

Recruitment of distinct UDP-glycosyltransferase families demonstrates dynamic evolution of chemical defense within *Eucalyptus* L'Hér

Cecilie Cetti Hansen¹ , Mette Sørensen¹ , Matteo Bellucci² , Wolfgang Brandt³ , Carl Erik Olsen¹ , Jason Q. D. Goodger⁴ , Ian E. Woodrow⁵ , Birger Lindberg Møller¹  and Elizabeth H. J. Neilson¹ 

¹Plant Biochemistry Laboratory, Department of Plant and Environmental Science, University of Copenhagen, 1871, Frederiksberg C, Denmark; ²Novo Nordisk Foundation Center for Protein Research, Protein Production and Characterization Platform, University of Copenhagen, 2200, Copenhagen, Denmark; ³Department of Bioorganic Chemistry, Leibniz-Institute of Plant Biochemistry, Halle 06120, Germany; ⁴School of BioSciences, The University of Melbourne, Parkville, Vic. 3052, Australia; ⁵School of Ecosystem and Forest Sciences, The University of Melbourne, Parkville, Vic. 3052, Australia

Summary

Author for correspondence:
Elizabeth H. J. Neilson
Email: en@plen.ku.dk

Received: 6 August 2022
Accepted: 18 October 2022

New Phytologist (2023) 237: 999–1013
doi: 10.1111/nph.18581

Key words: chemical defense, cyanogenic glucoside, cytochrome P450, *Eucalyptus*, evolution, plant-specialized metabolism, UDP-glycosyltransferase, UGT87.

- The economic and ecologically important genus *Eucalyptus* is rich in structurally diverse specialized metabolites. While some specialized metabolite classes are highly prevalent across the genus, the cyanogenic glucoside prunasin is only produced by c. 3% of species.
- To investigate the evolutionary mechanisms behind prunasin biosynthesis in *Eucalyptus*, we compared *de novo* assembled transcriptomes, together with online resources between cyanogenic and acyanogenic species. Identified genes were characterized *in vivo* and *in vitro*.
- Pathway characterization of cyanogenic *Eucalyptus camphora* and *Eucalyptus yarraensis* showed for the first time that the final glucosylation step from mandelonitrile to prunasin is catalyzed by a novel UDP-glycosyltransferase UGT87. This step is typically catalyzed by a member of the UGT85 family, including in *Eucalyptus cladocalyx*. The upstream conversion of phenylalanine to mandelonitrile is catalyzed by three cytochrome P450 (CYP) enzymes from the CYP79, CYP706, and CYP71 families, as previously shown. Analysis of acyanogenic *Eucalyptus* species revealed the loss of different ortholog prunasin biosynthetic genes.
- The recruitment of UGTs from different families for prunasin biosynthesis in *Eucalyptus* demonstrates important pathway heterogeneities and unprecedented dynamic pathway evolution of chemical defense within a single genus. Overall, this study provides relevant insights into the tremendous adaptability of these long-lived trees.

Introduction

Plants have acquired exceptional abilities to adapt and adjust to their biotic and abiotic environments. This adaptability is particularly impressive in long-lived trees that must withstand changing environmental conditions and stresses over hundreds of years. Abiotic and biotic interactions are significantly mediated by the biosynthesis of highly diverse specialized metabolites, such as terpenoids, alkaloids, phenylpropanoids, and cyanogenic glucosides. These metabolites play many essential roles in plants, including chemical defense against herbivores and pathogens, resistance to extreme temperatures, and the attraction of pollinators. Plant evolution is repeatedly marked by the dynamic acquisition and loss of associated biosynthetic genes, resulting in a large and diverse chemical landscape across the plant kingdom, and in plants' abilities to occupy and withstand highly variable environmental conditions and niches.

Eucalyptus L'Hér is a hyper-diverse and large genus of over 700 species (Nicolle, 2019), with high economic and ecological relevance. *Eucalyptus* trees are the most widely planted commercial hardwoods in the world and display remarkable adaptability across native Australia and nearby islands occupying many diverse and sometimes extreme ecosystems (Bennett, 2016). Tandem gene duplications are suggested to play an important role in the diversification of eucalypts (Butler *et al.*, 2017), with tandem duplications constituting 34% of the predicted protein-coding gene sequences in the genome of *Eucalyptus grandis* (Myburg *et al.*, 2014). Notably, the *E. grandis* genome is particularly enriched in genes encoding terpene synthases (TPSs), cytochromes P450 (CYP), and UDP-glycosyltransferases (UGT), which are enzyme families that are significantly involved in the biosynthesis of specialized metabolites (Chae *et al.*, 2014; Myburg *et al.*, 2014; Kulheim *et al.*, 2015; Bustos-Segura *et al.*, 2017; Wilson & Tian, 2019; Hansen *et al.*, 2021). Indeed,

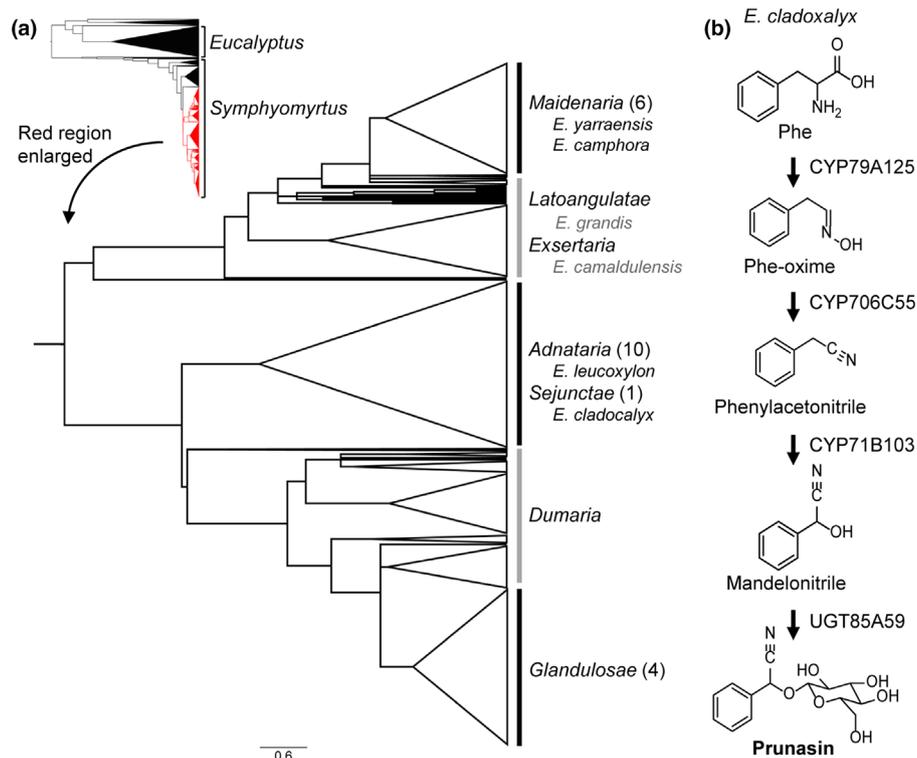


Fig. 1 Distribution and biosynthesis of cyanogenic glucosides in *Eucalyptus*. (a) Species phylogeny of the genus *Eucalyptus* adapted from Thornhill *et al.* (2019). The small phylogenetic tree shows the phylogeny of the entire *Eucalyptus* genus, where the two major subgenera (*Eucalyptus* and *Symphomyrtus*) are labeled. Subgenus *Eucalyptus* and the section *Bisectae* are colored black. The subtree marked in red is enlarged and consists of all sections belonging to *Symphomyrtus* except the large acyanogenic section *Bisectae*. *Eucalyptus* species are highly paraphyletic (Thornhill *et al.*, 2019), and the sections are labeled based on the most number of representative species for a given section. Sections encompassing cyanogenic species (black lines) and mayor sections without cyanogenic species (gray lines) are marked. The number in brackets denotes the number of identified cyanogenic species in a given section (Gleadow *et al.*, 2008). Cyanogenic and acyanogenic species studied in this article are written in black and gray, respectively. (b) Prunasin biosynthetic pathway in *Eucalyptus cladocalyx* (section *Sejunctae*) (Hansen *et al.*, 2018). phe, phenylalanine; phe-oxime, phenylacetaldoxime.

Eucalyptus species produce a corresponding panoply of specialized metabolites, with high intra and interspecies variation (Goodger & Woodrow, 2011; Neilson *et al.*, 2011; Bustos-Segura *et al.*, 2017; Marsh *et al.*, 2017; dos Santos *et al.*, 2019; Sørensen *et al.*, 2020). This variation is exemplified by the low proportion of species that display cyanogenic glucosides, with only some 3% of *Eucalyptus* species identified as cyanogenic (Gleadow *et al.*, 2008). All authenticated cyanogenic *Eucalyptus* species belong to the subgenus *Symphomyrtus* and are grouped into two distinct phylogenetic clusters: section *Maidenaria*; and sections *Adnataria*, *Sejunctae*, and *Glandulosae* (Gleadow *et al.*, 2008; Steane *et al.*, 2011). Most recent phylogenetic analysis of 732 species and subspecies indicates further stratification between the *Adnataria* and *Sejunctae* sections, and section *Glandulosae*, but resolution is challenged by a high degree of polyphyly (Fig. 1a; Thornhill *et al.*, 2019).

Cyanogenic glucosides are amino acid-derived α -hydroxynitrile glucosides (Gleadow & Møller, 2014). Upon tissue disruption and enzymatic hydrolysis, toxic hydrogen cyanide is liberated, thereby providing an effective chemical defense against generalist herbivores (Gleadow & Woodrow, 2002). In addition to defense, cyanogenic glucosides have increasingly been shown to provide reduced nitrogen for general metabolism via different recycling pathways (Jenrich *et al.*, 2007; Picmanova

et al., 2015; Bjarnholt *et al.*, 2018; Schmidt *et al.*, 2018). Cyanogenic *Eucalyptus* species produce L-Phe-derived cyanogenic glucosides, with prunasin being the major accumulating cyanogen. To date, characterization of prunasin biosynthesis in *Eucalyptus* has only been resolved for *Eucalyptus cladocalyx* (section *Sejunctae*), whereby prunasin is produced by the successive action of three CYP enzymes (CYP79A125, CYP706C55, and CYP71B103) and a UGT85A59 (Fig. 1b; Hansen *et al.*, 2018). Notably, this biosynthetic configuration contrasts with other characterized cyanogenic angiosperms, which only require two CYPs and a UGT (e.g. almond (*Prunus dulcis*)) (Thodberg *et al.*, 2018).

Given the distinct distribution of cyanogenic species within the subgenus *Symphomyrtus*, the atypical configuration of prunasin biosynthesis previously described for *E. cladocalyx*, and the known genetic complexity of the *Eucalyptus* genus in general, the functional characterization of cyanogenic glucoside biosynthesis within this genus constitutes an excellent model to investigate the dynamic evolution of specialized metabolism. We tested the following three hypotheses: (1) Due to the close relatedness within the subgenus *Symphomyrtus*, cyanogenic eucalypts would possess orthologous *CYP79A*, *CYP706C*, *CYP71B*, and *UGT85A* encoding genes to *E. cladocalyx* with conserved functionalities, (2) acyanogenic eucalypts would lack either one or more members of the genes encoding the biosynthetic enzymes in the prunasin

pathway, and/or (3) acyanogenic eucalypts possess orthologous genes, but their functionality is compromised. Here, we show that cyanogenic *Eucalyptus* species of the *Maidenaria* section have recruited a novel UGT87Y1 that catalyzes the glucosylation of mandelonitrile to prunasin. This represents the first UGT87 family member to be functionality-characterized in plants. We also show that the atypical three-CYP system is retained within cyanogenic *Eucalyptus* species, and through analysis of the acyanogenic species *E. camaldulensis* and *E. grandis*, and that there have been losses of orthologous prunasin biosynthetic genes in these species. Taken together, a highly dynamic evolution of chemical defense within the *Eucalyptus* genus is demonstrated, providing insights into the tremendous adaptability of these long-lived trees.

Materials and Methods

Plant material

Eucalyptus L'Hér seeds were purchased from CSIRO Australian Tree Seed Centre (Clayton, Vic., Australia). Seedlings were germinated in a glasshouse at the University of Copenhagen on 24 August 2013 and grown at minimum 22°C during the day (10-h light) and minimum 19°C at night. Leaf samples were harvested 1, 4, 7, and 10 months after sowing (33, 110, 215, and 285 d after sowing, respectively) from eight biological replicates. For the 1-month-aged seedlings, all plants were pooled to obtain sufficient leaf material. *Nicotiana benthamiana* Domin plants were grown in a glasshouse at 21°C : 19°C, day : night and infiltrated at c. 4 wk of age.

RNA extractions and transcriptomic analysis

Total RNA was extracted from a single representative expanding leaf from *Eucalyptus camphora* and *Eucalyptus yarraensis* plants at 4 and 7 months of age and a pooled sample at 1 month of age, using Spectrum™ Plant Total RNA Kit (Sigma) according to the manufacturer's instructions with a few modifications according to Hansen *et al.* (2018). The quantity and integrity of RNA was measured on an Agilent 2100 Bioanalyzer instrument using an RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions.

Transcriptomic sequencing (HiSeq2500, 100-bp paired-end) was performed by Macrogen Inc. (Seoul, Korea). Two independent transcriptome assemblies were combined for candidate gene identification. One assembly was performed by Sequentia Biotech (Bellaterra, Spain) as reported previously (Hansen *et al.*, 2018), using a minimum length setting of 35 bp and a minimum quality score of 35. The transcripts were clustered based on the coding sequence to remove redundancy (Cerveau & Jackson, 2016). In the second assembly, the transcriptomes were also *de novo* assembled using a genome-guided TRINITY pipeline (Haas *et al.*, 2013) after trimming the reads using ADAPTER-REMOVAL (Lindgreen, 2012), with a minimum length setting of 40 bp and a minimum quality score of 28. Expression levels were quantified using KALLISTO (Bray *et al.*, 2016). The first transcriptome assembly was used to guide UGT selection (Table S2),

while the second transcriptome was used to better quantify UGT expression due to a detailed assessment of intron retention and bicistronic assembly in the datasets.

Transient expression in *N. benthamiana*

The open reading frames (ORFs) of prunasin CYPs from *E. camphora* and *E. yarraensis* were cloned from cDNA. Primers used for the amplification of the nucleotide sequences are listed in Table S1. ORFs for genes from *E. grandis* and the UGT genes were ordered as synthetic constructs from Genscript (Piscataway, NJ, USA). The ORF of UGT87Y1 was in addition amplified from cDNA prepared from RNA from *E. camphora* and *E. yarraensis*, respectively.

Agrobacterium-mediated transient expression in *N. benthamiana* was carried out as described previously (Hansen *et al.*, 2018). Equal volumes of agrobacteria suspensions were introduced into *N. benthamiana* leaves using a blunt syringe, with an expression construct containing the *p19* gene (Hearne *et al.*, 1990) always co-infiltrated to prevent post-transcriptional gene silencing (Voinnet *et al.*, 1999; Lakatos *et al.*, 2004). Leaf disks (1.4 cm in diameter) were harvested 4–5 d after infiltration, snap-frozen in liquid nitrogen, and stored at –70°C until extraction. Due to the effect of leaf age and sampling position in the leaf on the variance in agroinfiltration experiments (Bashandy *et al.*, 2015), leaf disks were consistently harvested from the first expanded leaf, between the midrib and leaf edge. Leaf disks from minimum two replicate plants were harvested for each combination and verified in at least two independent experiments.

Metabolite extractions

Metabolites from frozen, homogenized *E. camphora* and *E. yarraensis* leaf tissue (15–50 mg) were extracted with cold aqueous 80% MeOH, while 85% MeOH was used for infiltrated *N. benthamiana* plants. Plant tissue was incubated on ice (5 min) and subsequently centrifuged (3 min, 10 000 g, 4°C). Extracts were diluted five times in MilliQ water and filtered through a Durapore® membrane with 22 µm pore size (Merck Millipore, Burlington, MA, USA).

Liquid chromatography tandem mass spectrometry

Extracts were analyzed either by an Agilent 1100 Series LC system coupled to a Bruker HCT-Ultra ion trap mass spectrometer (LC-ion trap) or using a Dionex Ultimate 3000RS UHPLC system (Thermo Fisher Scientific, Waltham, MA, USA) system coupled to a Compact™ (UPLC-QqToF) mass spectrometer (Bruker Daltonics, Bremen, Germany). Both instruments were equipped with an electrospray ionization source, DAD detector, and a temperature-controlled auto-sampler (10°C). Extracts of *Eucalyptus* and initial *N. benthamiana* transient expression experiments were analyzed according to the LC-ion trap methods reported in Hansen *et al.* (2018). Later, leaf extracts of *N. benthamiana* analyzed by UHPLC-QqToF were injected (2 µl) onto a Kinetex XB-C18 UHPLC column (150 × 2.1 mm, 1.7 µm; Phenomenex,

Torrance, CA, USA) at 40°C with a flow rate of 0.3 ml min⁻¹. The mobile phase A consisted of 0.05% formic acid in water, and solvent B consisted of 0.05% formic acid in acetonitrile. The initial composition was 95% A and 5% B, which was held for 1 min, followed by a linear gradient to 70% solvent B over 6 min, and a subsequent increase to 100% B over 1 min and held at 100% solvent B for 2 min. The QqToF mass spectrometer was operated in negative ion mode within a range of *m/z* 50–1200, and data-dependent MS/MS acquisition was triggered for the three most intense ions in the MS spectra. Liquid chromatography tandem mass spectrometry (LC-MS/MS) data were analyzed using the BRUKER DATA ANALYSIS 5.0 software. Prunasin was identified by comparison with an authentic standard (Møller *et al.*, 2016), and other compounds were putatively annotated based on their *m/z* value, MS2 fragmentation pattern, similar to the CFM-ID 4.0 library (for benzoic acid glucoside; Wang *et al.*, 2021) and comparison with the previous literature (Bak *et al.*, 2000; Kristensen *et al.*, 2005; Hansen *et al.*, 2018; Thodberg *et al.*, 2018).

Microbial protein expression and purification

Codon-optimized *UGT85A59* and *UGT87Y1* in pET-28a(+)-TEV were ordered from Genscript® and expressed in NiCo21 cells (New England Biolabs, Ipswich, MA, USA) in TB media with 50 µg ml⁻¹ kanamycin at 16°C for 20 h, 200 rpm, after induction with 0.5 mM IPTG. Cell pellets were sedimented by centrifugation (15 000 g, 15 min, 4°C) and stored at -20°C. Thawed cell pellets were suspended in 5 ml lysis buffer (100 mM Tris (pH 7.5), 300 mM NaCl, 1 mM MgCl₂, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), 0.04 lysozyme mg ml⁻¹, benzonase (25 U ml⁻¹; Sigma-Aldrich), 1 mM PMSF) per 1 g cells and lysed using a high-pressure homogenizer at 20 kPSI (Emulsi-Flex D20, Avestin, Ottawa, Canada). The lysate was cleared by centrifugation (9600 g, 40 min), filtered (0.45 µm pore size), and applied to an equilibrated 5 ml His-Trap FF column (GE Healthcare, Chicago, IL, USA) connected to an ÄKTAexpress (GE Healthcare). The column was washed with 10 column volumes (CV) of wash buffer (50 mM Na₃PO₄ (pH 7.5), 300 mM NaCl, 30 mM imidazole, and 0.5 mM TCEP). The protein was eluted using a combination of gradient and step elution: The column was washed with a linear gradient of 0–7.5% elution buffer (50 mM Na₃PO₄ (pH 7.5), 300 mM NaCl, 500 mM Imidazole, and 0.5 mM TCEP) followed by at least 2 CV at 7.5%. A second linear gradient of 7.5–25% for *UGT87Y1* and 7.5–35% for *UGT85A59* was followed by another plateau of at least 5 CV before a third linear increase to 100% elution buffer. The eluate was collected in 1.5 ml fractions, and fractions containing the target protein were pooled and concentrated to a final total volume of 6 ml by centrifugation using a pre-equilibrated membrane filter (30 kDa cut-off; Thermo Fisher Scientific). The concentrate was applied to an equilibrated HiLoad 16/60 Superdex 200 column (120 ml; GE Healthcare) connected to an ÄKTA Pure (GE Healthcare). Protein was eluted in 50 mM Na₃PO₄ (pH 7.5) with 100 mM NaCl, 0.5 mM TCEP, and 10% v/v glycerol. Fractions containing the target protein were combined and concentrated by centrifugation (30 kDa cutoff; Thermo Fisher Scientific). During the purification

process, selected fractions from all purification steps were analyzed by SDS-PAGE using 4–12% NuPAGE protein gels (Thermo Fisher Scientific). The final protein fraction was additionally analyzed by mass spectrometry on a micrOTOF-Q II instrument (Bruker Daltonics), which confirmed 87% purity for *UGT87Y1* and 92% purity for *UGT85A59* protein preparations. Isolated protein was frozen in liquid nitrogen and stored at -80°C.

Nicotiana benthamiana protein extract

Agrobacterium-infiltrated *N. benthamiana* leaves were harvested 4–5 d after infiltration and frozen in liquid nitrogen. Approximately 200 mg of homogenized leaf tissue was mixed with 0.1 g polyvinylpyrrolidone per gram fresh weight and 800 µl buffer composed of 100 mM Tricine (pH 7.9), 250 mM sucrose, 50 mM NaCl, 2 mM DTT, 1 tablet cOmplete™ protease inhibitor cocktail (Roche) per 50 ml buffer. Samples were centrifuged for 10 min at 15 000 g, 4°C. The supernatant was transferred to a clean tube and centrifuged again (10 min at 15 000 g, 4°C) with the resultant supernatant used for *in vitro* assays and Western blotting.

In vitro assays

The assay reactions consisted of 10 µl total protein extract or 1.2 µg purified protein, 0.2–10 mM of substrate, 8 µM ¹⁴C-UDP-glucose (0.8 µCi, specific activity 250 mCi mmol⁻¹; Perkin-Elmer, Waltham, MA, USA), and 50 mM Tris-HCl buffer (pH 7.5) with 100 mM NaCl in a total volume of 25 µl. Due to the instability of *p*-hydroxymandelonitrile and acetone cyanohydrin, these substrates were tested at 1 and 10 mM, respectively. Other substrates were tested at both 200 and 400 µM concentrations. Samples were incubated for 30 min at 30°C, 300 rpm, and then terminated by the addition of 2 µl 10% acetic acid. Samples were stored at -20°C until analysis by thin-layer chromatography (TLC). Eight microliter sample was applied to a TLC plate (Silica gel 60F₂₅₄ 0.2 mm thickness; Merck, Rahway, NJ, USA), and substrate and products were separated using a mobile phase consisting of EtOAc : Me₂CO : CH₂Cl₂ : MeOH : MilliQ H₂O in a ratio of 20 : 15 : 6 : 5 : 4 (v/v). Migration of prunasin was determined by the application of an unlabeled authentic standard and visualized by the UV absorbance. Analysis of ¹⁴C-labeled products was carried out by exposure of the TLC plate to a Storage Phosphor Screen (GE Healthcare) for 4 d followed by visualization using an Amersham Typhoon IP imager (GE Healthcare). The *in vitro* assays were conducted minimum two times.

Phylogenetic analysis

The resources listed in Table 1 were mined for CYP79, CYP706, CYP71, UGT85, and UGT87 sequences. Sequences were manually curated, and non-full-length sequences (i.e. containing deletions, insertions, and/or missing reads) were removed. CYPs involved in prunasin production and CYPs from *E. grandis* were kindly named by David Nelson according to the standardized CYP nomenclature system (Nelson, 2009). *EgUGT87* SNPs were assessed using the PHYTOZOME database (Goodstein

Table 1 Phylogenetic classification, authority, and source of data for *Eucalyptus* species investigated in this study. The classification follows Nicolle (2019).

Species and authority	Common name	Section	Series	Data source
<i>E. cladocalyx</i> F. von Müller	Sugar gum	<i>Sejunctae</i>	–	Hansen <i>et al.</i> (2018)
<i>E. leucoxydon</i> Müller	Yellow gum	<i>Adnataria</i>	Meliiodorae	ONEKP (sample AYMT)
<i>E. camaldulensis</i> Dehnh.	River red gum	<i>Exsertaria</i>	Exertae subser. Rostratae	<i>E. camaldulensis</i> Genome Database
<i>E. grandis</i> W. Hill ex Maiden	Flooded gum, rose gum	<i>Latoangulatae</i>	Transversae	PHYTOZOME
<i>E. camphora</i> R.T. Baker	Mountain swamp gum	<i>Maidenaria</i>	Foveolatae	This study
<i>E. yarraensis</i> Maiden & Cambage	Yarra Gum	<i>Maidenaria</i>	Foveolatae	This study

et al., 2012; Myburg *et al.*, 2014). Amino acid sequences were aligned in MEGA7 (Kumar *et al.*, 2016) using MUSCLE with default settings, and maximum likelihood trees were generated with $n = 1000$ bootstrap replicates.

Protein modeling

Homology modeling was performed with automated modeling integrated in Yet Another Scientific Artificial Reality Application (YASARA, v.16.4.16) (Krieger *et al.*, 2009; Krieger & Vriend, 2014, 2015). Nineteen homology models were created based on five templates (2PQ6 (Li *et al.*, 2007), 5U6S (George Thompson *et al.*, 2017), 2ACV (Shao *et al.*, 2005), and 2VG8 and 2VCH (Brazier-Hicks *et al.*, 2007)). A hybrid model, based mainly on the X-ray structure of 2VG8, with some better folded fragments from 2ACV and 2PQ6, was used for further modeling studies. The ligand, uridine-5'-diphosphate (UDP), was overtaken in the hybrid model based on their position in the X-ray structure of 2VG8. An alignment of the *Egr*UGT87Y with the sequence of UGT87Y1 showed 97.6% sequence identity and an almost identical model structure. The donor sugar was manually added to UDP while keeping a fixed protein structure and energy optimized with AMBER-EHT force field (Kumar *et al.*, 2016). One hundred docking runs were performed for mandelonitrile using MOE with the London dG fitness function (Molecular Operating Environment 2021; 2019.01; Chemical Computing Group ULC, Montreal, Canada). The fifth best docked arrangement fulfilled the requirements for an optimal catalytic reaction by forming a hydrogen bond of the hydroxyl group of mandelonitrile to the catalytic active histidine (His-21), which is part of the catalytic dyad with Asp-113, and by having a short distance of 3.8 Å between the carbon atoms of the ligand and the C1 of the sugar donor (dotted red line in Fig. 6b).

Results

The three-CYP system is retained in cyanogenic *Eucalyptus*

To compare the biosynthetic genes between different phylogenetic groups, prunasin biosynthesis was investigated in *E. camphora* and *E. yarraensis* from section *Maidenaria*, species known to accumulate this compound (Goodger *et al.*, 2002, 2007; Neilson *et al.*, 2006). Prunasin accumulation in expanding leaves from plants at 1, 4, 7, and 10 months of age was measured by LC-MS/MS (Fig. 2a). Prunasin was detected at all four ages in *E. yarraensis*

with up to 30 $\mu\text{g mg}^{-1}$ DW at 7 months of age. In *E. camphora*, the onset of prunasin formation occurred at a later age and with significantly lower quantities than in *E. yarraensis* (*c.* 100-fold less). Similar to *E. yarraensis*, the highest prunasin accumulation occurred in *E. camphora* plants at 7 months of age (0.36 $\mu\text{g mg}^{-1}$ DW). Compared with the high prunasin accumulation in *E. cladocalyx* at an equivalent age (Gleadow & Woodrow, 2000; Hansen *et al.*, 2018), *E. yarraensis* and *E. camphora* can be regarded as intermediate and low producers of prunasin, respectively.

To identify the biosynthetic genes responsible for prunasin production, leaves of three different ages (1, 4, and 7 months) were selected for transcriptomic analysis. CYP79A, CYP706C, and CYP71B transcripts, with expression patterns matching prunasin accumulation, were identified from *E. camphora* and *E. yarraensis* (Fig. 2b). The three candidate CYP genes from *E. yarraensis* showed higher expression than the candidates from *E. camphora*, consistent with the prunasin levels. The orthologous CYPs from *E. camphora* and *E. yarraensis* showed >99% amino acid sequence identity to each other and >94% sequence identity to the respective CYP sequences from *E. cladocalyx*. The CYPs from *E. yarraensis* and *E. camphora* were assigned the names CYP79A34, CYP706C55, and CYP71B103 according to the standardized CYP nomenclature system (Nelson, 2009). A previous study identified four CYP79 genes from *E. yarraensis* (Neilson, 2012), but only a single CYP79 sequence was present in our transcriptomic dataset.

The full-length CYP genes from *E. camphora* and *E. yarraensis* were functionally characterized by transient expression in *N. benthamiana*. Co-expression of *Ey*CYP79A34, *Ey*CYP706C55, and *Ey*CYP71B103 or *Eca*CYP79A34, *Eca*CYP706C55, and *Eca*CYP71B103, together with *UGT85A59* from *E. cladocalyx*, resulted in accumulation of prunasin in *N. benthamiana* leaves (Fig. 2c). Akin to similar studies (Hansen *et al.*, 2018; Thodberg *et al.*, 2018), when the prunasin CYPs were transiently expressed, some prunasin accumulated in the *N. benthamiana* leaves due to endogenous UGT activity (Fig. 2c). Putatively annotated identified malonated prunasin and other glucosylated and malonated pathway intermediates were also observed due to the presence of endogenous *N. benthamiana* enzyme activities (Fig. S1). In addition, a diagnostic peak with m/z 329.0873 $[\text{M} + \text{FA}]^-$ was identified, putatively annotated as benzoic acid glucoside and likely derived from the dissociation of mandelonitrile (Figs S2, S3). A drop-out experiment in which each CYP was omitted from the assay system confirmed that all three CYPs are necessary for prunasin production (Fig. S2). This indicates that the three-CYP

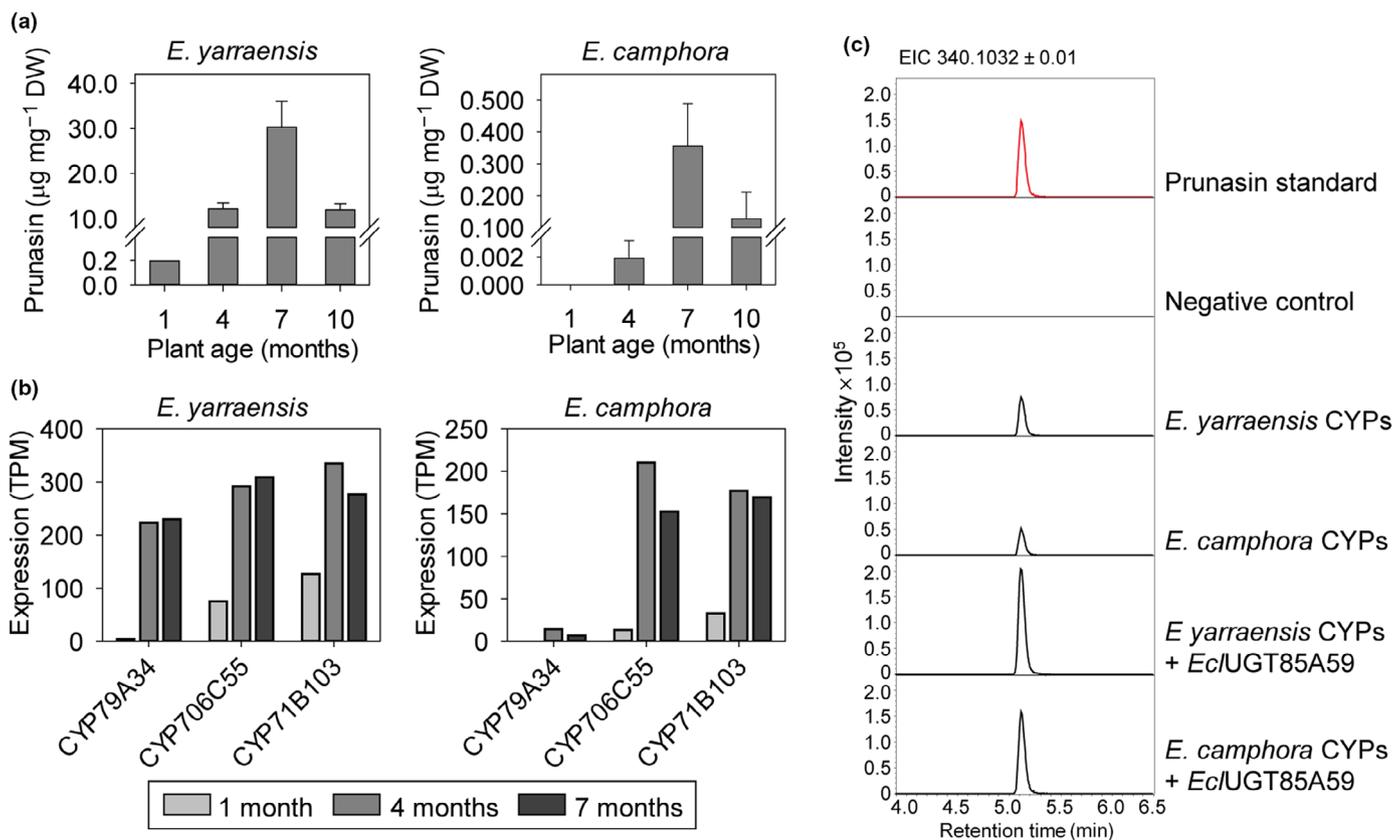


Fig. 2 Prunasin accumulation in *Eucalyptus yarraensis* and *Eucalyptus camphora*, and CYPs involved in prunasin production. (a) Mean prunasin content (\pm 1 SE) in *E. yarraensis* and *E. camphora* seedlings at 1, 4, 7, and 10 months of age ($n = 8$). (b) Transcript profiles of prunasin biosynthetic CYP genes in representative *E. yarraensis* and *E. camphora* individual tissue at different ontogenetic stages. Prunasin and transcript values at 1 month represent a pooled sample due to small plant size. (c) Extracted ion chromatograms (EIC) corresponding to prunasin (m/z 340.1032 [$M + \text{FA}$] $^-$) of leaf extracts from *Nicotiana benthamiana* plants transiently expressing the three candidate CYP genes from *E. yarraensis* and *E. camphora* with the UGT85A59 from *Eucalyptus cladocalyx*. DW, dry weight; TPM, transcripts per million.

system is conserved in subgenus *Symphyomyrtus* in contrast to the two-CYP system found in other flowering plant species.

Recruitment of a distinct UDP-glucosyltransferase for prunasin production in *E. camphora* and *E. yarraensis*

No *UGT85A59* ortholog was expressed in *E. camphora* and *E. yarraensis* at any of the three analyzed ages, and expression of other *UGT85s* showed either overall low expression or an expression pattern contrasting with the expression of the *CYPs* (Fig. 3a). A set of criteria were therefore defined to guide the selection of candidate UGTs for functional analysis: (1) The expression pattern should parallel prunasin accumulation or show overall high expression at all time points given that UGTs in cyanogenic glucoside biosynthesis do not necessarily co-express with the *CYPs* (Hansen *et al.*, 2018); (2) the UGT should be present in both *E. camphora* and *E. yarraensis* due to the close relationship of these species to each other; and (3) no ortholog match in the transcriptome from *E. cladocalyx* was expected since this species has *UGT85A59* that glucosylates mandelonitrile. These selection criteria resulted in three candidates from the families UGT88, UGT87, and UGT74, in addition to two *UGT85s* to include members of this family with the best matching expression pattern (Fig. 3a). These candidate

UGT genes were transiently co-expressed with the prunasin *CYPs* in *N. benthamiana* (Fig. 3b). When the *UGT85A59* from *E. cladocalyx* was co-expressed together with the *CYPs* as a positive control, prunasin levels were consistently 2.1–2.5 times higher, and we also observed the formation of prunasin malonate ester and the cyanogenic diglucoside amygdalin (Fig. S1). By contrast, markedly lower levels (18 times less) of the diagnostic m/z 329.0872 ion were detected when these four prunasin pathway members were co-expressed. As the pattern of m/z 329.0873 ion accumulation was consistently observed across multiple experiments with and without co-expression of a UGT candidate, the m/z 329.0873 ion was used to assess fluctuations in endogenous *N. benthamiana* UGT activity toward mandelonitrile. Accordingly, candidate UGT activity was assessed by total prunasin accumulation and by peak area ratio between prunasin and the m/z 329.0873 ion. The ratio calculation method showed consistent results across three independent experiments (Fig. S4). The inclusion of additional pathway derivatives (m/z 711 and m/z 761) in the calculation did not alter activity patterns across independent experiments (Fig. S4).

Upon co-expression with the five candidate UGTs, only co-expression of an *EyUGT87* resulted in high accumulation of prunasin in *N. benthamiana* leaves (Fig. 3b,c). Since the sequences of *EyUGT87Y* and *EcaUGT87Y* are identical at the amino acid

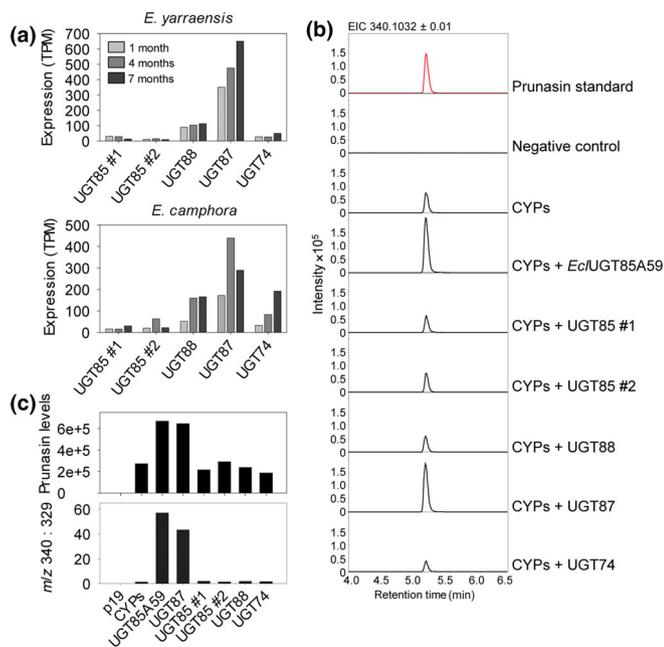


Fig. 3 UGT87 is involved in prunasin production in *Eucalyptus yarraensis* and *Eucalyptus camphora*. (a) Transcript profiles of UGT candidate genes from *E. yarraensis* and *E. camphora* ($n = 1$ per species, at 1, 4, and 7 months of age). UGT candidates were selected based on expression criteria within and between the two species. (b) Representative extracted ion chromatograms (EIC) corresponding to prunasin (m/z 340.1032 $[M + FA]^-$) of leaf extracts from *Nicotiana benthamiana* plants transiently expressing the *E. yarraensis* and *E. camphora* UGT candidate genes together with the prunasin *EyCYP79A34*, *EcCYP706C55*, and *EcCYP71B103*. (c) Corresponding levels of prunasin in *N. benthamiana* leaf extracts measured by peak area integration of prunasin (m/z 340; upper panel), and prunasin levels normalized to the pathway derivative putatively annotated as benzoic acid glucoside (m/z 329), shown here as a ratio (m/z 340 : 329; lower panel). All tested UGT candidates and the positive control UGT85A59 are co-expressed with the prunasin CYPs. TPM, transcripts per million.

level, only one *UGT87Y* gene was included in the experiments. Calculation of the prunasin to m/z 329.0872 ratio supported that only *EyUGT87* displayed mandelonitrile glucosylation activity above endogenous *N. benthamiana* UGT activity (Fig. 3c). Accordingly, while *E. cladocalyx* has a UGT85 that can glucosylate mandelonitrile, *E. camphora* and *E. yarraensis* have recruited a UGT87, which can catalyze the same reaction. This demonstrates that UGTs from distinct families and thus sharing < 45% amino acid sequence identity can glucosylate mandelonitrile (Fig. S5). The functionally active UGT87 was named UGT87Y1 by the UGT Nomenclature Committee (Burchell *et al.*, 1991; Mackenzie *et al.*, 1997, 2005).

To further characterize the activity of UGT87Y1 and UGT85A59, codon-optimized genes were heterologously expressed in *Escherichia coli* and the enzymes were purified for *in vitro* analysis using ^{14}C -UDP-glucose as the donor substrate. Both enzymes catalyzed the glucosylation of mandelonitrile, confirming the activity observed upon transient expression in *N. benthamiana* (Fig. S6). Substrate specificity of the two UGTs was investigated by testing other small metabolites as acceptors. UGT85A59 and UGT87Y1 also glucosylated benzoic acid but to a lower extent compared with mandelonitrile (Fig. S6).

UGT85A59 showed some activity toward the slightly larger cyanohydrin *p*-hydroxymandelonitrile, while UGT87Y1 did not glucosylate this compound. By contrast, UGT87Y1 glucosylated the smaller acetone cyanohydrin, whereas UGT85A59 did not. No activity was observed with gallic acid, coumaric acid, caffeic acid, ferulic acid, quercetin, genistein, naringenin, or kaempfeol.

Loss and gain of prunasin biosynthetic genes

To gain a broader understanding of prunasin evolution in the genus, we mined transcriptome and genome data from online sources available for three cyanogenic and acyanogenic *Eucalyptus* species (Table 1). A transcriptome from the cyanogenic species *Eucalyptus leucoxylo* (section *Adnataria*) is available from the ONEKP database (Carpenter *et al.*, 2019). A good-quality genome is available for the acyanogenic *E. grandis* (section *Latoangulatae*) (Myburg *et al.*, 2014), and a shallow genome library is available for the acyanogenic *E. camaldulensis* (section *Exsertaria*) (Hirakawa *et al.*, 2011).

In line with our hypotheses, orthologs to *EcCYP79A125*, *EcCYP706C55*, *EcCYP71B103*, and *EcUGT85A59* were found in the transcriptomic dataset from cyanogenic *E. leucoxylo* (Figs S5, S7–S9), consistent with *E. leucoxylo* and *E. cladocalyx* belonging to the same cyanogenic cluster in the phylogeny (Fig. 1a). The sequences share > 96% identity, with the CYP79 orthologs being most identical (99%) and the CYP71 orthologs being least identical (96%).

Assessment of the two acyanogenic species, *E. camaldulensis* and *E. grandis*, revealed different outcomes of prunasin biosynthesis gene conservation or loss. The *E. camaldulensis* genome harbors nine CYP79s, including a full-length ortholog (*Ec018270*) to *EyCYP79A34*, sharing 97.7% amino acid sequence identity (Fig. S7). In addition, a full-length ortholog (*EcC047195*) to *EcUGT85A59*, sharing 92% amino acid sequence identity, was also identified (Fig. S5). Only gene fragments with high identity to *EcCYP706C55* and *EcCYP71B103* were present in the *E. camaldulensis* genome, and no clear orthologs to UGT87Y1 were identified. These findings suggest that the lack of prunasin accumulation in *E. camaldulensis* is due to the lack of full-length CYP genes involved in oxime to cyanohydrin conversion, consistent with hypothesis 2.

In contrast to *E. camaldulensis*, several full-length prunasin pathway orthologs were identified in the *E. grandis* genome. A potential ortholog to *EyCYP79A34* was a pseudogenized CYP79 gene with a 9.76-kb insertion and a premature stop codon, but a total of 11 other full-length CYP79s were identified in *E. grandis*. Two of the full-length *E. grandis* CYP79s (*EgCYP79A37* and *EgCYP79A36*) show high identity (> 85%) to *EcCYP79A125* at the amino acid level and are apparent orthologous genes to *EyCYP79A36* and *EyCYP79A37*, respectively (Fig. S7). Full-length orthologs to *EcCYP706C55* (*EgCYP706C55*) and *EcCYP71B103* (*EgCYP71B115*) are also present in the *E. grandis* genome (Figs S8, S9). Interestingly, an ortholog to *EcUGT85A59* (*EgUGT85A*, > 92% identity) and an ortholog to UGT87Y1 (*EgUGT87Y*, > 97% identity) were identified (Fig. S5). Low but possible co-expression of *EgCYP79A37*,

EgCYP706C55, *EgCYP71B115*, and *EgUGT87Y* was detected in shoot tips and reproductive tissue, with no reported co-expression of *EgUGT85A* (Vining *et al.*, 2015).

To determine whether these homolog genes in *E. grandis* had the ability to produce prunasin, their functional activity was determined using transient expression in *N. benthamiana*. Functionality assessment of the two closest homologs to *EjCYP79A34* (*EgCYP79A37* and *EgCYP79A36*) by co-expression of downstream biosynthetic genes from *E. yarraensis* only resulted in traces of prunasin (Fig. 4), while targeted mining of the data identified a

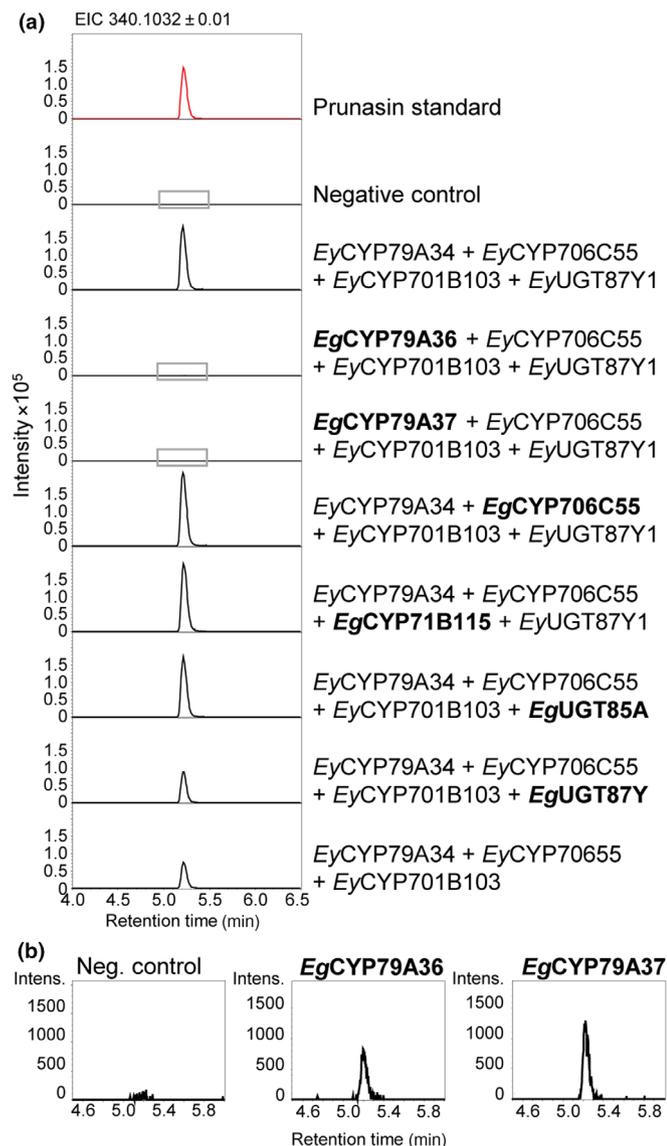


Fig. 4 Transient expression of homologous prunasin biosynthetic genes from acyanogenic *Eucalyptus grandis*. (a) Extracted ion chromatograms (EIC) of *Nicotiana benthamiana* leaf extracts prepared from plants transiently co-expressing homologs of prunasin biosynthetic genes from *E. grandis* (marked in bold) with the prunasin biosynthetic genes from *Eucalyptus yarraensis*. The chromatograms surrounded by a gray box for the combinations co-expressing *EgCYP79A36* and *EgCYP79A37*, respectively, are enlarged in panel B. (b) Zoom on two chromatograms from panel A where the CYP79s *EgCYP79A36* and *EgCYP79A37* are each co-expressed with downstream prunasin biosynthetic genes.

small peak with a *m/z* 308.1345 corresponding to a leucine/isoleucine oxime glucoside (Fig. S10). Substitutions of pathway members from *E. yarraensis* with *EgCYP706C55*, *EgCYP71B115*, or *EgUGT85A* resulted in prunasin accumulation in *N. benthamiana* leaves (Fig. 4a), suggesting that a CYP79 with phe-oxime biosynthetic activity is the limiting factor for prunasin production in *E. grandis*. In a separate experiment where *EgUGT85A* was co-expressed with the CYPs from *E. cladocalyx* and extracts analyzed on a LC-ion trap instrument, *EgUGT85A* co-expression resulted in more putatively glucosylated intermediates than *EcUGT85A59* (Fig. S11), suggesting that *EgUGT85A* is less specific toward mandelonitrile than *EcUGT85A59*. Transient co-expression of *EgUGT87Y* with the prunasin CYPs only resulted in prunasin levels slightly above background levels (1.4 times), with compromised *EgUGT87Y* activity also supported when normalized to *m/z* 329 (Figs 4a, 6c,d).

These data suggest that *E. grandis* has potential, but limited capacity to biosynthesize prunasin. Functional prunasin production, however, is constrained by several factors in line with our hypotheses: lack of a functional CYP79A34 ortholog (due to gene pseudogenization), limited activity of paralogous CYP79s, compromised activity of an orthologous *EgUGT87Y*, and the lack of *EgUGT85A* expression (Vining *et al.*, 2015). A summary of the prunasin pathway orthologs present in the analyzed species is shown in Fig. 5.

Site-directed mutagenesis of UGT87

Due to high amino acid sequence identity (97.6%) between the functional and functionally compromised UGT87Ys from *E. yarraensis*, *E. camphora* and *E. grandis*, respectively, key residues for mandelonitrile glucosylation were identified and tested. Homology modeling of UGT87Y1 showed that the phenyl ring

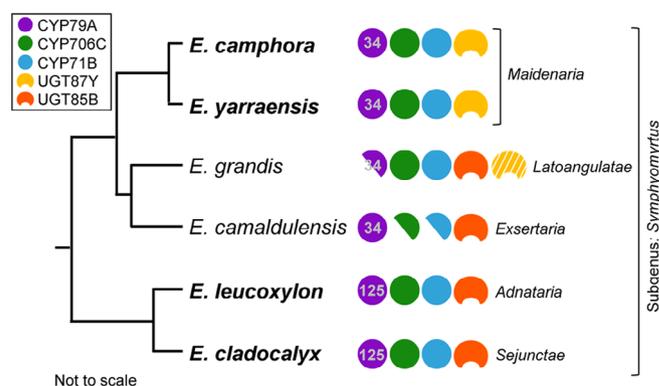


Fig. 5 Evolution of prunasin biosynthesis in *Eucalyptus* showing conservation of the three-CYP system but functional recruitment of two distinct UGT families. CYP79 subfamily names are denoted, showing divergence across sections. *Eucalyptus camphora*, *Eucalyptus yarraensis*, *Eucalyptus leucoxyton*, and *Eucalyptus cladocalyx* are cyanogenic (highlighted in bold). Ortholog genes identified in the acyanogenic *Eucalyptus grandis* and *Eucalyptus camaldulensis* are shown. The half circles illustrate pseudogenes. The ortholog CYP706C, CYP71B, and UGT85A from *E. grandis* demonstrated functional activity for prunasin production, while the ortholog UGT87Y showed severely compromised activity (illustrated with white stripes).

of mandelonitrile is primarily positioned in the active site by hydrophobic interactions with Phe-16, Phe-81, and Tyr-115. The rings in Phe-16 and Phe-81 may also form π stacks with the phenyl ring in mandelonitrile. Comparison between UGT87Y1 and *Eg*UGT87 identified two residues localized close to the catalytic His-21 that differed: a substitution of Phe-16 to a Tyr residue, and Gly-18 to an Ala (Fig. 6a). No SNPs are reported in these two amino acid positions for 36 sequenced *E. grandis* individuals (4). Modeled mutations of Y16F and G18A in *Eg*UGT87 resulted in enhanced hydrophobic interactions with the phenyl ring of mandelonitrile (Fig. 6b).

To investigate the importance of these residues on catalytic activity, mutants Y16F, G18A, and Y16F + G18A were generated for *Eg*UGT87. The mutant genes were transiently co-expressed with the prunasin CYPs in *N. benthamiana*. LC-MS/MS analysis of

infiltrated leaves showed increased prunasin levels for mutants Y16F and Y16F + G18A compared with *Eg*UGT87Y but with varying levels across three independent experiments (Figs 6c,d, S4). These data suggest partial to full recovery of activity. However, when prunasin levels were normalized to the levels of the *m/z* 329 ion, only partial recovery of activity was observed for the mutant Y16F and Y16F + G18A (Figs 6d, S4), with the normalized data showing a consistent pattern across the three experiments. To independently test the activity of the *E. grandis* enzyme mutants, we performed *in vitro* assays using total protein extracts from *N. benthamiana* leaves and 14 C-UDP-glucose and mandelonitrile as substrates (Fig. 6e). In this experiment, UGT87 activity levels showed a pattern highly similar to the normalized prunasin levels. These data support the reliability of the *m/z* 340 : 329 normalization method. The increased prunasin accumulation observed for the

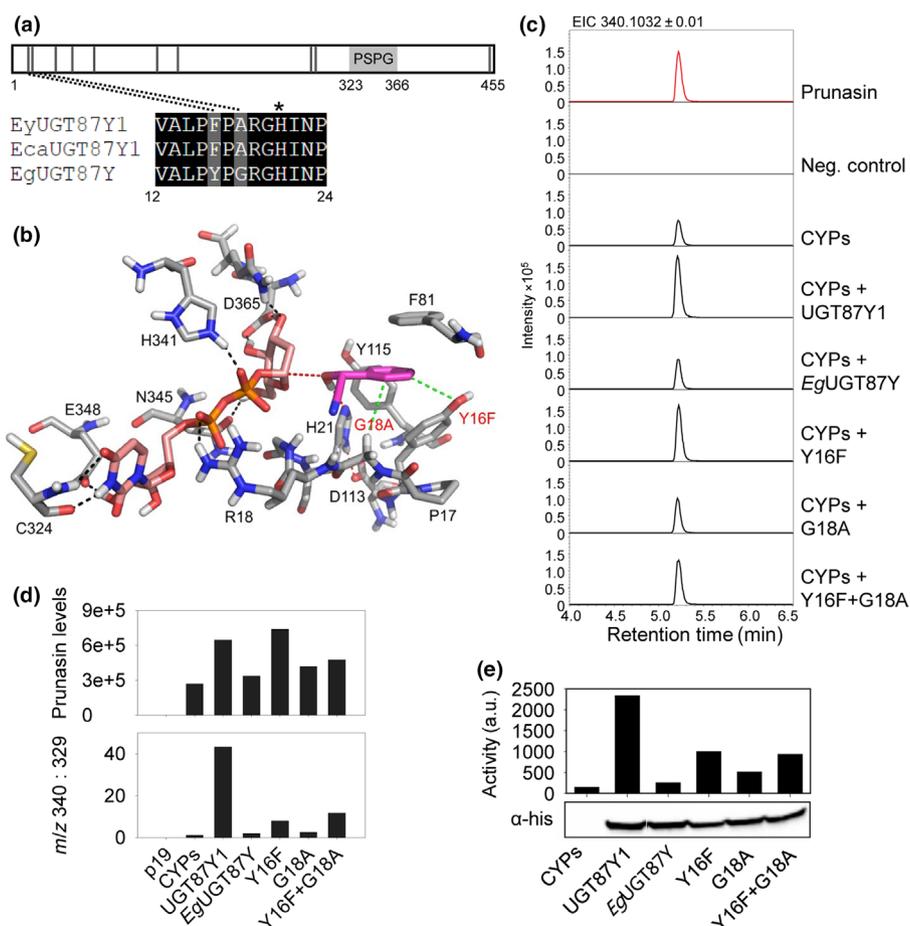


Fig. 6 Site-directed mutagenesis of *Eg*UGT87. (a) Schematic of the UGT87 amino acid sequence indicating the positions of amino acids (black vertical lines) that differ between UGT87Y1 and *Eg*UGT87Y. The PSPG motif is marked with a light gray box. The catalytic His is marked with an asterisk. (b) Active site of *Eg*UGT87 (Eucgr.B03993) (grey carbon atoms) with docked mandelonitrile (magenta carbon atoms) and UDP-glucose (pink carbon atoms). Nitrogen, oxygen, sulfur and phosphorus atoms are shown in blue, red, yellow and orange, respectively. The benzene moiety of the substrate is mainly recognized by hydrophobic interactions with Y16, F81, and Y115. Mutations of Y16F and G18A (red text) enhance the hydrophobic interactions (green dotted lines) with the benzene ring of mandelonitrile. The catalytic dyad H21 - D113 is shown (red dotted lines) and hydrogen bonds indicated (black dotted lines). (c) Representative extracted ion chromatograms (EIC) corresponding to prunasin (*m/z* 340.1032 [M + FA]⁻) of leaf extracts from *Nicotiana benthamiana* plants transiently expressing *Eucalyptus* UGTs together with the prunasin CYPs from *Eucalyptus yarraensis*. (d) Prunasin accumulation in *N. benthamiana* leaf extracts measured by peak area integration of prunasin (*m/z* 340; upper panel), and prunasin levels normalized to the pathway derivative putatively annotated as benzoic acid glucoside (*m/z* 329), shown here as a ratio (*m/z* 340 : 329; lower panel). (e) *In vitro* assays using protein extracts prepared from *N. benthamiana* leaves transiently expressing prunasin CYPs and the UGT variants (top) and Western blot of the His-tagged UGTs in the protein extract that was used for activity assay (bottom). The UGT activity is plotted as the intensity of the prunasin signal on a TLC plate measured by phospho-imaging.

*Eg*UGT87Y mutants, but not the *m/z* 340 : 329 ratio, suggests possible stimulation of prunasin CYP expression by the UGT enzymes *in vivo* and a higher flux toward mandelonitrile production.

The partial recovery of the mutants Y16F and Y16F + G18A suggests that other amino acid differences between UGT87Y1 and *Eg*UGT87 are important for catalytic activity. The other residues that differ between *Eg*UGT87 and UGT87Y1 are not localized within the catalytic pocket or in the conserved PSPG motif and may be involved in the stabilization of the enzyme structure. For example, the homology model shows that the differences between R77Q and S139P cause a slightly altered structure, potentially causing a different conformation of K79, which is positioned close to the substrate entry channel.

Discussion

Dynamic and conserved features of prunasin CYP evolution

Cyanogenic *Eucalyptus* species within the subgenus *Symphomyrtus* group into distinct phylogenetic clades. The evolutionary genetic patterns behind prunasin biosynthesis were achieved using a metabolite-guided transcriptomic approach in the species *E. yarraensis* and *E. camphora*. Both species contain the functional CYP79A, CYP706C, and CYP71B necessary for prunasin production (Figs 2, S2), similar to *E. cladocalyx* (23). *Eucalyptus* species that do not accumulate cyanogenic glucosides (i.e. *E. grandis* and *E. camaldulensis*) were characterized by an apparent lack of either one or two full-length CYP orthologs associated with prunasin biosynthesis (Fig. 5).

CYP79s are the entry point to cyanogenic glucoside production in seed plants, and regulation of gene expression and enzyme activity thus control the flux into the pathway (Busk & Møller, 2002; Thodberg *et al.*, 2018). In agreement with this, *CYP79A34* from *E. camphora* and *E. yarraensis* was the lowest expressed pathway member, which was also observed for *Ecl-CYP79A125* (Hansen *et al.*, 2018). Moreover, the expression of *EcaCYP79A34* was markedly lower than *EjCYP79A34*, consistent with significantly lower prunasin levels in *E. camphora* (Fig. 2).

Several CYP79s were identified in the acyanogenic species *E. grandis* and *E. camaldulensis*, with *E. grandis* harboring among the highest CYP79 gene copy number within diploid species sequenced to date (Fig. S7; Nelson, 2009). The physiological functions of these CYP79s are unknown. Published expression data from various *E. grandis* tissues show that the CYP79s are either expressed at low levels or not at all (Vining *et al.*, 2015), although it is possible that CYP79s are expressed at different developmental stages or induced upon exposure to environmental factors. Two of the CYP79s from *E. grandis* (*EgCYP79A36* and *EgCYP79A37*) showed trace activity with L-Phe and potentially L-Leu/L-Ile upon transient expression in *N. benthamiana* (Fig. 4). While we did not detect other amino acid-derived oxime glucosides in this study, L-Phe, L-Leu, L-Ile, L-Trp, and L-Tyr have been identified as substrates for CYP79s present in the acyanogenic tree *Populus trichocarpa* (Irmisch *et al.*, 2013). In *P. trichocarpa*, the formation of volatile oximes is induced upon herbivore

damage. It is also possible that CYP79s may perform N-hydroxylation reactions on other substrates beyond amino acids, and the high copy number and diversity of eucalypt CYP79s provide an interesting system to study the role of this CYP gene family in more detail.

In other seed plants, the conversion of an amino acid-derived oxime into the corresponding cyanohydrin in cyanogenic glucoside biosynthesis is performed by a single, multifunctional CYP usually from the 71 family (Bak *et al.*, 1998; Jørgensen *et al.*, 2011; Yamaguchi *et al.*, 2014; Thodberg *et al.*, 2018; Lai *et al.*, 2020). By contrast, this step is catalyzed by two distinct CYPs, a CYP706C and a CYP71B, in *Eucalyptus*. The functional characterization of a CYP79A, CYP706C, and CYP71B in different *Eucalyptus* species shows that the three-CYP system is conserved within the subgenus *Symphomyrtus*, thereby requiring a total of four biosynthetic enzymes for prunasin production. Recent work suggests that a four-step pathway for prunasin biosynthesis is also operating in the cyanogenic fern *Phlebodium aureum* (Thodberg *et al.*, 2021), demonstrating that this biosynthetic configuration is not unique to *Eucalyptus*.

Repeated recruitment of UGTs for two-component chemical defense systems in plants

In contrast to the conservation of CYPs in prunasin biosynthesis, we found evidence of novel UGT87Y1 recruitment for mandelonitrile glucosylation within the *Maidenaria* section. Measurement of UGT87Y1 activity above endogenous *N. benthamiana* enzyme levels was facilitated by calculating a ratio between prunasin and a pathway derivative formed upon the expression of the prunasin CYPs (Fig. S3). This is the first time that a UGT87 has been shown to be involved in cyanogenic glucoside biosynthesis and the first member of the UGT87 family to be functionally characterized. *UGT87Y1* expression showed a pattern similar to CYP expression and was the highest expressed pathway gene in both species at all time points (Figs 2, 3), supporting a role in cyanohydrin stabilization.

Very little is known about the UGT87 family in plants. UGT87A2 from *Arabidopsis thaliana* was reported to regulate flowering time (Wang *et al.*, 2012) and play a role in adaptation to abiotic stress including salt tolerance (Li *et al.*, 2017). A UGT87A from *Carex rigescens* has also been linked to increased salt tolerance (Zhang *et al.*, 2021). Several metabolite acceptors for UGT87A2 have been experimentally tested *in vitro* including abscisic acid, auxins, cytokinins, various phenylpropanoids, and benzoates, but the *in planta* substrate has not yet been identified (Lim *et al.*, 2002; Li *et al.*, 2017).

Similar to the glucosylation of mandelonitrile by *Ecl*UGT85A59, the glucosylation of a cyanohydrin to form the cyanogenic glucoside is typically catalyzed by UGT85 members in various angiosperms (Franks *et al.*, 2008; Kannangara *et al.*, 2011; Takos *et al.*, 2011). A recent exception is a UGT94AF3 from almond that was able to glucosylate mandelonitrile (Thodberg *et al.*, 2018) in addition to the previously characterized UGT85A19 (Franks *et al.*, 2008). However, substrate specificity of UGT94AF3 was not examined and its

physiological function awaits further investigation. Most of the cyanogenic UGT85s belong to distinct phylogenetic subfamilies demonstrating that homologous, but not orthologous, *UGT* genes have been repeatedly recruited for cyanogenic glucoside biosynthesis (Fig. S5). This is similar to the CYP79 and CYP71 members involved in cyanogenic glucoside biosynthesis in different plant species (Takos *et al.*, 2011).

Beyond the biosynthesis of cyanogenic glucosides, UGTs from different families able to catalyze the same reaction are also known from plants with other classes of specialized metabolites, including benzoxazinoids. Benzoxazinoids are a small class of defense compounds found in multiple monocot Poaceae species including maize (*Zea mays*) and wheat (*Triticum aestivum*), and a few dicot species (e.g. *Consolida orientalis*) (Sicker *et al.*, 2000). Glucosylation of the benzoxazinoid DIBOA is catalyzed by UGT710Es in Poaceae species and by UGT85N1 in *C. orientalis* (von Rad *et al.*, 2001; Dick *et al.*, 2012). Together, these examples point to the remarkable plasticity of UGTs that have been recruited for lineage-specific functions.

Eucalyptus contains a rich repertoire of UGTs

UDP-glucosyltransferases constitute a large gene family in plants, often with > 100 members present in angiosperm species (Wilson & Tian, 2019). A phylogenomic analysis of UGTs from 65 plant genomes representing taxonomic diversity from green algae to angiosperms showed that *E. grandis* contains the highest number of UGT genes (379) among the species that were investigated. For comparison, 123 UGT genes were identified in Arabidopsis (*A. thaliana*) and 136 in pomegranate (*Punica granatum*), the latter species belonging to the same order (Myrtales) as *Eucalyptus*. Notably, the genomes of oak (*Quercus suber*) and loblolly pine (*Pinus taeda*), two long-lived trees, also contain a high number of UGT genes (312 and 243, respectively). Long-lived trees experience diverse environmental conditions over their lifespan, and it can be speculated that a rich UGT diversity is important for tree longevity.

Little is known about UGT functionalities in *Eucalyptus*, thereby leaving a large catalog of uncharted UGTs to be investigated. In addition to the UGTs reported in this and our previous study (Hansen *et al.*, 2018), two UGT84As from *E. camaldulensis* were shown to glucosylate gallic acid in the biosynthesis of hydrolysable tannins (Tahara *et al.*, 2018). The genome of *E. grandis* contains numerous CYP and TPS genes (Myburg *et al.*, 2014; Kulheim *et al.*, 2015; Hansen *et al.*, 2021), which are involved in the production of various specialized metabolites in addition to some general metabolites. Therefore, the rich UGT resource of 379 genes may complement the large collection of TPSs and CYPs by serving to stabilize, detoxify, and facilitate the transport and storage of the metabolites produced by these and other biosynthetic enzymes.

Many of the *UGT* genes in the genome of *E. grandis* occur in clusters, consistent with a high proportion of tandemly duplicated genes in this species (Myburg *et al.*, 2014). The multiplicity of duplicated genes provides a repertoire of genes ready to undergo neofunctionalization by positive selection of advantageous mutations. Sometimes very few mutations result

in gain of a new function. This is exemplified by the site-directed mutagenesis study of the UGT87Y1 ortholog, *EgUGT87*, where a single-point mutation in the active site resulted in partial gain of the ability to glucosylate mandelonitrile compared with UGT87Y1 (Fig. 6). Interestingly, increased prunasin accumulation was observed for the *EgUGT87Y* mutants, while no-to-partial recovery was observed for the *m/z* 340 : 329 ratio. This could be explained by variation in the delivery of different T-DNA constructs to the same cell upon co-expression of multiple genes (Montague *et al.*, 2011) and/or by stimulation of CYP activity (Laursen *et al.*, 2016). For example, there may be an increased flux of phenylalanine to mandelonitrile by the presence of certain UGT proteins as observed for the dhurrin metabolon.

Studies into the functionality of tandemly duplicated UGT genes are relatively rare. A key study in the grass *Brachypodium distachyon* showed that two UGTs in a cluster comprising six similar UGT genes conferred resistance against trichothecene mycotoxins from the fungal pathogen *Fusarium graminearum* (Schweiger *et al.*, 2013). The authors found no evidence of diversifying selection and therefore speculated that such a cluster constituted a 'trench in the warfare' between grasses and fungal pathogens. This strategy equips the plant with a collection of genes readily available to be activated for defense upon fungal infection or herbivore attack. Other studies demonstrate similar differential functionality of tandemly duplicated UGT genes in terms of sugar donor and acceptor specificity (Chen *et al.*, 2021; Sayama *et al.*, 2012), as well as regio-selectivity for glycosidic linkage (Ono *et al.*, 2019). Beyond UGTs, other defense gene families including CYPs and TPSs are heavily expanded in *Eucalyptus* and are positioned in tandem repeats (Myburg *et al.*, 2014). Further studies into the regulation and evolution of these tandemly duplicated genes in *Eucalyptus* would shed light on genetic strategies to secure the metabolic and physiological plasticity as required by these long-lived trees.

Evolution of chemical defense in *Eucalyptus*

The biosynthesis of cyanogenic glucosides is known from several distantly related plant species. Nonetheless, little is known about the evolution of cyanogenic glucoside biosynthesis in closely related plant species (Lai *et al.*, 2020). Here, we show on the molecular level how the distribution and biosynthesis of cyanogenic glucosides occurs within a subgenus. The prunasin biosynthetic gene patterns in the six species analyzed here suggest that the capacity to produce prunasin, and the three-CYP system, arose early in the diversification of the subgenus *Symphyomyrtus*, whereby the trait was retained by some species (i.e. within the *Sejunctae* and *Adnataria* series), and lost by others. Gene duplication of an ancestral *CYP79A125* likely gave rise to the *CYP79A34* and *CYP79A35* where prunasin functionality was then reobtained in *E. camphora* and *E. yarraensis* (section *Maidenaria*) but then lost in *E. grandis* (section *Latoangulatae*) via gene pseudogenization. Pseudogenization is an apparent driver for biosynthetic capacity loss in *E. camaldulensis*, but here it is due to mutations

in the CYP71 and CYP706 pathway genes (Fig. 5). Cyanogenic glucoside loss in these *Eucalyptus* species differs from the well-characterized system in sweet almond (*P. dulcis*), where the lack of prunasin biosynthesis is due to a mutation in the BHLH transcription factor activating CYP79 and CYP71 transcriptions, resulting in a lack of CYP expression (Sánchez-Pérez *et al.*, 2019). These CYP evolution patterns, coupled to the recruitment of different UGT families in distinct phylogenetic clades, demonstrate the dynamic evolution of cyanogenic glucoside biosynthesis in *Eucalyptus*. To understand the full extent of pathway gain and loss, access to much greater genomic resources of high quality across this important genus is required.

Based on the high diversity of CYPs and UGTs in the sequenced *Eucalyptus* species to date, it is reasonable to suggest that there may be many members of the prunasin gene families that possess some capacity to catalyze a reaction relevant for cyanogenic glucoside formation, especially as many other characterized CYPs and UGTs show substrate promiscuity (Hansen *et al.*, 2003, 2021; Zhang *et al.*, 2022). The *Eucalyptus* genus is large and genetically complex, and it is possible that some species previously classified as acyanogenic may possess a limited degree of capacity to produce cyanogenic glucosides. It is interesting to speculate that the remarkable adaptivity of *Eucalyptus* species may be an evolutionary outcome where certain defense pathways are more readily activated or somatically mutated to provide a fitness advantage under particular environmental conditions or pressures, as shown in the so-called ‘mosaic trees’ (Padovan *et al.*, 2015; Orr *et al.*, 2020).

Eucalyptus are predominantly native to the Australian continent, where many soils are depleted in nitrogen and phosphorous, and the apparent ‘cost’ related to the formation and storage of these nitrogen-containing specialized metabolites might be an important factor driving selection for or against the presence of cyanogenic glucosides (Neilson *et al.*, 2013). Studies into the geochemical distribution of cyanogenic glucosides in *Eucalyptus* species are required to link soil conditions to the accumulation of cyanogenic glucosides. Notably, trees belonging to the section *Adnataria* generally grow on younger and fertile soil (Pryor, 1976). Coincidentally, this section and the sections *Bisectae*, *Exsertaria*, *Glandulosae*, and *Maidenaria* were shown to have accelerated net species diversification (Thornhill *et al.*, 2019), and the authors speculated that a driver of diversification could be the ability to quickly adapt to newer geochemical environments.

The evolution of specific chemotypes within the genus may also be driven by different abiotic and biotic factors. For example, an analysis of *E. camaldulensis* seedlings sourced from a wide geographical range linked the variation in mono- and sesquiterpene abundance to temperature and humidity (Bustos-Segura *et al.*, 2017), while formylated phloroglucinol compound concentration was implicated in the preferred tree selection of koalas (*Phascolarctos cinereus*) (Moore & Foley, 2005). The genetic complexity of *Eucalyptus* species, characterized by numerous tandem duplications and great diversity of specialized metabolite gene families (Myburg *et al.*, 2014), provides a rich genetic space from which to garner information on metabolic and physiological plasticity within plants.

Acknowledgements

We thank David R. Nelson for naming the CYPome from *Eucalyptus grandis* and David Vernon for his generous support for wildlife and early career researchers. This work was supported by the VILLUM Center for Plant Plasticity (VKR023054) (BLM); the European Research Council Advanced Grant (ERC-2012-ADG_20120314) (BLM); VILLUM Young Investigator Grant (VKR013167) (EHJN), a Novo Nordisk Emerging Investigator Grant (0054890) (EHJN) and funding from the Danish Independent Research Council (6111-00379B, 1051-00083B, and 1131-00002B) (EHJN).

Author contributions

CCH, MS, JQDG, IEW, BLM and EHJN conceptualized the study. CCH, MS, MB, CEO and EHJN designed the study methodology. CCH, MS, MB, WB and EHJN involved in investigation, and curated and analyzed the data. CCH and EHJN involved in writing – original draft. CCH, MS, MB, WB, CEO, JQDG, IEW, BLM and EHJN involved in writing – review and editing. BLM, IEW and EHJN acquired fund for the study.

ORCID

Matteo Bellucci  <https://orcid.org/0000-0002-1392-5091>

Wolfgang Brandt  <https://orcid.org/0000-0002-0825-1491>

Jason Q. D. Goodger  <https://orcid.org/0000-0001-7392-3891>

Cecilie Cetti Hansen  <https://orcid.org/0000-0002-0973-6466>

Birger Lindberg Møller  <https://orcid.org/0000-0002-3252-3119>

Elizabeth H. J. Neilson  <https://orcid.org/0000-0002-8279-9906>

Carl Erik Olsen  <https://orcid.org/0000-0002-2275-0596>

Mette Sørensen  <https://orcid.org/0000-0001-9787-8386>

Ian E. Woodrow  <https://orcid.org/0000-0002-6397-8378>

Data availability

The transcriptomic data that support the findings of this study are openly available in BioProject at www.ncbi.nlm.nih.gov/bioproject, ID: PRJNA765518.

References

- Bak S, Kahn RA, Nielsen HL, Møller BL, Halkier BA. 1998. Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. *Plant Molecular Biology* 36: 393–405.
- Bak S, Olsen CE, Halkier BA, Møller BL. 2000. Transgenic tobacco and Arabidopsis plants expressing the two multifunctional sorghum cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. *Plant Physiology* 123: 1437–1448.

- Bashandy H, Jalkanen S, Teeri TH. 2015. Within leaf variation is the largest source of variation in agroinfiltration of *Nicotiana benthamiana*. *Plant Methods* 11: 47.
- Bennett AF. 2016. Eucalypts, wildlife and nature conservation: from individual trees to landscape patterns. *Proceedings of the Royal Society of Victoria* 128: 71–86.
- Bjarnholt N, Neilson EHJ, Crocoll C, Jørgensen K, Motawia MS, Olsen CE, Dixon DP, Edwards R, Møller BL. 2018. Glutathione transferases catalyze recycling of auto-toxic cyanogenic glucosides in sorghum. *The Plant Journal* 94: 1109–1125.
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34: 525–527.
- Brazier-Hicks M, Offen WA, Gershter MC, Revett TJ, Lim E-K, Bowles DJ, Davies GJ, Edwards R. 2007. Characterization and engineering of the bifunctional *N*- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants. *Proceedings of the National Academy of Sciences, USA* 104: 20238–20243.
- Burchell B, Nebert DW, Nelson DR, Bock KW, Iyanagi T, Jansen PL, Lancet D, Mulder GJ, Chowdhury JR, Siest G *et al.* 1991. The UDP glucuronosyltransferase gene superfamily: suggested nomenclature based on evolutionary divergence. *DNA and Cell Biology* 10: 487–494.
- Busk PK, Møller BL. 2002. Dhurrin synthesis in sorghum is regulated at the transcriptional level and induced by nitrogen fertilization in older plants. *Plant Physiology* 129: 1222–1231.
- Bustos-Segura C, Dillon S, Keszei A, Foley WJ, Külheim C. 2017. Intraspecific diversity of terpenes of *Eucalyptus camaldulensis* (Myrtaceae) at a continental scale. *Australian Journal of Botany* 65: 257–269.
- Butler JB, Vaillancourt RE, Potts BM, Lee DJ, King GJ, Baten A, Shepherd M, Freeman JS. 2017. Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. *BMC Genomics* 18: 397.
- Carpenter EJ, Matasci N, Ayyampalayam S, Wu S, Sun J, Yu J, Jimenez Vieira FR, Bowler C, Dorrell RG, Gitzendanner MA *et al.* 2019. Access to RNA-sequencing data from 1,173 plant species: the 1000 Plant transcriptomes initiative (1KP). *Gigascience* 8: gzl126.
- Cerveau N, Jackson DJ. 2016. Combining independent *de novo* assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms. *BMC Bioinformatics* 17: 525.
- Chae L, Kim T, Nilo-Poyanco R, Rhee SY. 2014. Genomic signatures of specialized metabolism in plants. *Science* 344: 510–513.
- Chen X, Chen J, Feng J, Wang Y, Li S, Xiao Y, Diao Y, Zhang L, Chen W. 2021. Tandem UGT71B5s catalyze lignan glycosylation in *Isatis indigotica* with substrates promiscuity. *Frontiers in Plant Science* 12. doi: 10.3389/fpls.2021.637695
- Dick R, Rattei T, Haslbeck M, Schwab W, Gierl A, Frey M. 2012. Comparative analysis of benzoxazinoid biosynthesis in monocots and dicots: independent recruitment of stabilization and activation functions. *Plant Cell* 24: 915–928.
- Franks TK, Yadollahi A, Wirthensohn MG, Guerin JR, Kaiser BN, Sedgley M, Ford CM. 2008. A seed coat cyanohydrin glucosyltransferase is associated with bitterness in almond (*Prunus dulcis*) kernels. *Functional Plant Biology* 35: 236–246.
- George Thompson AM, Iancu CV, Neet KE, Dean JV, Choe J-Y. 2017. Differences in salicylic acid glucose conjugations by UGT74F1 and UGT74F2 from *Arabidopsis thaliana*. *Scientific Reports* 7: 46629.
- Gleadow RM, Haburjak J, Dunn JE, Conn ME, Conn EE. 2008. Frequency and distribution of cyanogenic glucosides in *Eucalyptus* L'Herit. *Phytochemistry* 69: 1870–1874.
- Gleadow RM, Møller BL. 2014. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annual Review of Plant Biology* 65: 155–185.
- Gleadow RM, Woodrow IE. 2000. Temporal and spatial variation in cyanogenic glycosides in *Eucalyptus cladocalyx*. *Tree Physiology* 20: 591–598.
- Gleadow RM, Woodrow IE. 2002. Constraints on effectiveness of cyanogenic glycosides in herbivore defense. *Journal of Chemical Ecology* 28: 1301–1313.
- Goodger JQD, Capon RJ, Woodrow IE. 2002. Cyanogenic polymorphism in *Eucalyptus polyanthemos* Schauer subsp. *vestita* L. Johnson and K. Hill (Myrtaceae). *Biochemical Systematics and Ecology* 30: 617–630.
- Goodger JQD, Choo TY, Woodrow IE. 2007. Ontogenetic and temporal trajectories of chemical defence in a cyanogenic eucalypt. *Oecologia* 153: 799–808.
- Goodger JQD, Woodrow IE. 2011. α,β -Unsaturated monoterpene acid glucose esters: structural diversity, bioactivities and functional roles. *Phytochemistry* 72: 2259–2266.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al.* 2012. PHYTOZOME: a comparative platform for green plant genomics. *Nucleic Acids Research* 40: D1178–D1186.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M *et al.* 2013. *De novo* transcript sequence reconstruction from RNA-seq using the TRINITY platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.
- Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. 2021. Plant cytochrome P450 plasticity and evolution. *Molecular Plant* 14: 1244–1265.
- Hansen CC, Sørensen M, Veiga TAM, Zibrandtsen JFS, Heskes AM, Olsen CE, Boughton BA, Møller BL, Neilson EHJ. 2018. Reconfigured cyanogenic glucoside biosynthesis in *Eucalyptus cladocalyx* involves a cytochrome P450 CYP706C55. *Plant Physiology* 178: 1081–1095.
- Hansen KS, Kristensen C, Tattersall DB, Jones PR, Olsen CE, Bak S, Møller BL. 2003. The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. *Phytochemistry* 64: 143–151.
- Hearne PQ, Knorr DA, Hillman BI, Morris TJ. 1990. The complete genome structure and synthesis of infectious RNA from clones of tomato bushy stunt virus. *Virology* 177: 141–151.
- Hirakawa H, Nakamura Y, Kaneko T, Isohe S, Sakai H, Kato T, Hibino T, Sasamoto S, Watanabe A, Yamada M *et al.* 2011. Survey of the genetic information carried in the genome of *Eucalyptus camaldulensis*. *Plant Biotechnology* 28: 471–480.
- Irmisch S, McCormick AC, Boeckler GA, Schmidt A, Reichelt M, Schneider B, Block K, Schnitzler JP, Gershenzon J, Unsicker SB *et al.* 2013. Two herbivore-induced cytochrome P450 enzymes CYP79D6 and CYP79D7 catalyze the formation of volatile aldoximes involved in poplar defense. *Plant Cell* 25: 4737–4754.
- Jenrich R, Trompetter I, Bak S, Olsen CE, Møller BL, Piotrowski M. 2007. Evolution of heteromeric nitrilase complexes in Poaceae with new functions in nitrile metabolism. *Proceedings of the National Academy of Sciences, USA* 104: 18848–18853.
- Jørgensen K, Morant AV, Morant M, Jensen NB, Olsen CE, Kannangara R, Motawia MS, Møller BL, Bak S. 2011. Biosynthesis of the cyanogenic glucosides linamarin and lotaustralin in cassava: isolation, biochemical characterization, and expression pattern of CYP71E7, the oxime-metabolizing cytochrome P450 enzyme. *Plant Physiology* 155: 282–292.
- Kannangara R, Motawia MS, Hansen NK, Paquette SM, Olsen CE, Møller BL, Jørgensen K. 2011. Characterization and expression profile of two UDP-glucosyltransferases, UGT85K4 and UGT85K5, catalyzing the last step in cyanogenic glucoside biosynthesis in cassava. *The Plant Journal* 68: 287–301.
- Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* 77(Suppl 9): 114–122.
- Krieger E, Vriend G. 2014. YASARA view – molecular graphics for all devices – from smartphones to workstations. *Bioinformatics* 30: 2981–2982.
- Krieger E, Vriend G. 2015. New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry* 36: 996–1007.
- Kristensen C, Morant M, Olsen CE, Ekstrøm CT, Galbraith DW, Møller BL, Bak S. 2005. Metabolic engineering of dhurrin in transgenic *Arabidopsis* plants with marginal inadvertent effects on the metabolome and transcriptome. *Proceedings of the National Academy of Sciences, USA* 102: 1779–1784.
- Kulheim C, Padovan A, Hefer C, Krause ST, Kollner TG, Myburg AA, Degenhardt J, Foley WJ. 2015. The *Eucalyptus* terpene synthase gene family. *BMC Genomics* 16: 450.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis v.7.0 for bigger datasets. *Molecular Biology and Evolution* 33: 1870–1874.
- Lai D, Maimann AB, Macea E, Ocampo CH, Cardona G, Pičmanová M, Darbani B, Olsen CE, Debouck D, Raatz B *et al.* 2020. Biosynthesis of

- cyanogenic glucosides in *Phaseolus lunatus* and the evolution of oxime-based defenses. *Plant Direct* 4: e00244.
- Lakatos L, Szittyá G, Silhavy D, Burgyán J. 2004. Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *EMBO Journal* 23: 876–884.
- Laursen T, Borch J, Knudsen C, Bavishi K, Torta F, Martens HJ, Silvestro D, Hatzakis NS, Wenk MR, Dafforn TR *et al.* 2016. Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science* 354: 890–893.
- Li L, Modolo LV, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X. 2007. Crystal structure of *Medicago truncatula* UGT85H2 – insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *Journal of Molecular Biology* 370: 951–963.
- Li P, Li YJ, Wang B, Yu HM, Li Q, Hou BK. 2017. The Arabidopsis UGT87A2, a stress-inducible family 1 glycosyltransferase, is involved in the plant adaptation to abiotic stresses. *Physiologia Plantarum* 159: 416–432.
- Lim E-K, Doucet CJ, Li Y, Elias L, Worrall D, Spencer SP, Ross J, Bowles DJ. 2002. The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *The Journal of Biological Chemistry* 277: 586–592.
- Lindgreen S. 2012. ADAPTERREMOVAL: easy cleaning of next-generation sequencing reads. *BMC Research Notes* 5: 337.
- Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Bélanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T *et al.* 1997. The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* 7: 255–269.
- Mackenzie PI, Walter Bock K, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW. 2005. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenetics and Genomics* 15: 677–685.
- Marsh KJ, Külheim C, Blomberg SP, Thornhill AH, Miller JT, Wallis IR, Nicolle D, Salminen JP, Foley WJ. 2017. Genus-wide variation in foliar polyphenolics in eucalypts. *Phytochemistry* 144: 197–207.
- Møller BL, Olsen CE, Motawia MS. 2016. General and stereocontrolled approach to the chemical synthesis of naturally occurring cyanogenic glucosides. *Journal of Natural Products* 79: 1198–1202.
- Montague NP, Thuenemann EC, Saxena P, Saunders K, Lenzi P, Lomonosoff GP. 2011. Recent advances of Cowpea mosaic virus-based particle technology. *Human Vaccines* 7: 383–390.
- Moore BD, Foley WJ. 2005. Tree use by koalas in a chemically complex landscape. *Nature* 435: 488–490.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* 510: 356–362.
- Neilson EH. 2012. *Characterisation of cyanogenic glucoside synthesis in Eucalyptus*. PhD thesis, University of Melbourne, Melbourne, Vic., Australia.
- Neilson EH, Goodger JQD, Motawia MS, Bjarnholt N, Frisch T, Olsen CE, Møller BL, Woodrow IE. 2011. Phenylalanine derived cyanogenic diglucosides from *Eucalyptus camphora* and their abundances in relation to ontogeny and tissue type. *Phytochemistry* 72: 2325–2334.
- Neilson EH, Goodger JQD, Woodrow IE. 2006. Novel aspects of cyanogenesis in *Eucalyptus camphora* subsp. *humeana*. *Functional Plant Biology* 33: 487–496.
- Neilson EH, Goodger JQD, Woodrow IE, Møller BL. 2013. Plant chemical defense: at what cost? *Trends in Plant Science* 18: 250–258.
- Nelson DR. 2009. The cytochrome P450 homepage. *Human Genomics* 4: 59.
- Nicolle D. 2019. *Classification of the eucalypts* (Angophora, Corymbia and Eucalyptus) v.4. [WWW document] URL <http://www.dn.com.au/Classification-Of-The-Eucalypts.pdf>
- Ono E, Waki T, Oikawa D, Murata J, Shiraiishi A, Toyonaga H, Kato M, Ogata N, Takahashi S, Yamaguchi M *et al.* 2019. Glycoside-specific glycosyltransferases catalyze regio-selective sequential glucosylations for a sesame lignan, sesaminol triglucoside. *The Plant Journal* 101: 1221–1233.
- Orr AJ, Padovan A, Kainer D, Külheim C, Lindell Bromham L, Bustos-Segura C, Foley W, Haff T, Hsieh J-F, Morales-Suarez A *et al.* 2020. A phylogenomic approach reveals a low somatic mutation rate in a long-lived plant. *Proceedings of the Royal Society B: Biological Sciences* 287: 20192364.
- Padovan A, Patel HR, Chuah A, Huttley GA, Krause ST, Degenhardt J, Foley WJ, Külheim C. 2015. Transcriptome sequencing of two phenotypic mosaic *Eucalyptus* trees reveals large scale transcriptome re-modelling. *PLoS ONE* 10: e0123226.
- Picmanova M, Neilson EH, Motawia MS, Olsen CE, Agerbirk N, Gray CJ, Flitsch S, Meier S, Silvestro D, Jørgensen K *et al.* 2015. A recycling pathway for cyanogenic glucosides evidenced by the comparative metabolic profiling in three cyanogenic plant species. *The Biochemical Journal* 469: 375–389.
- Pryor LD. 1976. *The biology of eucalypts*. London, UK: Edward Arnold.
- von Rad U, Hüttel R, Lottspeich F, Gierl A, Frey M. 2001. Two glycosyltransferases are involved in detoxification of benzoxazinoids in maize. *The Plant Journal* 28: 633–642.
- Sánchez-Pérez R, Pavan S, Mazzeo R, Moldovan C, Aiese Cigliano R, Del Cueto J, Ricciardi F, Lotti C, Ricciardi L, Dicenta F *et al.* 2019. Mutation of a bHLH transcription factor allowed almond domestication. *Science* 364: 1095–1098.
- dos Santos BM, Zibrandtsen JFS, Gunbilig D, Sørensen M, Cozzi F, Boughton BA, Heskes AM, Neilson EHJ. 2019. Quantification and localization of formylated phloroglucinol compounds (FPCs) in *Eucalyptus* species. *Frontiers in Plant Science* 10: 186.
- Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H *et al.* 2012. The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *The Plant Cell* 24: 2123–2138.
- Schmidt FB, Cho SK, Olsen CE, Yang SW, Møller BL, Jørgensen K. 2018. Diurnal regulation of cyanogenic glucoside biosynthesis and endogenous turnover in cassava. *Plant Direct* 2: e00038.
- Schweiger W, Pasquet J-C, Nussbaumer T, Paris MPK, Wiesnerberger G, Macadré C, Ametz C, Berthiller F, Lemmens M, Saindrenan P *et al.* 2013. Functional characterization of two clusters of *Brachypodium distachyon* UDP-glycosyltransferases encoding putative deoxynivalenol detoxification genes. *Molecular Plant-Microbe Interactions* 26: 781–792.
- Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X. 2005. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* 17: 3141–3154.
- Sicker D, Frey M, Schulz M, Gierl A. 2000. Role of natural benzoxazinones in the survival strategy of plants. *International Review of Cytology* 198: 319–346.
- Sørensen M, Rinnan R, Woodrow I, Møller BL, Neilson EHJ. 2020. The entangled dynamics of eucalypt leaf and flower volatile emissions. *Environmental and Experimental Botany* 176: 104032.
- Steane DA, Nicolle D, Sansaloni CP, Petrolí CD, Carling J, Kilian A, Myburg AA, Grattapaglia D, Vaillancourt RE. 2011. Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping. *Molecular Phylogenetics and Evolution* 59: 206–224.
- Tahara K, Nishiguchi M, Frolov A, Mittasch J, Milkowski C. 2018. Identification of UDP glycosyltransferases from the aluminum-resistant tree *Eucalyptus camaldulensis* forming beta-glucogallin, the precursor of hydrolyzable tannins. *Phytochemistry* 152: 154–161.
- Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K *et al.* 2011. Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *The Plant Journal* 68: 273–286.
- Thodberg S, Del Cueto J, Mazzeo R, Pavan S, Lotti C, Dicenta F, Jakobsen Neilson EH, Møller BL, Sanchez-Perez R. 2018. Elucidation of the amygdalin pathway reveals the metabolic basis of bitter and sweet almonds (*Prunus dulcis*). *Plant Physiology* 178: 1096–1111.
- Thodberg S, Hansen CC, Takos AM, Pičmanová M, Møller BL, Nelson DR, Neilson EHJ. 2021. The fern CYPome: fern-specific cytochrome P450 family involved in convergent evolution of chemical defense. *bioRxiv*. doi: [10.1101/2021.03.23.436569](https://doi.org/10.1101/2021.03.23.436569).
- Thornhill AH, Crisp MD, Külheim C, Lam KE, Nelson LA, Yeates DK, Miller JT. 2019. A dated molecular perspective of eucalypt taxonomy, evolution and diversification. *Australian Systematic Botany* 32: 29–48.
- Vining KJ, Romanel E, Jones RC, Klocko A, Alves-Ferreira M, Hefer CA, Amarasinghe V, Dharmawardhana P, Naithani S, Ranik M *et al.* 2015.

- The floral transcriptome of *Eucalyptus grandis*. *New Phytologist* 206: 1406–1422.
- Voinnet O, Pinto YM, Baulcombe DC. 1999. Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences, USA* 96: 14147–14152.
- Wang B, Jin SH, Hu HQ, Sun YG, Wang YW, Han P, Hou BK. 2012. UGT87A2, an Arabidopsis glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. *New Phytologist* 194: 666–675.
- Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart DS. 2021. CFM-ID 4.0: more accurate ESI-MS/MS spectral prediction and compound identification. *Analytical Chemistry* 93: 11692–11700.
- Wilson AE, Tian L. 2019. Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. *The Plant Journal* 100: 1273–1288.
- Yamaguchi T, Yamamoto K, Asano Y. 2014. Identification and characterization of CYP79D16 and CYP71AN24 catalyzing the first and second steps in L-phenylalanine-derived cyanogenic glycoside biosynthesis in the Japanese apricot, *Prunus mume* Sieb. et Zucc. *Plant Molecular Biology* 86: 215–223.
- Zhang K, Sun Y, Li M, Long R. 2021. CrUGT87A1, a UDP-sugar glycosyltransferases (UGTs) gene from *Carex rigescens*, increases salt tolerance by accumulating flavonoids for antioxidation in *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 159: 28–36.
- Zhang L-J, Wang D-G, Zhang P, Wu C, Li Y-Z. 2022. Promiscuity characteristics of versatile plant glycosyltransferases for natural product glycodiversification. *ACS Synthetic Biology* 11: 812–819.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Metabolites produced upon transient expression of prunasin biosynthetic genes in *Nicotiana benthamiana*.

Fig. S2 Drop-out experiment of the prunasin biosynthetic genes from *Eucalyptus yarraensis* and *Eucalyptus camphora*.

Fig. S3 Observation of a diagnostic fragment m/z 329 putatively annotated as benzoic acid glucoside.

Fig. S4 Assessment of UGT activity using diagnostic prunasin pathway derivatives provides a consistent measure of prunasin accumulation above purported *Nicotiana benthamiana* endogenous UGT activity.

Fig. S5 Phylogenetic tree of UGT85 and UGT87 sequences identified from five different *Eucalyptus* species.

Fig. S6 *In vitro* assays of UGT87Y1 and UGT85A59 showing activity toward different metabolite acceptors.

Fig. S7 CYP79 phylogeny with sequences identified from five different *Eucalyptus* species.

Fig. S8 CYP706 phylogeny with sequences identified from five different *Eucalyptus* species.

Fig. S9 CYP71 phylogeny with sequences identified from five different *Eucalyptus* species.

Fig. S10 Transient expression of *EgCYP79A36* and *EgCYP79A37* from the acyanogenic *Eucalyptus grandis*.

Fig. S11 Comparison of product profiles upon transient co-expression of *EcUGT85A59* or *EgUGT85A* with the prunasin CYPs from *Eucalyptus cladocalyx*.

Table S1 Primers used to isolate target open reading frames from cDNA.

Table S2 UGT gene expression at 1, 4, and 7 months in *Eucalyptus camphora* and *Eucalyptus yarraensis*.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.