the complete biosynthesis of MbA in *N. benthamiana*. Engineered *N. benthamiana* produced small amounts of MbA but predominantly produced MbB, which is not an HPA inhibitor. Thus, as previously reported, additional engineering of the acyl group, which is caffeoyl in MbA and coumaroyl in MbB, will be required to enable *N. benthamiana* as a heterologous production system (Irmisch et al., 2018;...
Irmisch et al., 2019a; Irmisch et al., 2019b). The discovery of genes and enzymes for the final two steps of MbA biosynthesis has filled essential missing components for the reconstruction of MbA biosynthesis in a plant or microbial system to solve the MbA supply challenge.

**Enzyme specificity dictates linear MbA assembly**

MbA is formed by stepwise assembly of its seven building blocks, the core flavonol myricetin, the five sugars from UDP-Glc, UDP-Rha and UDP-Xyl, and the acyl group from caffeoyl-CoA (Irmisch et al., 2018). In the MbA molecule, the myricetin core is decorated with two glycosyl chains: One is the acylated trisaccharide attached to the 3-hydroxy (O3) of the myricetin C-ring, and the other is the disaccharide chain attached to the 4’-hydroxy (O4’) of the myricetin B-ring (Figure 1). Formation of these side chains occurs in a linear sequence (Irmisch et al., 2018; Irmisch et al., 2019b), starting with a 3-hydroxy rhamnosylation of myricetin and a 1,2-glucosylation carried out by CcUGT1 and CcUGT2, respectively. The glucose of the resulting myricetin disaccharide is acylated by either CcAT1 or CcAT2 on O6 to form mini-MbA (Irmisch et al., 2018). CcUGT3 completes the acylated trisaccharide on the 3-hydroxy through a terminal 1,2-glucosylation (Irmisch et al., 2019b). The 4’-hydroxy group of myricetin is glycosylated after completion of the acylated trisaccharide chain (Irmisch et al., 2019b). Here we showed that CcUGT4 and CcUGT5 complete the MbA molecule by catalyzing a 4’-hydroxy xylosylation of the acylated myricetin trisaccharide, MbA-XR², followed by a 1,4-rhamnosylation of the xyloside (MbA-R²) to yield MbA (Figure 1). The linearity of MbA assembly is achieved through substrate specificity of the six biosynthetic enzymes. The acyltransferases CcAT1 and CcAT2 are specific for the myricetin disaccharide (Irmisch et al., 2018), and the next enzyme, CcUGT3, catalyzes the glucosylation of mini-MbA but does not glucosylate the myricetin disaccharide MRG (Irmisch et al., 2019b). Similarly, CcUGT4 and CcUGT5 require the glycosyl chain on O3 for efficient catalysis. A linear assembly pathway has also been described for acylated anthocyanins in lobelia (Lobelia erinus), known as lobelinins, and the steroidal glycoalkaloid α-tomatine in tomato (Solanum lycopersicum) (Itkin et al., 2013; Hsu et al., 2017). In contrast, the glycosylation of the stevioside diterpenoid glycosides in sweetleaf (Stevia rebaudiana) and the triterpenoid glycoside mogroside V in monk fruit (Siraitia grosvenorii) involve UGTs with broader substrate compatibility, and as a consequence, biosynthesis may occur...
through multiple routes in a metabolic grid instead of a linear pathway (Richman et al., 2005; Itkin et al., 2016). The molecular basis for UGTs to either be substrate specific or to accept substrates with varying sizes remains to be elucidated. Recently, the first crystal structure of a glycoside glycosyltransferase (GGT), UGT76G1 from sweetleaf, was determined (Lee et al., 2019; Liu et al., 2020). The crystal structure reveals features by which the enzyme accommodates a two-glucosyl side chain and provides a site for the regio-specific addition of a third sugar molecule (Lee et al., 2019). It is worth noting that the spacious pocket of UGT76G1 for sugar-acceptor binding is also observed to be the reason for the enzyme’s promiscuity (Liu et al., 2020). GGTs in MbA production are rather substrate specific, and future studies on the structure-function relationship of montbretia UGTs could help understand the enzymatic mechanisms of accepting complex glycosylated substrates.

**Clade D UGTs produce the complex glycosylation pattern of MbA**

The majority of known glycosyltransferases act on flavonoids on the O3 (ring C) or O7 (ring A) position (Hofer, 2016). On the B-ring the most frequently hydroxylated position is the 4′-hydroxy group (Hofer, 2016). Flavonoid 4′-O-glycosylation activity is commonly observed as a minor UGT side activity, and few UGTs have been described to catalyze the predominant formation of 4′-O-glucosides (Witte et al., 2009; Kim et al., 2010; Hall et al., 2011; Ruby et al., 2014; Funaki et al., 2015; Song et al., 2015). Montbretia CcUGT4 (UGT703H1) catalyzes a 4′-xylosylation of the myricetin core of a flavonoid glycoside but not of the aglycon. A few other UGTs that specifically act on the core of a glycoside have been described (Jones et al., 2003; Imayama et al., 2004; Ogata et al., 2005; Richman et al., 2005; Yonekura-Sakakibara et al., 2007; Itkin et al., 2016). For example, Arabidopsis (Arabidopsis thaliana) UGT89C1 is a flavonol 7-O-rhamnosyltransferase, specific for 3-O-glycosylated flavonols (Yonekura-Sakakibara et al., 2007). Notably, four of the five UGTs in MbA biosynthesis accept glycosides as substrates. While CcUGT4 glycosylates the myricetin core of the flavonol 3-O-glycoside, CcUGT2, CcUGT3 and CcUGT5 are GGTs mediating a glycosidic chain elongation (Irmisch et al., 2018; Irmisch et al., 2019b).

GGTs accepting flavonoids as substrates cluster in clade A of the UGT superfamily (Bowles et al., 2005), and this clade contains the large majority of known GGTs that act on a variety of substrates, including flavonoid, lignan and lignan.
glycosides (Noguchi et al., 2008; Shibuya et al., 2010; Jung et al., 2014; Di et al., 2015). However, the number of GGTs reported from other clades is increasing, with GGTs reported from clade D, E, H, O and P (Richman et al., 2005; Shibuya et al., 2010; Sayama et al., 2012; Trapero et al., 2012; Itkin et al., 2013; Irmisch et al., 2018; Irmisch et al., 2019b). The three characterized GGTs in MbA biosynthesis fall into clade P (CcUGT2) (Irmisch et al., 2018) and clade D (CcUGT3 (Irmisch et al., 2019b) and CcUGT5 (this work)). CcUGT4 also belongs to clade D (this work), and other UGTs glycosylating the core of a glycoside have been identified from clade D, L and E (Jones et al., 2003; Imayama et al., 2004; Ogata et al., 2005; Richman et al., 2005; Yonekura-Sakakibara et al., 2007; Itkin et al., 2016) (Figure 2). These observations highlight that prediction of UGT function by clade association is not currently possible (Yonekura-Sakakibara and Hanada, 2011). Clade D is one of the largest groups of plant UGTs covering a wide array of different functions (Fukuchi-Mizutani et al., 2003; Moraga et al., 2009; Shibuya et al., 2010; Caputi et al., 2012). It comprises, for example, flavonoid 7-O-glycosyltransferases and contains a subcluster of various triterpene/phytosterol-related UGTs (Sayama et al., 2012). In an analysis of 65 plant genomes, clade D was overrepresented in the monocots, particularly in the order Asparagales which includes montbretia (Wilson and Tian, 2019). In agreement with these numbers, the majority of montbretia UGTs are members of clade D including 56 (or more than one third) of all corm-expressed UGTs (Irmisch et al., 2018). Three of the five UGTs of the MbA biosynthesis belong to clade D and two of these belong to the UGT703 family, a group which appears to be specific for monocots (Irmisch et al., 2018; Irmisch et al., 2019b; Wilson and Tian, 2019). Wilson and Tian (2019) propose that the duplication and subsequent diversification of UGTs in group D may correlate with the expansion of metabolism specific to the monocots, representing an example of either functional convergence with other UGT groups or sub-functionalization within group D. It is tempting to speculate that the expansion in the montbretia UGT D clade is related to the production of complex glycosylated specialized metabolites, specifically the montbretins and crocosmiosides (Asada et al., 1989). The fact that flavonoid disaccharides are common in plants supports the notion that the more complex glycosylation pattern of MbA might have evolved from a more conserved flavonoid disaccharide (here MRG) upon the recruitment of three D clade UGTs.
UGTs in MbA biosynthesis utilize three different UDP-sugars

Glucosides are by far the most common form of sugar-conjugated specialized metabolites, and the majority of known UGTs are glucosyltransferases utilizing UDP-Glc as the sugar donor (Hofer, 2016). In addition to two glucosyltransferases (CcUGT2 and CcUGT3), the biosynthesis of MbA requires two rhamnosyltransferases (CcUGT1 and CcUGT5) and one xylosyltransferase (CcUGT4). CcUGT1, a myricetin 3-O-rhamnosyltransferase (Irmisch et al., 2018), and CcUGT5, a flavon glycoside 1,4-rhamnosyltransferase, catalyze the first and final glycosylations in MbA biosynthesis, respectively, show strong preference for UDP-Rha as the sugar donor. A few flavonoid rhamnosyltransferases utilizing UDP-Rha have been characterized in other plants, mainly catalyzing the formation of flavonoid 3-O- or 7-O-rhamnosides, and rutinosides or neohesperidosides (Jones et al., 2003; McCue et al., 2007; Yonekura-Sakakibara et al., 2007; Frydman et al., 2013; Rojas Rodas et al., 2014; Hsu et al., 2017). For example, Arabidopsis UGT78D1 mediates a 3-hydroxy rhamnosylation of kaempferol or quercetin, and lobelia ABRT2/4 catalyzes a 1,6-rhamnosylation to form delphinidin 3-O-rhamnosyl 1,6-glucoside (delphinidin rutinoside) (Jones et al., 2003; Hsu et al., 2017). CcUGT4 is a myricetin 3-O-glycoside 4'-O-xylosyltransferase with specificity for UDP-Xyl as the sugar donor. Interestingly, CcUGT3 can also utilize UDP-Xyl as a sugar donor in vitro, although in planta UDP-Glc appears to be the more accessible sugar donor (Irmisch et al., 2019b). A handful of UGTs using UDP-Xyl have been described in other plants (Sayama et al., 2012; Itkin et al., 2013). For example, GAME2 (glycoalkaloid metabolism 2) is a branch-chain GGT involved in α-tomatine biosynthesis in tomato, while Arabidopsis UGT79B1 and kiwi (Actinidia deliciosa) F3GGT1 (flavonoid 3-O-glucoside GT 1) catalyze 1,2-xylosylation of anthocyanidin 3-O-glycosides (Montefiori et al., 2011; Yonekura-Sakakibara et al., 2012; Itkin et al., 2013). A few studies investigated the molecular basis of sugar-donor specificity of UGTs, but much more remains to be learned (Osmani et al., 2009; Chen and Li, 2017; Zong et al., 2019). For example, a single amino acid was identified to confer a change in sugar donor specificity in two UGT89A2 isoforms in Arabidopsis (Chen and Li, 2017). Specifically, an isoleucine-to-serine change in position 153 altered the specificity for UDP-Xyl to also accept UDP-Glc. CcUGT4, which is specific for UDP-Xyl, contains a serine in this position. Using the first crystal structure of a rhamnosyltransferase, UGT89C1, Zong et al. (2019) identified histidine in position
357 in the PSPG-motif as a key amino acid for UDP-Rha specificity (Zong et al., 2019). Exchange with glutamine resulted in an enzyme that accepted both UDP-Rha and UDP-Glc. Interestingly, all CcUGTs in MbA biosynthesis possess a glutamine in this position (Supplemental Figure S1). Currently there is no known common feature that defines UDP-sugar-donor specificity. It is possible that sugar-donor specificity evolved independently in different plant lineages following evolution of substrate- and regio-specificity for the sugar acceptor (Noguchi et al., 2009). Structure/function analyses of a larger variety of UGTs will be necessary to understand UGT sugar-donor specificity.

Conclusion

This work concludes the discovery of the complete biosynthetic pathway for the anti-diabetic metabolite MbA from montbretia. We describe the gene discovery and functional characterization of the final two UDP-glycosyltransferases in the MbA biosynthetic pathway. We successfully reconstructed MbA biosynthesis in *N. benthamiana* demonstrating the potential of this plant system as a heterologous production host. The complete set of genes for MbA biosynthesis also enables the synthetic biology of this compound through metabolic engineering in other systems to ensure supply of MbA for drug development and application.

Material and Methods

Plant material

Montbretia (*Crocosmia x crocosmiiflora*) of the variety Emily McKenzie were obtained, propagated and harvested as described in Irmisch et al. (2018). *Nicotiana benthamiana* were grown from seed in potting soil in a controlled environment chamber (day, 26 °C; night, 22 °C; 16 h/8 h light/dark cycle).

Transcriptome sequencing, *de novo* assembly

RNA samples (RNA Integrity Number ≥ 9) prepared separately from yC of the June 27th, July 22nd, August 16th, September 12th and October 6th 2016 time points, each with two biological replicates, were sequenced at the McGill University & Génome Québec Innovation Centre (http://gqinnovationcenter.com). RNA-Seq was performed on the Illumina HiSeq 4000 platform using 100-bp paired-end (PE) strand-specific libraries multiplexed over two lanes, generating approximately 376 million PE reads.
Data for yC of the June 10\textsuperscript{th}, 2016 time point was recently acquired and utilized (Irmisch et al., 2018; SRP108844). Sequence quality was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences were trimmed with BBDuk of the BBTools software suite (v 38.32, sourceforge.net/projects/bbmap/). Sequences were then assembled by RNA-Bloom (version 0.9.8) (Nip et al., 2019) for each individual time point. TransDecoder (version 5.4.0, https://transdecoder.github.io/) was used to predict peptides for each assembly. In the end, predicted peptides for each time point were then amalgamated and clustered at 96% identity using CD-Hit (version 4.6.8) resulting in 40,565 contigs with average length of 331 aa.

**Generation of transcript abundance and heatmap**

Transcript expression data in counts per million (cpm) were extracted from the transcriptome assembly using the voom/limma package in R with quantification results from Salmon (v 0.11.3) with numBootStrap = 100 (Law et al., 2014; Patro et al., 2017). Contigs with less than 100 cpm were discarded. Expression data for the 190 UGTs can be found in the Supplemental Data Set 2. The heatmap was generated with the R package heatmaply.

**Identification of target UGTs**

Putative montbretia UGT sequences were identified in the yC-time course (yC-TC) transcriptome assembly by BLASTP and reciprocal BLASTP search of the montbretia yC-TC translated protein database as described previously (Irmisch et al., 2018). 298 putative UGT sequences could be identified with 190 UGTs larger than 250 amino acids. UGTs were filtered based on their cpm values. Parameters for filtering of candidate UGTs were based on the recovery of known genes in MbA-XR\textsuperscript{2} biosynthesis (> 30 cpm for the June 10\textsuperscript{th} time point and > 5-fold higher expression June 10\textsuperscript{th} compared to October). A heatmap was constructed utilizing the remaining 19 UGTs (including UGT77B2, UGT709G2 and UGT703E1) and CcAT1 and CcAT2. A list of target UGTs was compiled based on clustering of candidates with characterized genes in MbA biosynthesis.

**UGT cDNA cloning and heterologous expression in *Escherichia coli***

Target UGTs were amplified from cDNA prepared from yC of the June 10\textsuperscript{th}, 2016 time point and cloned into the pJET1.2/blunt vector (ThermoFisher Scientific) for sequencing (Supplemental Table S6 for primer information). Complete open reading...
frames of the target UGTs were cloned as *Bsa*I fragments into the pASK-IBA37
vector (IBA-GmbH, Göttingen, Germany). The *E. coli* TOP10 strain (Invitrogen) was
used for heterologous UGT expression. Cultures were grown at 21 °C, induced at an
OD600 = 0.5 with 200 µg/L anhydrotetracycline (Sigma-Aldrich, Germany) and then
placed at 18 °C and grown for another 20 h. Cells were collected by centrifugation
and disrupted by five freeze and thaw cycles in chilled extraction buffer [50 mM Tris-
HCl, pH 7.5, 10 mM MgCl₂, 5 mM dithiothreitol, 10% (v/v) glycerol, 1x Pierce™
protease inhibitor (EDTA-free, ThermoFisher Scientific), 25 U Benzonase Nuclease
(Merck, Germany), 0.2 mg x mL⁻¹ lysozyme]. Cell fragments were removed by
centrifugation at 14,000 x g and the supernatant was desalted into assay buffer (10
mM Tris-HCl, pH = 7.5, 1 mM dithiothreitol, 10% (v/v) glycerol) using Econopac
10DG columns (BioRad, Hercules, CA, USA). For protein purification using Ni-NTA, a
modified extraction buffer was used [50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 5 mM
dithiothreitol, 2% (v/v) glycerol, 150 mM NaCl₂, 20 mM imidazole, 1x Pierce™
protease inhibitor (EDTA-free, ThermoFisher Scientific), 25 U Benzonase Nuclease
(Merck, Germany), and 0.2 mg x mL⁻¹ lysozyme] and the lysate was directly loaded
onto a Ni-NTA agarose column (Qiagen, Hilden, Germany). Protein was eluted with
elution buffer (10 mM Tris-HCl, pH = 7.5, 500 mM imidazole, 1 mM dithiothreitol, 10%
(v/v) glycerol) and desalted into assay buffer using Illuma NAP-5 Columns (GE
Healthcare). Enzyme concentrations were determined using UV absorption at 280
nm. The Monoclonal Anti-polyHistidine–Alkaline Phosphatase antibody (Sigma-
Aldrich) and the 1-Step™ NBT/BCIP Substrate Solution (Thermo Fisher Scientific)
were used to ensure successful heterologous enzyme expression (Supplemental
Figure S1B).

**Enzyme assays with recombinant UGTs**

To test for UGT activity, initial enzyme assays were performed with 100 µL of the
bacterial extract, 50 µM MbA-XR² and 1 mM UDP-Xyl or 50 µM MbA-R² and UDP-
Rha in a Teflon-sealed, screw-capped 1 ml GC glass vial. Unless stated otherwise,
assays were performed in assay buffer in a final volume of 100 µL and incubated at
25 °C. Assays were incubated for 2 h and stopped by placing on ice after the addition
of an equal volume of MeOH. To characterize UGTs and determine enzyme
parameters, UGT4 and UGT5 were Ni-purified and incubated with different sugar
donors and sugar acceptors. For sugar donor specificity 1.7 µg of UGT4 or 4 µg of
UGT5 were incubated with 50 µM MbA-XR² or 50 µM MbA-R², respectively, and 1
mM of either UDP-Glc, UDP-Rha or UDP-Xyl in a final volume of 50 µL for 40 min. Sugar acceptor specificity was tested using 1.7 µg of UGT4 or 4 µg of UGT5, 1 mM UDP-Xyl or UDP-Rha, respectively, and 50 µM of the different substrates (listed in Supplemental Table S2) and incubated for 1.5 h.

For step wise reconstruction of MbA and MbB biosynthesis, CcUGT1, CcUGT2, CcUGT3, CcUGT4, CcUGT5 and CcAT1 were heterologously expressed in E. coli and Ni-purified (Irmisch et al., 2018; Irmisch et al., 2019b). Myricetin (50 µM) was used as the flavonol acceptor and incubated with the respective enzymes (approximately 10 µg each), sugar donors (1mM each) and acyl donors (50 µM, caffeoyl-CoA for MbA or coumaroyl-CoA for MbB) for 2 h at 25 °C. Myricetin, MR, spiraeoside, caffeic acid and UDP-Glc were obtained from Sigma-Aldrich, Germany; kaempferol and quercetin were obtained from TRC-Canada, MRG and UDP-Rha (HPLC-purified) were prepared as described in Irmisch et al., 2018; MbA, mini-MbA, MbA-XR², MbA-R² and MbA-CR² were prepared as described in Tarling et al., 2008 and Williams et al., 2015; UDP-Xyl was obtained from Biosynth-Carbosynth; caffeoyl-CoA and coumaroyl-CoA were obtained from TransMIT GmbH.

**LC-MS analysis**

LC was performed on an Agilent 1100 HPLC (Agilent Technologies GmbH, Waldbronn, Germany) with Agilent ZORBAX SB-C18 column (50 x 4.6 mm, 1.8 µm particle size) (Merck, Darmstadt, 370 Germany) using aqueous formic acid (0.2% v/v) (mobile phase A) and acetonitrile plus formic acid (0.2% v/v) (mobile phase B). Different methods were established to yield peak separation. Method A: elution profile was: 0 - 0.2 min, 95% A; 0.2 - 1 min, 5 - 17% B in A; 1 - 8 min, 17 - 20% B in A; 8 - 9 min 20 - 90% B in A; 9 - 10 min 90% B in A and 10.1 - 11 min 95% A. The flow rate was 1 mL x min⁻¹ at a column temperature of 45 °C. Method B: elution profile was: 0 - 0.2 min, 90% A; 0.2 - 1 min, 10 - 17% B in A; 1 - 8 min, 17 - 20% B in A; 8 - 9 min 20 - 90% B in A; 9 - 10 min 90% B in A and 10.1 - 11 min 95% A. The flow rate was 1 mL x min⁻¹ at a column temperature of 30 °C. Method C: elution profile: 0 - 7 min, 95% A; 7 - 9 min, 5 - 15% B in A; 9 - 20 min, 15 - 18% B in A; 20 - 25 min 18 - 90% B in A; 25 - 27 min 90% B in A and 27.1 - 30 min 95% A. The flow rate was 1 mL x min⁻¹ at a column temperature of 40 °C. LC was coupled to an Agilent MSD Trap XCT-Plus mass spectrometer equipped with an electro-spray operated in negative ionization mode (capillary voltage, 4000 eV; temp, 350 °C; nebulizing gas, 50 psi; dry gas 12 L/min) and an Agilent 1100 Diode Array Detector.
(DAD, detection 200 - 700 nm, J&M Analytik AG, Aalen, Germany). MS\textsuperscript{n} was conducted to analyse fragmentation patterns for compound identification. The LC/MSD Trap Software 5.2 (Bruker Daltonik, GmbH) was used for data acquisition and processing. Enzyme products were quantified using an external MbA standard curve. Compounds were tentatively identified using their molecular masses and specific fragmentation patterns. Authentic standards were available for mini-MbA, MbA-XR\textsuperscript{2}, MbA-R and MbA (Williams et al., 2015).

Accurate mass measurement was performed on an Agilent 1290 Infinity UHPLC (Agilent Technologies GmbH, Waldbronn, Germany) utilizing the same column, mobile phase and method C as described. The LC was coupled to an Agilent 6530 Accurate Mass Q-ToF mass spectrometer equipped with an electrospray ion source operated in negative ionization mode (capillary voltage, 4000 eV; temp, 350 °C; nebulizing gas, 60 psi; dry gas 12 L/min) and an Agilent 1290 Diode Array Detector (DAD, detection 190-400 nm, J&M Analytik AG, Aalen, Germany).

Reverse transcription and RT-qPCR
RNA was extracted and cDNA synthesis done as previously described (Irmisch et al., 2018). For RT-qPCR the cDNA was diluted 1:5 with water. For the amplification of CcUGT4 (UGT703H1) and CcUGT5 (UGT729A2) gene fragments of 169 and 141 bp length respectively, a primer pair was designed with a Tm ≥ 60 °C, a GC content of 50 - 60%, and a primer length of 20 to 22 nt (Supplemental Table S6). Primer specificity was confirmed by agarose gel electrophoresis, melting curve analysis, standard curve analysis and by sequence verification of cloned PCR products. RT-qPCR reactions were performed in duplicate on a Bio-Rad CFX96\textsuperscript{TM} instrument (Bio-Rad Laboratory, Hercules, CA, USA) in optical 96-well plates using SsoFast\textsuperscript{TM} EvaGreen® Supermix (BioRad) with the following PCR conditions: Initial incubation at 95 °C for 30 sec followed by 40 cycles of amplification (95 °C for 5 sec, 60 °C for 10 sec). RT-qPCR analyses were performed with three biological replicates for each of the six different time points of yC and oC collections in 2016 (June 10\textsuperscript{th}, June 27\textsuperscript{th}, July 22\textsuperscript{nd}, August 16\textsuperscript{th}, September 12\textsuperscript{th}, October 06\textsuperscript{th}) as described in Irmisch et al.
Serin-incorporator (MEP) and zinc-finger protein (ZF) were used as reference genes.

**Transient expression in *N. benthamiana***

For expression in *N. benthamiana*, the coding regions of *CcUGT4* and *CcUGT5* were separately cloned into the pCAMBiA2300U vector. After sequence verification, the pCAMBiA vectors carrying *CcUGT4* or *CcUGT5* as well as pCAMBiA vectors carrying the previously described genes for myricetin biosynthesis, *CcFLS*, *CcCYP2* and *CcMYB4* (Irmisch et al., 2019a), and the genes for MbA-XR² biosynthesis, *CcUGT1*, *CcUGT2*, *CcAT1* and *CcUGT3* (Irmisch et al., 2018; Irmisch et al., 2019b), and the pBIN:p19 were separately transferred into *Agrobacterium tumefaciens* strain C58pMP90. One mL of overnight cultures (220 rpm, 28 °C) was used to inoculate 10 mL LB-media containing 50 µg x mL⁻¹ kanamycin, 25 µg x mL⁻¹ rifampicin and 25 µg x mL⁻¹ gentamicin for overnight growth. The following day the cultures were centrifuged (4,000 x g, 5 min), and cells were re-suspended in infiltration buffer (10 mM MES, pH 5.6, 10 mM MgCl₂, 100 µM acetosyringone) to a final OD₆₀₀ of 0.5. After shaking for 3 h at RT, the following combinations of transformed *A. tumefaciens* were prepared for leaf infiltration using: (i) *A. tumefaciens* 3Sₚₜ: (*CcMYB4* + *CcFLS* + *CcCYP2* + *CcUGT1* + *CcUGT2* + *CcAT1* + *CcUGT3*) + *A. tumefaciens* pBIN:p19; (ii) *A. tumefaciens* 3Sₚₜ: (*CcMYB4* + *CcFLS* + *CcCYP2* + *CcUGT1* + *CcUGT2* + *CcAT1* + *CcUGT3*) + *A. tumefaciens* pBIN:p19; iii) *A. tumefaciens* 3Sₚₜ: (*CcMYB4* + *CcFLS* + *CcCYP2* + *CcUGT1* + *CcUGT2* + *CcAT1* + *CcUGT3* + *CcUGT4* + *CcUGT5*) + *A. tumefaciens* pBIN:p19. Equal volumes of each line of transformed *A. tumefaciens* were used to prepare the mixtures. The leaves of four-week-old *N. benthamiana* plants were infiltrated with *A. tumefaciens* solution using a 1-mL needle-free syringe to gently push the bacterial mixture into the abaxial surface. Infiltrated leaves were labeled with tape and harvested five days after infiltration and directly frozen in liquid nitrogen and stored at -80°C until further analysis. Plant material was ground in liquid nitrogen into a fine powder, and 100 mg were extracted with 1 mL 50% (v/v) MeOH for 2 h at RT. The extracts were analyzed using LC-MS.

**Alignment and phylogenetic tree construction**

An amino acid alignment of montbretia UGTs and other plant UGTs (Supplemental Table S1) was constructed using the MUSCLE algorithm implemented in MEGA6 (Tamura et al., 2011). Based on this alignment, a phylogenetic tree was constructed using the **phylogenetic tree construction**
reconstructed with MEGA6 using a neighbor-joining algorithm (Poisson model). A bootstrap resampling analysis with 1000 replicates was performed to evaluate the tree topology. Selection of plant UGTs depicted in the phylogeny was based on UGT enzymes involved in the biosynthesis of more complex specialized metabolites, GGTs and UGTs utilizing sugar donors other than UDP-glucose.

**Statistical analysis**

To test for significant differences in *CcUGT4* and *CcUGT5* transcript abundance in yC and oC at different time points, data were log transformed to meet statistical requirements and a two-way analysis of variance (ANOVA) was performed followed by a Tukey-Test using SigmaPlot 11.0 for Windows (Systat Software Inc. 2008) (Supplemental Table S3).

**Accession numbers**

Previously published transcriptome libraries as well as time course data described in the manuscript are available in the NCBI/GenBank Sequence Read Archive (SRA) under the project PRJNA389589 (SRP108844). UGT DE data of yC and oC were previously published (Irmisch et al., 2018). New UGT nucleotide sequences were deposited in GenBank with the accession numbers MT386071 (*CcUGT4, UGT703H1*) and MT386072 (*CcUGT5, UGT729A2*). Accession numbers for all other genes/proteins used in this work are listed in Supplemental Table S1, and sequence information of the 190 montbretia UGTs can be found in the Supplemental data file.

**Supplemental Data**

**Supplemental Figure S1.** Alignment and immunoblot of montbretia UGTs and screening of candidate UGTs for MbA-R² and MbA formation activity.

**Supplemental Figure S2.** MbA-R² and MbA formation with *CcUGT4* and *CcUGT5*, respectively.

**Supplemental Figure S3.** Activity of purified *CcUGT4* with different sugar donor and acceptor substrates.

**Supplemental Figure S4.** Activity of purified *CcUGT5* with MbA-CR² and MbA-R² and different sugar donors.

**Supplemental Figure S5.** Reconstruction of MbA and MbB biosynthesis using identified MbA biosynthesis enzymes.
Supplemental Figure S6. MbA-R^2, MbB-R^2, MbA and MbB production in N. benthamiana transiently expressing montbretia MbA biosynthesis genes.

Supplemental Figure S7. MbA-C production in N. benthamiana.

Supplemental Table S1. NCBI identification numbers of sequences used in the phylogenetic tree.

Supplemental Table S2. Activity assays of CcUGT4 (UGT703H1) and CcUGT5 (UGT729A2) with different acceptors.

Supplemental Table S3. Statistical analysis of gene expression data.

Supplemental Table S4. Accurate mass data.

Supplemental Table S5. MbA and MbB production in N. benthamiana.

Supplemental Table S6. Oligonucleotides

Supplemental Data Set 1. Amino acid sequences of the 190 putative UGTs.

Supplemental Data Set 2. UGT expression data

Acknowledgements

We thank Dr. Stephen G. Withers (UBC) for discussions and collaboration, Dr. Carol Ritland (UBC) for excellent project management and Michael Court and the UGT nomenclature committee for naming of montbretia UGTs. The research was supported with funds to JB from the GlycoNet Networks of Centres of Excellence (GlycoNet NCE) and the Natural Sciences and Engineering Research Council of Canada (NSERC, Discovery Grant). SI was supported by the Alexander-von-Humboldt foundation through a Feodor Lynen Research Fellowship. JB is a Distinguished University Scholar.

Figure Legends

Figure 1. Schematic structures of MbA and intermediates in the MbA biosynthetic pathway and expression profiles of candidate UGTs and previously characterized MbA pathway genes. (A) The six steps of the MbA pathway are indicated by numbers below the arrows. Genes and enzymes for steps 1
to 4 have previously been characterized (Irmisch et al., 2018; Irmisch et al., 2019b). The two final steps from MbA-XR$_2$ to MbA-R$_2$ and MbA are illustrated. Schematic of MbA-XR$_2$ (myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside), MbA-R$_2$ (myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-xyloloside) and MbA, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-1,2-β-D-glucosyl 1,2-α-L-rhamnoside 4'-O-α-L-rhamnosyl 1,4-β-D-xyloloside. M, myricetin (pink); R, rhamnose (yellow); G, glucose (green); X, xylolose (orange); C, caffeic acid (blue). (B) Heatmap showing relative transcript abundance of UGTs and two ATs in montbretia young corms at six different time points of corm development (see Irmisch et al, 2018 for images of developing corms). Genes previously characterized as being involved in MbA biosynthesis are shown in red. Candidate UGTs selected for characterization in this work are shown within the green frame. Identification numbers or gene names are given, and sampling dates in 2016 are indicated at the bottom.

**Figure 2. Phylogeny of montbretia UGT4 (UGT703H1) and UGT5 (UGT729A2).**
Amino acid sequences of montbretia UGTs were aligned with selected UGTs from other plant species and a neighbor joining tree was constructed using MEGA6. Phylogeny was used to cluster montbretia UGTs with known UGT clades. UGTs depicted in the phylogeny were chosen with emphasis on glycoside glycosyltransferases (GGT) and UGTs accepting sugar donors different from UDP-glucose. Those characteristics are indicated by the colored circles and UGT clades are labeled. Montbretia UGTs are bold, UGTs involved in MbA biosynthesis are red and bold, and UGTs characterized in this work are additionally underlined. Species abbreviations and accession numbers can be found in Supplemental Table S1.

**Figure 3. Enzyme activity of purified CcUGT4 and CcUGT5.** UGTs were heterologously expressed in *E. coli*, and Ni-purified protein was assayed for activity with MbA-XR$_2$ and UDP-Xyl for CcUGT4 (A) or with MbA-R$_2$ and UDP-Rha for CcUGT5 (B). As controls, assays were performed with protein extracts of *E. coli* transformed with the empty vector. Products were analysed using LC-MS and LC-UV (method B). UV detection was monitored at 350 - 370 nm. MS/MS fragmentation patterns of the reaction products and respective standards formed in A (C) and B (D) are shown. Peak 1, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside (MbA-XR$_2$); peak 2, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-xyloloside (MbA-R$_2$); peak 3, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-α-L-rhamnosyl 1,4-β-D-xyloloside (MbA-R$_2$).
rhamnoside 4’-O-rhamnosyl xyloside (MbA). Asterisks indicate mother ion for MS/MS fragmentation.

**Figure 4. Transcript abundance of CcUGT4 and CcUGT5 in yC and oC over a time course of corm development.** RNA was isolated from yC and oC harvested at different time points of corm development. Transcript abundance was determined by RT-qPCR for CcUGT4 (A) and CcUGT5 (B). Means and standard errors are shown (n = 3). Different letters above the data points indicate significant differences (p ≤ 0.05) between harvest points in yC. Asterisks (*) indicate the statistical significance (p ≤ 0.05) between yC and oC for time points.

**Figure 5: Reconstruction of MbA biosynthesis in N. benthamiana using montbretia genes.** N. benthamiana leaves were infiltrated with different combinations of A. tumefaciens transformed with plasmids carrying the 35S-promoter-gene constructs for myricetin and MbA biosynthesis genes. Leaves were collected at day five after infiltration. Metabolites were extracted with 50% MeOH, analyzed by LC-MS (method B) and identified based on comparison of their retention times and fragmentation patterns to authentic standards for MbA-XR², MbA-R², MbA or enzyme products for MbB-XR², MbB-R², MbB. The extracted ion chromatograms (EIC) (A) and schematic structure of products (B) are shown. Combinations of genes used for transient expression in N. benthamiana (C). Peak 1, unidentified; peak 2, MbB-XR²; peak 3, tentatively identified as myricetin 3-O-(glucosyl-6’-O-coumaroyl)-glucosyl glucoside; peak 4, MbA-XR²; peak 5-8, unidentified; peak 9, MbB-R²; peak 10, MbA-R²; peak 11, MbA-C; peak 12-13, unidentified; peak 14, MbB; peak 15, unidentified; peak 16, MbA. MbB-XR², myricetin 3-O-(glucosyl-6’-O-coumaroyl)-glucosyl rhamnoside; MbA-R², myricetin 3-O-(glucosyl-6’-O-caffeoyl)-glucosyl rhamnoside; MbB-R², myricetin 3-O-(glucosyl-6’-O-coumaroyl)-glucosyl rhamnoside 4’-O-xyloside; MbA-R², myricetin 3-O-(glucosyl-6’-O-caffeoyl)-glucosyl rhamnoside 4’-O-xyloside; MbB, myricetin 3-O-(glucosyl-6’-O-coumaroyl)-glucosyl rhamnoside 4’-O-rhamnosyl xyloside; MbA, myricetin 3-O-(glucosyl-6’-O-caffeoyl)-glucosyl rhamnoside 4’-O-rhamnosyl xyloside.

Downloaded on November 5, 2020. - Published by https://plantphysiol.org
Copyright (c) 2020 American Society of Plant Biologists. All rights reserved.
Asada Y, Ueoka T, Furuya T (1989) Novel acylated saponins from montbretia (Crocosmia crocosmiiflora). isolation of saponins and the structures of Crocosmiosides A, B and H. Chem & Pharm Bull 37: 2139-2146

Asada Y, Hirayama Y, Furuya T (1988) Acylated flavonols from Crocosmia-Crocosmiiflora. Phytochemistry 27: 1497-1501

Bowles D, Isayenkova J, Lim EK, Poppenberger B (2005) Glycosyltransferases: managers of small molecules. Curr Opin Plant Biol 8: 254-263

Caputi L, Malnoy M, Goremykin V, Nikiforova S, 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

Chen H-Y, Li X (2017) Identification of a residue responsible for UDP-sugar donor selectivity of a dihydroxybenzoic acid glycosyltransferase from Arabidopsis natural accessions. Plant J 89: 195-203

Di S, Yan F, Rodas FR, Rodriguez TO, Murai Y, Iwashina T, Sugawara S, Mori T, Nakabayashi R, Yonekura-Sakakibara K, Saito K, Takahashi R (2015) Linkage mapping, molecular cloning and functional analysis of soybean gene Fg3 encoding flavonol 3-O-glucoside/galactoside (1 --> 2) glucosyltransferase. BMC Plant Biol 15: 126

Frydman A, Liberman R, Huhman DV, Carmeli-Weissberg M, Sapir-Mir M, Ophir R, Summer LW, Eyal Y (2013) The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. Plant J 73: 166-178

Funaki A, Waki T, Noguchi A, Kawai Y, Yamashita S, Takahashi S, Nakayama T (2015) Identification of a highly specific isoflavone 7-O-glucosyltransferase in the soybean (Glycine max (L.) Merr.). Plant Cell Physiol 56: 1512-1520

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

Hall D, Kim KH, De Luca V (2011) Molecular cloning and biochemical characterization of three Concord grape (Vitis labrusca) flavonol 7-O-glucosyltransferases. Planta 234: 1201-1214

Hofer B (2016) Recent developments in the enzymatic O-glycosylation of flavonoids. Appl Microbiol Biotechnol 100: 4269-4281

Hsu Y-H, Tagami T, Matsunaga K, Okuyama M, Suzuki T, Noda N, Suzuki M, Shimura H (2017) Functional characterization of UDP-rhamnose-dependent rhamnosyltransferase involved in anthocyanin modification, a key enzyme determining blue coloration in Lobelia erinus. Plant J 89: 325-337
Imayama T, Yoshihara N, Fukuchi-Mizutani M, Tanaka Y, Ino I, Yabuya T (2004) Isolation and characterization of a cDNA clone of UDP-glucose: anthocyanin 5-O-glucosyltransferase in Iris hollandica. Plant Sci 167: 1243-1248

Irmisch S, Jancsik S, Yuen MMS, Madilao LL, Bohlmann J (2019b) Biosynthesis of the anti-diabetic metabolite montbretin A: glucosylation of the central intermediate mini-MbA. Plant J 100: 879-891

Irmisch S, Ruebsam H, Jancsik S, Man Saint Yuen M, Madilao LL, Bohlmann J (2019a) Flavonol biosynthesis genes and their use in engineering the plant antidiabetic metabolite montbretin A. Plant Physiol 180: 1277-1290

Irmisch S, Jo S, Roach CR, Jancsik S, Man Saint Yuen M, Madilao LL, O’Neil-Johnson M, Williams R, Withers SG, Bohlmann J (2018) Discovery of UDP-glucosyltransferases and BAHD-acyltransferases involved in the biosynthesis of the antidiabetic plant metabolite montbretin A. Plant Cell 30: 1864-1886

Itkin M, Davidovich-Rikanati R, Cohen S, Portnoy V, Doron-Faigenboim A, Oren E, Freilich S, Tzuri G, Baranes N, Shen S, Petreikov M, Sertchook R, Ben-Dor S, Gottlieb H, Hernandez A, Nelson DR, Paris HS, Tadmor Y, Burger Y, Lewinsohn E, Katzir N, Schaffer A (2016) The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from Siraitia grosvenorii. PNAS 113: E7619-E7628

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitisky S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP, Aharoni A (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science 341: 175-179

Jones P, Messner B, Nakajima J, Schaffner AR, Saito K (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. J Biol Chem 278: 43910-43918

Jung S-C, Kim W, Park SC, Jeong J, Park MK, Lim S, Lee Y, Im W-T, Lee JH, Choi G, Kim SC (2014) Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. Plant Cell Physiol 55: 2177-2188

Kim B-G, Sung SH, Jung NR, Chong Y, Ahn J-H (2010) Biological synthesis of isorhamnetin 3-O-glucoside using engineered glucosyltransferase. J Mol Catal B Enzym 63: 194-199

Kubo A, Arai Y, Nagashima S, Yoshikawa T (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys 429: 198-203

Lee SG, Salomon E, Yu O, Jez JM (2019) Molecular basis for branched steviol glucoside biosynthesis. PNAS 116: 13131-13136

Liu Z, Li J, Sun Y, Zhang P, Wang Y (2020) Structural insights into the catalytic mechanism of a plant diterpene glycosyltransferase SrUGT76G1. Plant Com 1: 100004
Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42: 21

McCue KF, Allen PV, Shepherd LVT, Blake A, Malendia Maccree M, Rockhold DR, Novy RG, Stewart D, Davies HV, Belknap WR (2007) Potato glycosyl rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. Phytochemistry 68: 327-334

Montefiori M, Espley RV, Stevenson D, Cooney J, Datson PM, Saiz A, Atkinson RG, Hellens RP, Allan AC (2011) Identification and characterisation of F3G1T1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (Actinidia chinensis). Plant J 65: 106-118

Nip KM, Chiu R, Yang C, Chu J, Mohamadi H, Warren RL, Birol I (2019) RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. bioRxiv 701607.

Noguchi A, Horikawa M, Fukui Y, Fukuchi-Mizutani M, Iuchi-Okada A, Ishiguro M, Kiso Y, Nakayama T, Ono E (2009) Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in lamiales. Plant Cell 21: 1556-1572

Noguchi A, Fukui Y, Iuchi-Okada A, Kakutani S, Satake H, Iwashita T, Nakao M, Umezawa T, Ono E (2008) Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by Sesamum indicum UGT71A9 and UGT94D1 glucosyltransferases. Plant J 54: 415-427

Ogata J, Kanno Y, Itoh Y, Tsugawa H, Suzuki M (2005) Anthocyanin biosynthesis in roses. Nature 435: 757-758

Ono E, Homma Y, Horikawa M, Kunikane-Doi S, Imai H, Takahashi S, Kawai Y, Ishiguro M, Fukui Y, Nakayama T (2010) Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (Vitis vinifera). Plant Cell 22: 2856-2871

Osmani SA, Bak S, Moller BL (2009) Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70: 325-347

Richman A, Swanson A, Humphrey T, Chapman R, McGarvey B, Pocs R, Brandle J (2005) Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of Stevia rebaudiana. Plant J 41: 56-67

Rojas Rodas F, Rodriguez TO, Murai Y, Iwashina T, Sugawara S, Suzuki M, Nakabayashi R, Yonekura-Sakakibara K, Saito K, Kitajima J, Toda K, Takahashi R (2014) Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-O-glucoside (1 --> 6) rhamnosyltransferase. Plant Mol Biol 84: 287-300

Ruby, Santosh Kumar RJ, Vishwakarma RK, Singh S, Khan BM (2014) Molecular cloning and characterization of genistein 4'-O-glucoside specific glycosyltransferase from Bacopa monniera. Mol Biol Rep 41: 4675-4688
Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S, Tatsuizaka N, Tsukamoto C, Ishimoto M (2012) The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell 24: 2123-2138

Scheen AJ (2003) Is there a role for α-glucosidase inhibitors in the prevention of type 2 diabetes mellitus? Drugs 63: 933-951

Scheen AJ (1997) Drug treatment of non-insulin-dependent diabetes mellitus in the 1990s - Achievements and future developments. Drugs 54: 355-368

Scott LJ, Spencer CM (2000) Miglitol: a review of its therapeutic potential in type 2 diabetes mellitus. Drugs 59: 521-549

Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in Glycine max. FEBS Letters 584: 2258-2264

Song CK, Gu L, Liu JY, Zhao S, Hong XT, Schulenburg K, Schwab W (2015) Functional characterization and substrate promiscuity of UGT71 glycosyltransferases from Strawberry (Fragaria x ananassa). Plant Cell Physiol 56: 2478-2493

Tarling CA, Woods K, Zhang R, Brastianos HC, Brayer GD, Andersen RJ, Withers SG (2008) The search for novel human pancreatic alpha-amylase inhibitors: High-throughput screening of terrestrial and marine natural product extracts. Chembiochem 9: 433-438

Trapero A, Ahrazem O, Rubio-Moraga A, Jimeno ML, Gomez MD, Gomez-Gomez L (2012) Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in Crocus sativus. Plant Phys 159: 1335-1354

Williams LK, Zhang X, Caner S, Tysoe C, Nguyen NT, Wicki J, Williams DE, Coleman J, McNeill JH, Yuen V, Andersen RJ, Withers SG, Brayer GD (2015) The amylase inhibitor montbretin A reveals a new glycosidase inhibition motif. Nat Chem Biol 11: 691-696

Wilson AE, Tian L (2019) Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. Plant J 100: 1273-1288

Witte S, Moco S, Vervoort J, Matern U, Martens S (2009) Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from Hieracium pilosella L. Planta 229: 1135-1146

Yonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi M, Kitajima M, Takayama H, Saito K (2014) A flavonoid 3-O-glucoside:2"-O-glycosyltransferase responsible for terminal modification of pollen-specific flavonols in Arabidopsis thaliana. Plant J 79: 769-782
Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, Inoue E, Kuromori T, Ito T, Shinozaki K, Wangwattana B, Yamazaki M, Saito K (2012) Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in Arabidopsis thaliana. Plant J 69: 154-167

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

Yonekura-Sakakibara K, Tohge T, Niida R, Saito K (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. J Biol Chem 282: 14932-14941

Yuen VG, Coleman J, Withers SG, Andersen RJ, Brayer GD, Mustafa S, McNeill JH (2016) Glucose lowering effect of montbretin A in Zucker Diabetic Fatty rats. Mol Cell Biochem 411: 373-381

Zong G, Fei S, Liu X, Li J, Gao Y, Yang X, Wang X, Shen Y (2019) Crystal structures of rhamnosyltransferase UGT89C1 from Arabidopsis thaliana reveal the molecular basis of sugar donor specificity for UDP-β-L-rhamnose and rhamnosylation mechanism. Plant J 99: 257-269
Figure 1. Schematic structures of MbA and intermediates in the MbA biosynthetic pathway and expression profiles of candidate UGTs and previously characterized MbA pathway genes. (A) The six steps of the MbA pathway are indicated by numbers below the arrows. Genes and enzymes for steps 1 to 4 have previously been characterized (Irmisch et al., 2018; Irmisch et al., 2019b). The two final steps from MbA-XR² to MbA-R² and MbA are illustrated. Schematic of MbA-XR² (myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside), MbA-R² (myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-xyloside) and MbA, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-1,2-glucosyl 1,2-rhamnoside 4'-O-rhamnosyl 1,4-xyloside. M, myricetin (pink); R, rhamnose (yellow); G, glucose (green); X, xylose (orange); C, caffeic acid (blue). (B) Heatmap showing relative transcript abundance of UGTs and two ATs in montbretia young corms at six different time points of corm development (see Irmisch et al., 2018 for images of developing corms). Genes previously characterized as being involved in MbA biosynthesis are shown in red. Candidate UGTs selected for characterization in this work are shown within the green frame. Identification numbers or gene names are given and sampling dates in 2016 are indicated at the bottom.
Figure 2. Phylogeny of montbretia UGT4 (UGT703H1) and UGT5 (UGT729A2). Amino acid sequences of montbretia UGTs were aligned with selected UGTs from other plant species and a neighbor joining tree was constructed using MEGA6. Phylogeny was used to cluster montbretia UGTs with known UGT clades. UGTs depicted in the phylogeny were chosen with emphasis on glycoside glycosyltransferases (GGT) and UGTs accepting sugar donors different from UDP-glucose. Those characteristics are indicated by the colored circles and UGT clades are labeled. Montbretia UGTs are bold, UGTs involved in MbA biosynthesis are red and bold, and UGTs characterized in this work are additionally underlined. Species abbreviations and accession numbers can be found in Supplemental Table S1.
Figure 3. Enzyme activity of purified CcUGT4 and CcUGT5. UGTs were heterologously expressed in E. coli, and Ni-purified protein was assayed for activity with MbA-XR² and UDP-Xyl for CcUGT4 (A) or with MbA-R² and UDP-Rha for CcUGT5 (B). As controls, assays were performed with protein extracts of E. coli transformed with the empty vector. Products were analysed using LC-MS/MS and LC-UV (method B). UV detection was monitored at 350 - 370 nm. MS/MS fragmentation patterns of the reaction products and respective standards formed in A (C) and B (D) are shown. Peak 1, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside (MbA-XR²); peak 2, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-xyloside (MbA-R²); peak 3, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-rhamnosyl xyloside (MbA). Asterisks indicate mother ion for MS/MS fragmentation.
Figure 4. Transcript abundance of *CcUGT4* and *CcUGT5* in yC and oC over a time course of corm development. RNA was isolated from yC and oC harvested at different time points of corm development. Transcript abundance was determined by RT-qPCR for *CcUGT4* (A) and *CcUGT5* (B). Means and standard errors are shown (n = 3). Different letters above the data points indicate significant differences (p ≤ 0.05) between harvest points in yC. Asterisks (*) indicate the statistical significance (p ≤ 0.05) between yC and oC for time points.
Figure 5. Reconstruction of MbA biosynthesis in *N. benthamiana* using montbretia genes. *N. benthamiana* leaves were infiltrated with different combinations of *A. tumefaciens* transformed with plasmids carrying the 35S-promoter-gene constructs for myricetin and MbA biosynthesis genes. Leaves were collected at day five after infiltration. Metabolites were extracted with 50% MeOH, analyzed by LC-MS (method B) and identified based on comparison of their retention times and fragmentation patterns to authentic standards for MbA-XR<sup>2</sup>, MbA-R<sup>2</sup>, MbA or enzyme products for MbB-XR<sup>2</sup>, MbB-R<sup>2</sup>, MbB. The extracted ion chromatograms (EIC) (A) and schematic structure of products (B) are shown. Combinations of genes used for transient expression in *N. benthamiana* (C). Peak 1, unidentified; peak 2, MbB-XR<sup>2</sup>; peak 3, tentatively identified as myricetin 3-O-(glucosyl-6'-coumaroyl)-glucosyl glucoside; peak 4, MbA-XR<sup>2</sup>; peak 5-8, unidentified; peak 9, MbB-R<sup>2</sup>; peak 10, MbA-R<sup>2</sup>; peak 11, MbA-C; peak 12-13, unidentified; peak 14, MbB; peak 15, unidentified; peak 16, MbA. MbB-XR, myricetin 3-O-(glucosyl-6'-coumaroyl)-glucosyl rhamnoside; MbB-R<sup>2</sup>, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside; MbB-R<sup>2</sup>, myricetin 3-O-(glucosyl-6'-O-coumaroyl)-glucosyl rhamnoside 4'-O-rhamnosyl xyloside; MbA, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-rhamnosyl xyloside; MbA, myricetin 3-O-(glucosyl-6'-O-coumaroyl)-glucosyl rhamnoside 4'-O-rhamnosyl xyloside; MbA, myricetin 3-O-(glucosyl-6'-O-caffeoyl)-glucosyl rhamnoside 4'-O-rhamnosyl xyloside.

**Combination of genes used for transformation**

| Target       | Combination of genes used for transformation                      |
|--------------|------------------------------------------------------------------|
| MbA-XR<sup>2</sup> | MYB4+CYP2+FLS+UGT1+UGT2+AT1+UGT3                               |
| MbA-R<sup>2</sup>   | MYB4+CYP2+FLS+UGT1+UGT2+AT1+UGT3+UGT4                          |
| MbA           | MYB4+CYP2+FLS+UGT1+UGT2+AT1+UGT3+UGT4+UGT5                      |
Asada Y, Ueoka T, Furuya T (1989) Novel acylated saponins from montbretia (Crocosmia crocosmiiflora). isolation of saponins and the structures of Crocosmioids A, B and H. Chem & Pharm Bull 37: 2139-2146

Pubmed: Author and Title
Google Scholar: Author Only Title Only Author and Title

Asada Y, Hirayama Y, Furuya T (1988) Acylated flavonols from Crocosmia-Crocosmiiflora. Phytochemistry 27: 1497-1501

Pubmed: Author and Title
Google Scholar: Author Only

Bowles D, Isayenkov J, Lim EK, Poppenberger B (2005) Glycosyltransferases: managers of small molecules. Curr Opin Plant Biol 8: 254-263

Pubmed: Author and Title
Google Scholar: Author Only

Caputi L, Malnoy M, Goremykin V, Nikiforova S, 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

Pubmed: Author and Title
Google Scholar: Author Only

Chen H-Y, Li X (2017) Identification of a residue responsible for UDP-sugar donor selectivity of a dihydroxybenzoic acid glycosyltransferase from Arabidopsis natural accessions. Plant J 89: 195-203

Pubmed: Author and Title
Google Scholar: Author Only

Di S, Yan F, Rodas FR, Rodriguez TO, Murai Y, Iwashina T, Sugawara S, Mori T, Nakabayashi R, Yonekura-Sakakibara K, Saito K, Takahashi R (2015) Linkage mapping, molecular cloning and functional analysis of soybean gene Fg3 encoding flavonol 3-O-galactoside/galactoside (1 → 2) glycosyltransferase. BMC Plant Biol 15: 126

Pubmed: Author and Title
Google Scholar: Author Only

Frydman A, Liberman R, Huhman DV, Carmeli-Weissberg M, Sapir-Mir M, Ophir R, Summer LW, Eyal Y (2013) The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. Plant J 73: 166-178

Pubmed: Author and Title
Google Scholar: Author Only

Funaki A, Waki T, Noguchi A, Kawai Y, Yamashita S, Takahashi S, Nakayama T (2015) Identification of a highly specific isoflavone 7-O-glycosyltransferase in the soybean (Glycine max (L.) Merr.). Plant Cell Physiol 56: 1512-1520

Pubmed: Author and Title
Google Scholar: Author Only

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

Pubmed: Author and Title
Google Scholar: Author Only

Hall D, Kim KH, De Luca V (2011) Molecular cloning and biochemical characterization of three Concord grape (Vitis labrusca) flavonol 7-O-glucosyltransferases. Planta 234: 1201-1214

Pubmed: Author and Title
Google Scholar: Author Only

Hofer B (2016) Recent developments in the enzymatic O-glycosylation of flavonoids. Appl Microbiol Biotechnol 100: 4269-4281

Pubmed: Author and Title
Google Scholar: Author Only

Hsu Y-H, Tagani T, Matsunaga K, Okuyama M, Suzuki T, Noda N, Suzuki M, Shimura H (2017) Functional characterization of UDP-rhamnose-dependent rhamnosyltransferase involved in anthocyanin modification, a key enzyme determining blue coloration in Lobelia erinus. Plant J 89: 325-337

Pubmed: Author and Title
Google Scholar: Author Only

Imayama T, Yoshihara N, Fukuchi-Mizutani M, Tanaka Y, Ito I, Yabuya T (2004) Isolation and characterization of a cDNA clone of UDP-glucose: anthocyanin 5-O-glucosyltransferase in Iris hollandica. Plant Sci 167: 1243-1248

Pubmed: Author and Title
Google Scholar: Author Only

Irmisch S, Jancsik S, Yuen MMS, Madilao LL, Bohlmann J (2019b) Biosynthesis of the anti-diabetic metabolite montbretin: a glucosylation of the central intermediate mini-MbA. Plant J 100: 879-891

Pubmed: Author and Title
Google Scholar: Author Only

Irmisch S, Ruebsam H, Jancsik S, Man Saint Yuen M, Madilao LL, Bohlmann J (2019a) Flavonol biosynthesis genes and their use in...
engineering the plant antidiabetic metabolite montbretin A. Plant Physiol 180: 1277-1290

Irmisch S, Jo S, Roach CR, Jancsik S, Man Saint Yuen M, Madlao LL, O'Neil-Johnson M, Williams R, Wthers SG, Bohlmann J (2018) Discovery of UDP-glycosyltransferases and BAHD-acyltransferases involved in the biosynthesis of the antidiabetic plant metabolite montbretin A. Plant Cell 30: 1864-1886

Itkin M, Davidovich-Rikanati R, Cohen S, Portnoy V, Doron-Faigenboim A, Oren E, Freilich S, Tzuri G, Baranes N, Shen S, Petreikov M, Sertchook R, Ben-Dor S, Gottlieb H, Hernandez A, Nelson DR, Paris HS, Tadmor Y, Burger Y, Lewinsohn E, Katzir N, Schaffer A (2016) The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from Siraitia grosvenorii. PNAS 113: E7619-E7628

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP, Aharoni A (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science 341: 175-179

Jones P, Messner B, Nakajima J, Schaffner AR, Saito K (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. J Biol Chem 278: 43910-43918

Jung S-C, Kim W, Park SC, Jeong J, Park MK, Lim S, Lee Y, Im WT, Lee JH, Choi G, Kim SC (2014) Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd. Plant Cell Physiol 55: 2177-2188

Kubo A, Arai Y, Nagashima S, Yoshikawa T (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys 429: 198-203

Lee SG, Salomon E, Yu O, Jez JM (2019) Molecular basis for branched steviol glucoside biosynthesis. PNAS 116: 13131-13136

Liu Z, Li J, Sun Y, Zhang P, Wang Y (2020) Structural insights into the catalytic mechanism of a plant diterpene glycosyltransferase SrUGT76G1. Plant Com 1: 100004

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZY) in 2013. Nucleic Acids Res 42: 21

McCue KF, Allen PV, Shepherd LVT, Blake A, Malendia Maccree M, Rockhold DR, Novy RG, Stewart D, Davies HV, Belknap WR (2007) Potato glycososterol rhamnosyltransferase, the terminal step in triose side-chain biosynthesis. Phytochemistry 68: 327-334

Montefiori M, Espley RV, Stevenson D, Cooney J, Datson PM, Saiz A, Atkinson RG, Hellens RP, Allan AC (2011) Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (Actinidia chinensis). Plant J 65: 106-118

Nip KM, Chiu R, Yang C, Chu J, Mohamadi H, Warren RL, Birol I (2019) RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. bioRxiv 701607

Noguchi A, Horikawa M, Fukui Y, Fukuchi-Mizutani M, luchi-Okada A, Ishiguro M, Kiso Y, Nakayama T, Ono E (2009) Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in lamiales. Plant Cell 21: 1556-1572
Noguchi A, Fukui Y, Iuchi-Okada A, Satake H, Iwashita T, Nakao M, Urezawa T, Ono E (2008) Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by Sesamum indicum UGT71A9 and UGT94D1 glucosyltransferases. Plant J 54: 415-427

Ogata J, Kanno Y, Itoh Y, Tsugawa H, Suzuki M (2005) Anthocyanin biosynthesis in roses. Nature 435: 757-758

Ono E, Homma Y, Horikawa M, Kunikane-Doi S, Imai H, Takahashi S, Kawai Y, Ishiguro M, Fukui Y, Nakayama T (2010) Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (Vitis vinifera). Plant Cell 22: 2856-2871

Osmani SA, Bak S, Moller BL (2009) Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70: 325-347

Richman A, Swanson A, Humphrey T, Chapman R, McGarvey B, Pocs R, Brandle J (2005) Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of Stevia rebaudiana. Plant J 41: 56-67

Rojas Rodas F, Rodriguez TO, Murai Y, Iwashina T, Sugawara S, Suzuki M, Nakabayashi R, Yonekura-Sakakibara K, Saito K, Kitajima J, Toda K, Takahashi R (2014) Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-O-glucoside (1 --> 6) rhamnosyltransferase. Plant Mol Biol 84: 287-300

Ruby, Santosh Kumar RJ, Vishwakarma RK, Singh S, Khan BM (2014) Molecular cloning and characterization of genistein 4’-O-glucoside specific glycosyltransferase from Bacopa monniera. Mol Biol Rep 41: 4675-4688

Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S, Tatsuzaki N, Tsukamoto C, Ishimoto M (2012) The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell 24: 2123-2138

Scheen AJ (2003) Is there a role for α-glucosidase inhibitors in the prevention of type 2 diabetes mellitus? Drugs 63: 933-951

Scheen AJ (1997) Drug treatment of non-insulin-dependent diabetes mellitus in the 1990s - Achievements and future developments. Drugs 54: 355-368

Scott LJ, Spencer CM (2000) Miglitol: a review of its therapeutic potential in type 2 diabetes mellitus. Drugs 59: 521-549

Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in Glycine max. FEBS Letters 584: 2258-2264

Song CK, Gu L, Liu JY, Zhao S, Hong XT, Schulenburg K, Schwab W (2015) Functional characterization and substrate promiscuity of UGT71 glycosyltransferases from Strawberry (Fragaria x ananassa). Plant Cell Physiol 56: 2478-2493

Tarling CA, Woods K, Zhang R, Brastianos HC, Brayer GD, Andersen RJ, Wthers SG (2008) The search for novel human pancreatic alpha-amylase inhibitors: High-throughput screening of terrestrial and marine natural product extracts. Chembiochem 9: 433-438

Trapero A, Ahrazem O, Rubio-Moraga A, Jimeno ML, Gomez MD, Gomez-Gomez L (2012) Characterization of a glucosyltransferase enzyme involved in the formation of kaempferol and quercetin sophorosides in Crocus sativus. Plant Physiol 159: 1335-1354
Williams LK, Zhang X, Caner S, Tysoe C, Nguyen NT, Wicki J, Williams DE, Coleman J, McNeill JH, Yuen V, Andersen RJ, Withers SG, Brayer GD (2015) The amylase inhibitor montbretin A reveals a new glycosidase inhibition motif. Nat Chem Biol 11: 691-696

Wilson AE, Tian L (2019) Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. Plant J 100: 1273-1288

Witte S, Moco S, Vervoort J, Matern U, Martens S (2009) Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from Hieracium pilosella L. Planta 229: 1135-1146

Wonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi M, Kitajima M, Takayama H, Saito K (2014) A flavonoid 3-O-glucoside:2″-O-glucosyltransferase responsible for terminal modification of pollen-specific flavonols in Arabidopsis thaliana. Plant J 79: 769-782

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

Yonekura-Sakakibara K, Tohge T, Niida R, Saito K (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. J Biol Chem 282: 14932-14941

Yuen VG, Coleman J, Withers SG, Andersen RJ, Brayer GD, Mustafa S, McNeill JH (2016) Glucose lowering effect of montbretin A in Zucker Diabetic Fatty rats. Mol Cell Biochem 411: 373-381

Zong G, Fei S, Liu X, Li J, Gao Y, Yang X, Wang X, Shen Y (2019) Crystal structures of rhamnosyltransferase UGT89C1 from Arabidopsis thaliana reveal the molecular basis of sugar donor specificity for UDP-β-L-rhamnose and rhamnosylation mechanism. Plant J 99: 257-269