New Phytologist Supporting Information

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

Cecilie Cetti Hansen, Mette Sørensen, Matteo Bellucci, Wolfgang Brandt, Carl Erik Olsen, Jason Q. D. Goodger, Ian E. Woodrow, Birger Lindberg Møller, Elizabeth H. J. Neilson

Article acceptance date: 18th October 2022

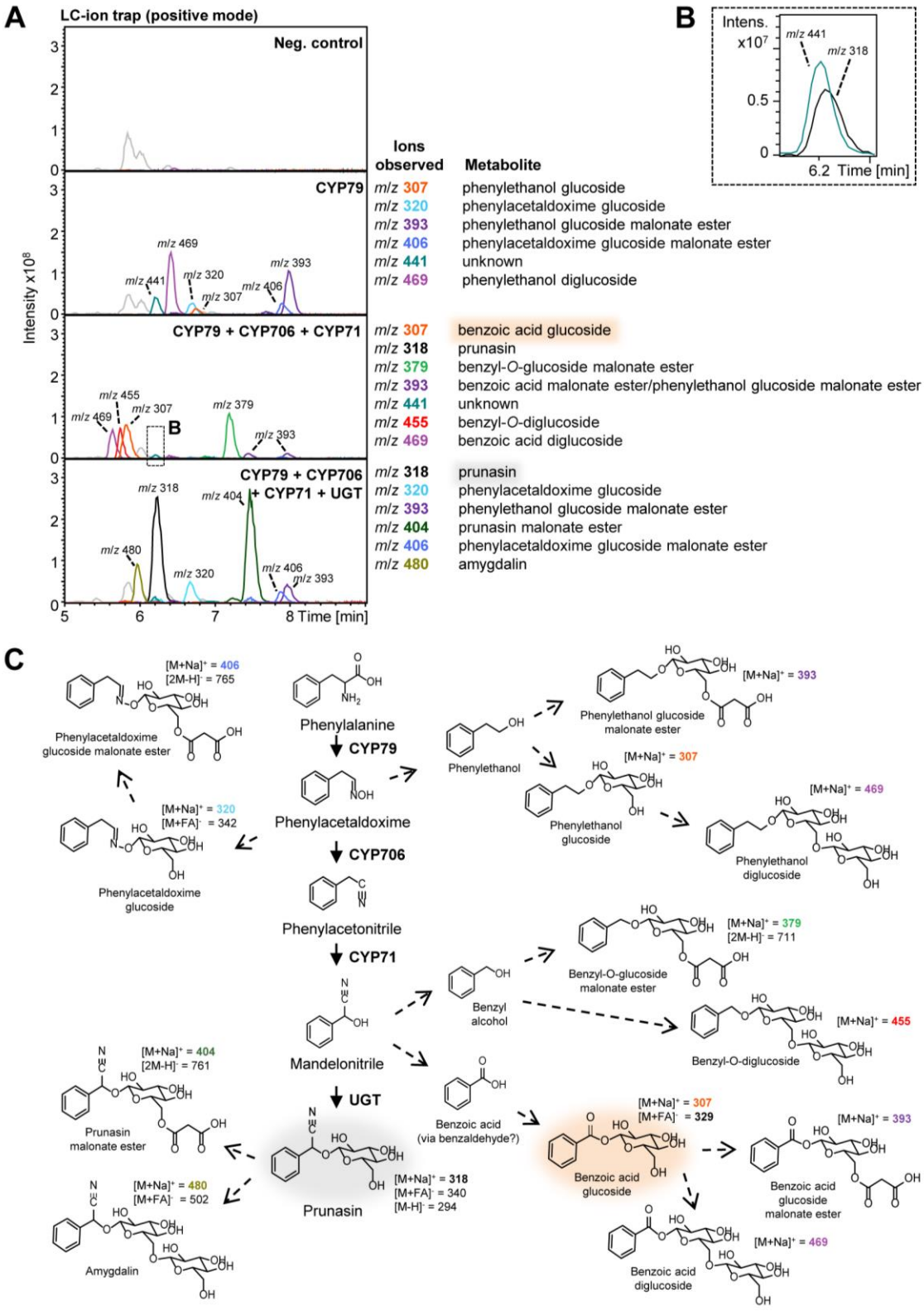

Fig. S1. Metabolites produced upon transient expression of prunasin biosynthetic genes in *N. benthamiana*. A. LC-ion trap chromatograms with the base peak chromatogram shown in light grey and extracted ion chromatograms (based on $[M+Na]^+$) shown in color. Each color corresponds

to distinct ions that appears upon transient expression of heterologous genes. **B.** Zoom on the two peaks corresponding to ions *m*/*z* 441 and 318 in CYP79, CYP706 and CYP71 combination, and marked by the dashed box. Small amounts of prunasin (*m/z* 318) accumulate when only the CYPs are co-expressed due to endogenous *N. benthamiana* UGT activity. **C.** Putatively annotated compounds related to transient expression of the prunasin pathway based on fragmentation pattern, their observed *m*/*z* values for the molecular ion related ions, biochemical knowledge and prior identification in literature (Bak *et al.*, 2000; Kristensen *et al.*, 2005; Hansen *et al.*, 2018; Thodberg *et al.*, 2018). Prunasin and putatively annotated benzoic acid glucoside are highlighted. An extended example with putative benzoic acid glucoside is shown in Figure S3. Prunasin and amygdalin were confirmed by authentic standards. For LC-ion trap spectra, only $[M+Na]^+$ was observed. For UHPLC-QqTOF spectra, one or more of the ions $[M+FA]^-$, $[2M-H]^-$ and $[M-H]^-$ were observed. Some derivatives (i.e. ions only displaying the $[M+Na]^+$ adduct) were only detected by the LC-ion trap method. The aglycons in this pathway grid (i.e. mandelonitrile, benzyl alcohol etc.) were not detected by either LC-MS method.

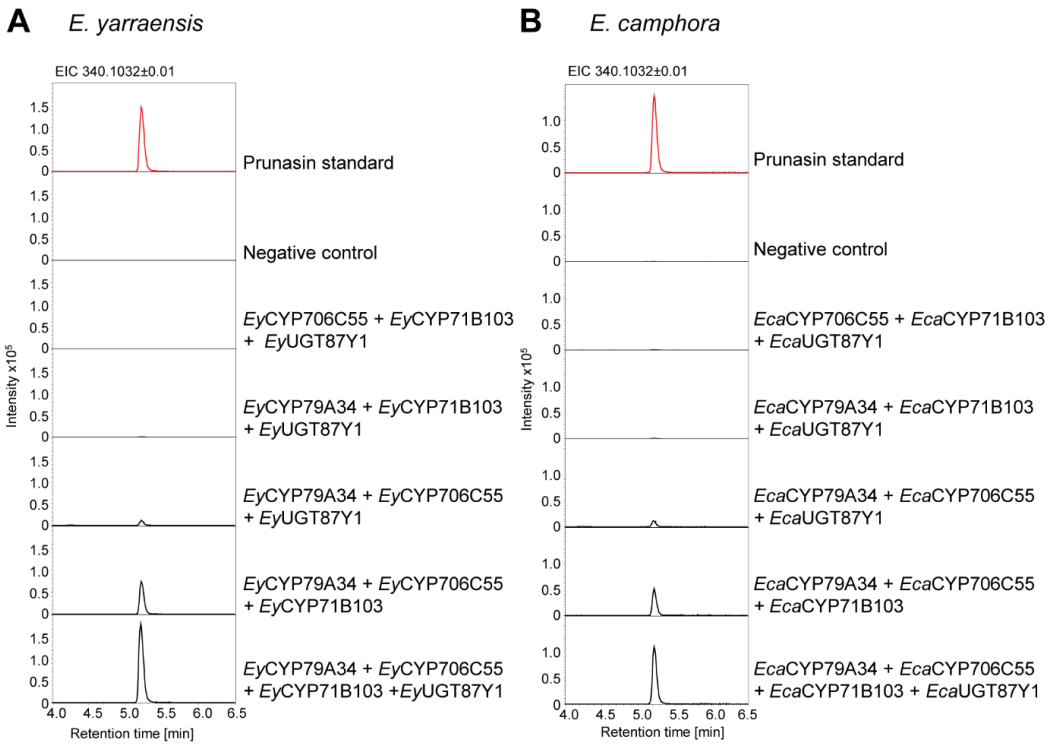

Fig. S2. Drop out experiment of the prunasin biosynthetic genes from *E. yarraensis* and *E. camphora*. Representative extracted ion chromatograms (EICs) showing accumulation of prunasin (*m/z* 340.1032; $[M+FA]^-$) in *N. benthamiana* plants transiently expressing some or all prunasin biosynthetic genes from **A.** *E. yarraensis* and **B.** *E. camphora*.

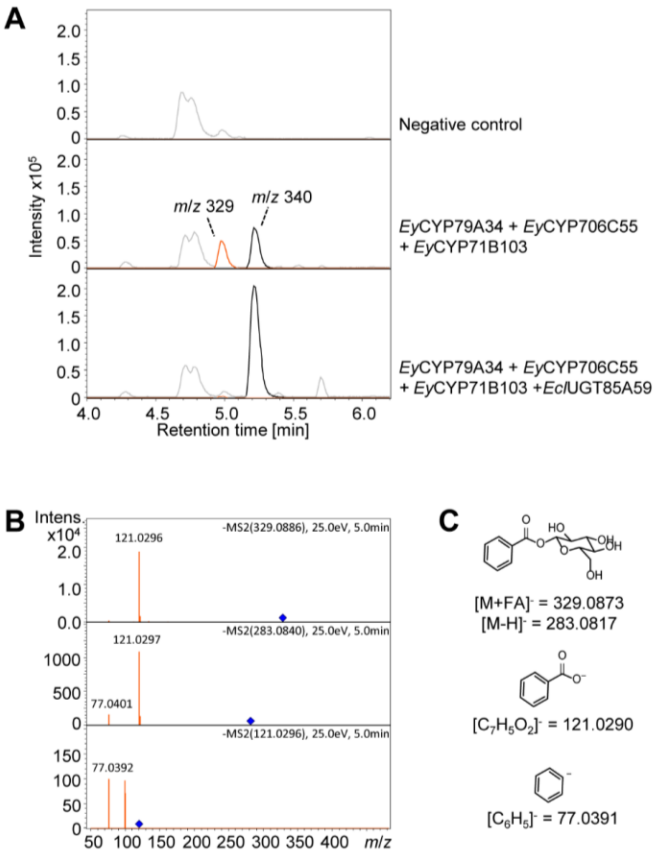

Fig. S3. Observation of a diagnostic fragment *m/z* 329 putatively annotated as benzoic acid glucoside. **A.** Extracted ion chromatograms (EICs) of *N. benthamiana* extracts shows accumulation of both prunasin (*m/z* 340.1032, $[M+FA]^-$; black) and *m/z* 329.0872 (orange) when only prunasin CYP genes are introduced indicating endogenous *N. benthamiana* UGT activity, but with metabolic leakage. Upon introduction of a UGT85A59, higher levels of prunasin are detected, while levels of *m/z* 329.0873 are negligible indicating metabolic channeling when the full pathway is expressed. The base peak chromatogram is shown in light grey. **B.** MS2 spectra of the diagnostic *m/z* 329 peak with rt = 5 min. **C.** Calculated *m/z* values and predicted fragments for benzoic acid glucoside. Compund identification was supported by CFM-ID 4.0 (Wang *et al.*, 2021).

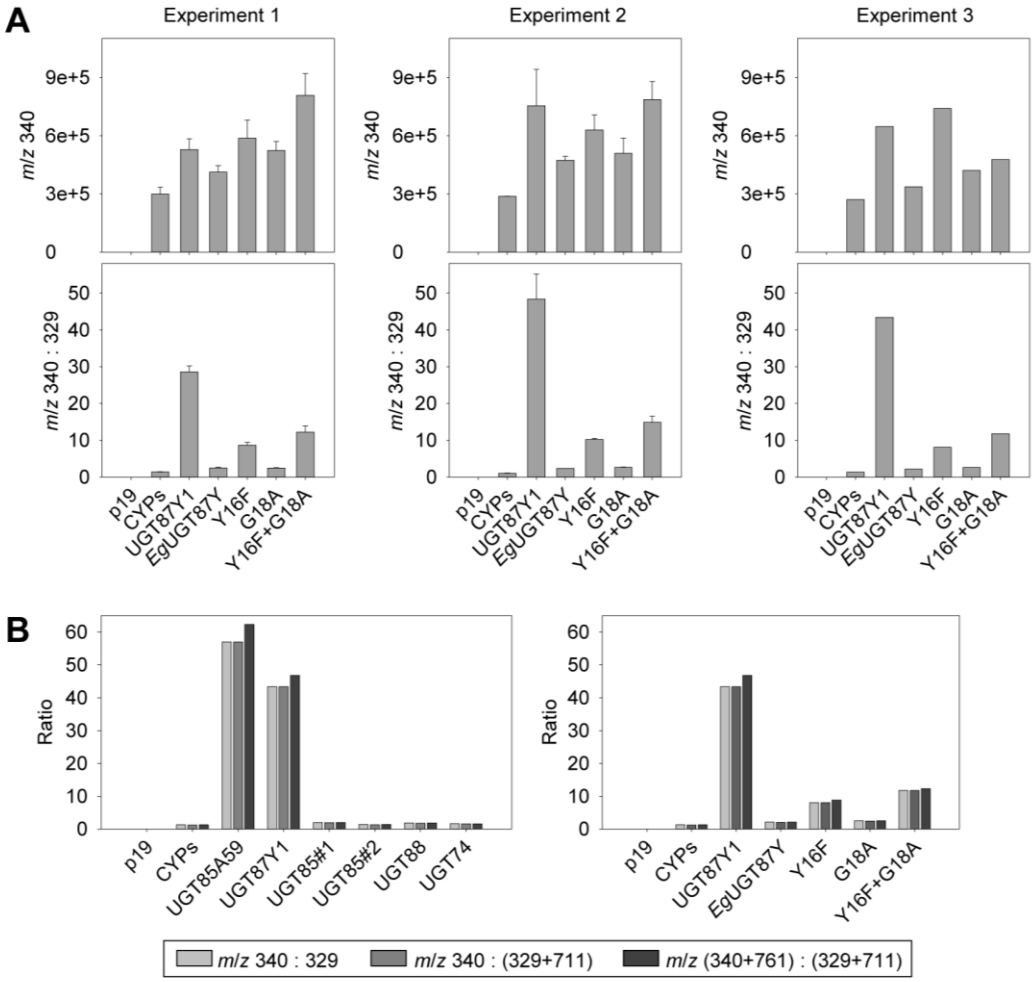

Fig. S4. Assessment of UGT activity using diagnostic prunasin pathway derivatives provides a consistent measure of prunasin accumulation above the background level, which is attributed to *N. benthamiana* endogenous UGT activity. **A.** Data from three independent transient expression experiments are compared, where in all experiments, UGT wild type or mutant variants were co-expressed the prunasin CYPs. Error bars show the standard error of technical replicates. Experiment three show the average of technical duplicates. Top panel: Prunasin levels (*m/z* 340.1032, $[M+FA]^-$) estimated as integrated area under the peak. Bottom panel: Prunasin levels normalized to the levels of putative benzoic acid glucoside (*m/z* 329.0873, $[M+FA]^-$) depict a high level of reproducibility between technical replicates (i.e. small error bars) and across three independent experiments. **B.** Calculation of UGT activity remains consistent when additional putative derivatives (*m/z* 711 and m/z 761) are included in the calculation. The left panel shows the test of candidate UGTs for the prunasin pathway and right panel shows the test of UGT87 mutants. The negative control (p19), the sample with the CYPs without a UGT co-expressed and the sample with UGT87Y1 co-expressed are the same in the two plots.

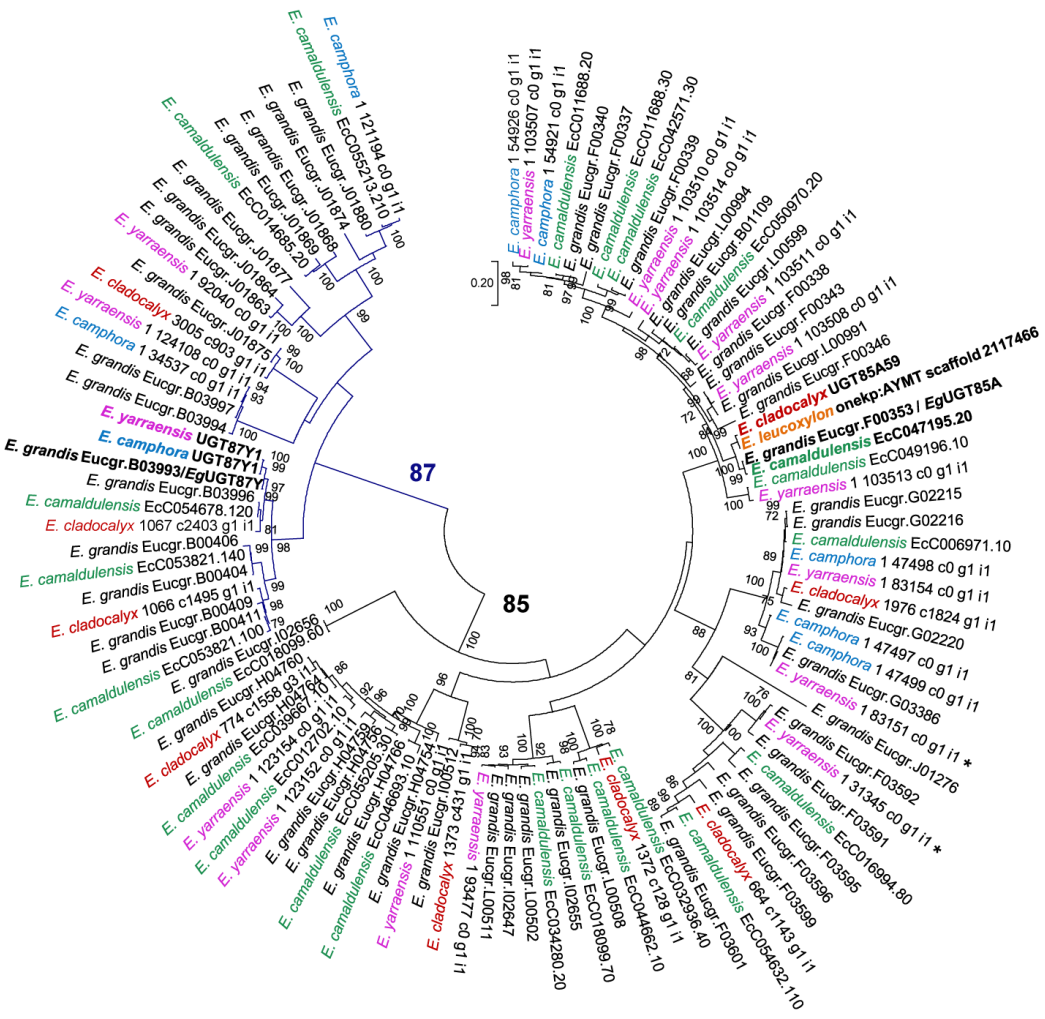

Fig. S5. Phylogenetic tree of UGT85 (black branches) and UGT87 (navy branches) sequences from five *Eucalyptus* species, with different colours used to differentiate species names. Amino acid sequences were aligned using MUSCLE and phylogeny was inferred using the Maximum Likelihood method with n=1000 bootstrap replicates. The branch lengths are measured in number of substitutions per site. Bootstrap values ≥70% are shown. UGT genes characterized in this study and their closest orthologs are marked in bold. The two *E. yarraensis* UGT85s that were tested as candidates for prunasin production are marked with an asterisk.

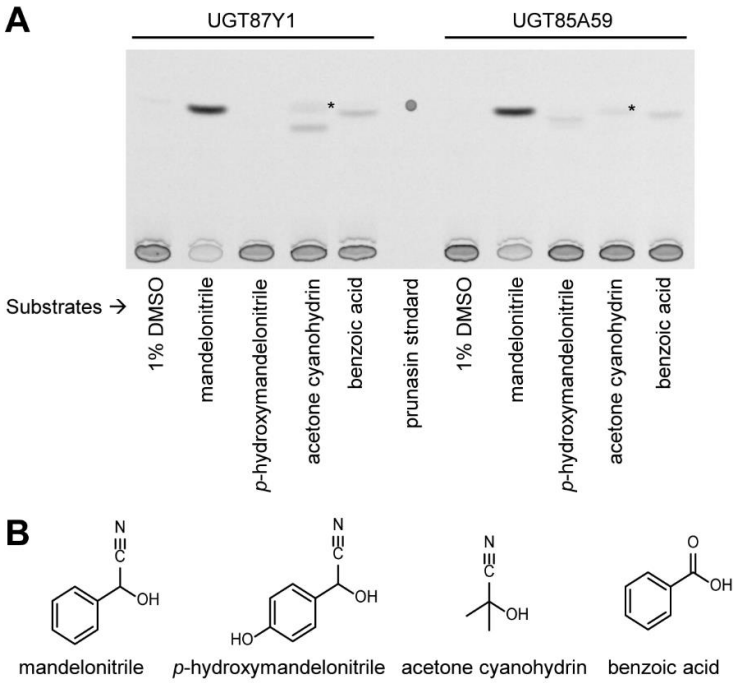

Fig. S6. *In vitro* assays with UGT87Y1 and UGT85A59. **A.** Thin layer chromatography plate of activity assays where UGT87Y1 and/or UGT85A59 showed activity towards the metabolite acceptor. The asterisk indicates a background signal in the assays with acetone cyanohydrin due to high substrate concentration. **B.** Structures of mandelonitrile, *p*-hydroxy mandelonitrile, acetone cyanohydrin and benzoic acid.

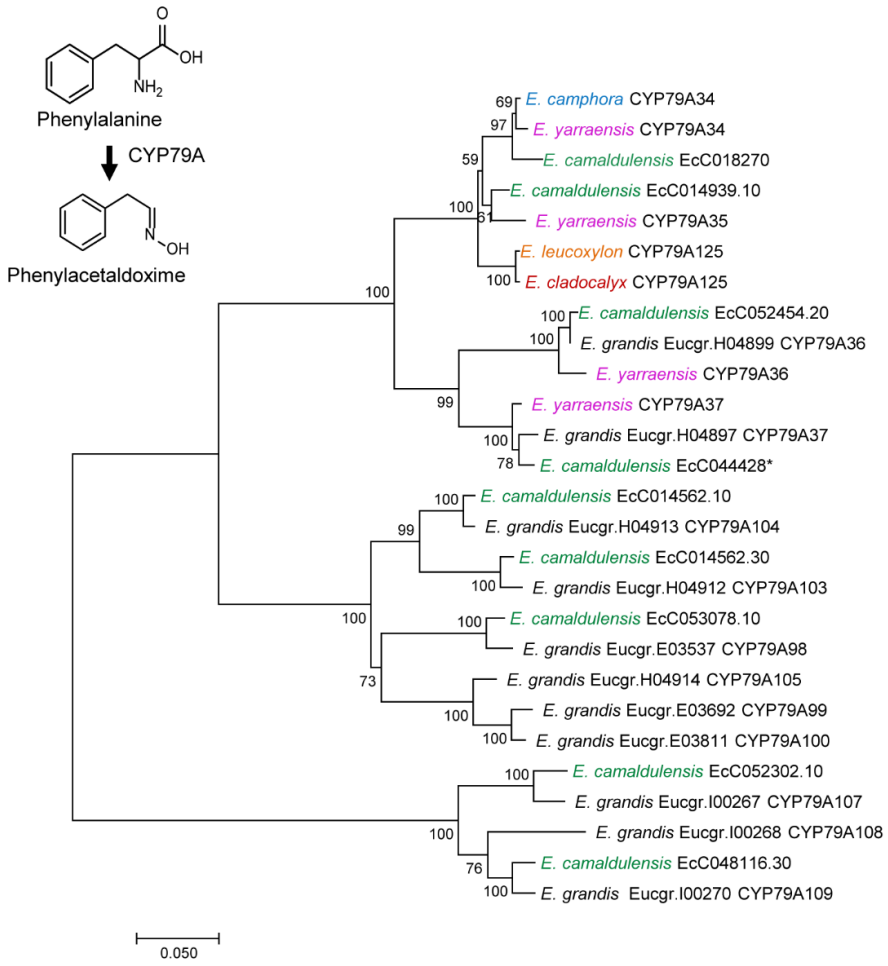

Fig. S7. *Eucalyptus* CYP79 phylogeny using sequences from five different species. All CYP79 members belong to the subfamily CYP79A. The first step in prunasin biosynthesis catalyzed by CYP79A enzymes is shown next to the phylogenetic tree. Full-length amino acid sequences were aligned using MUSCLE and phylogeny was inferred using the Maximum Likelihood method with n=1000 bootstrap replicates. The branch lengths are measured in number of substitutions per site. *Some reads in EcC044428 starting from amino acid #392 are missing due to low sequence quality.

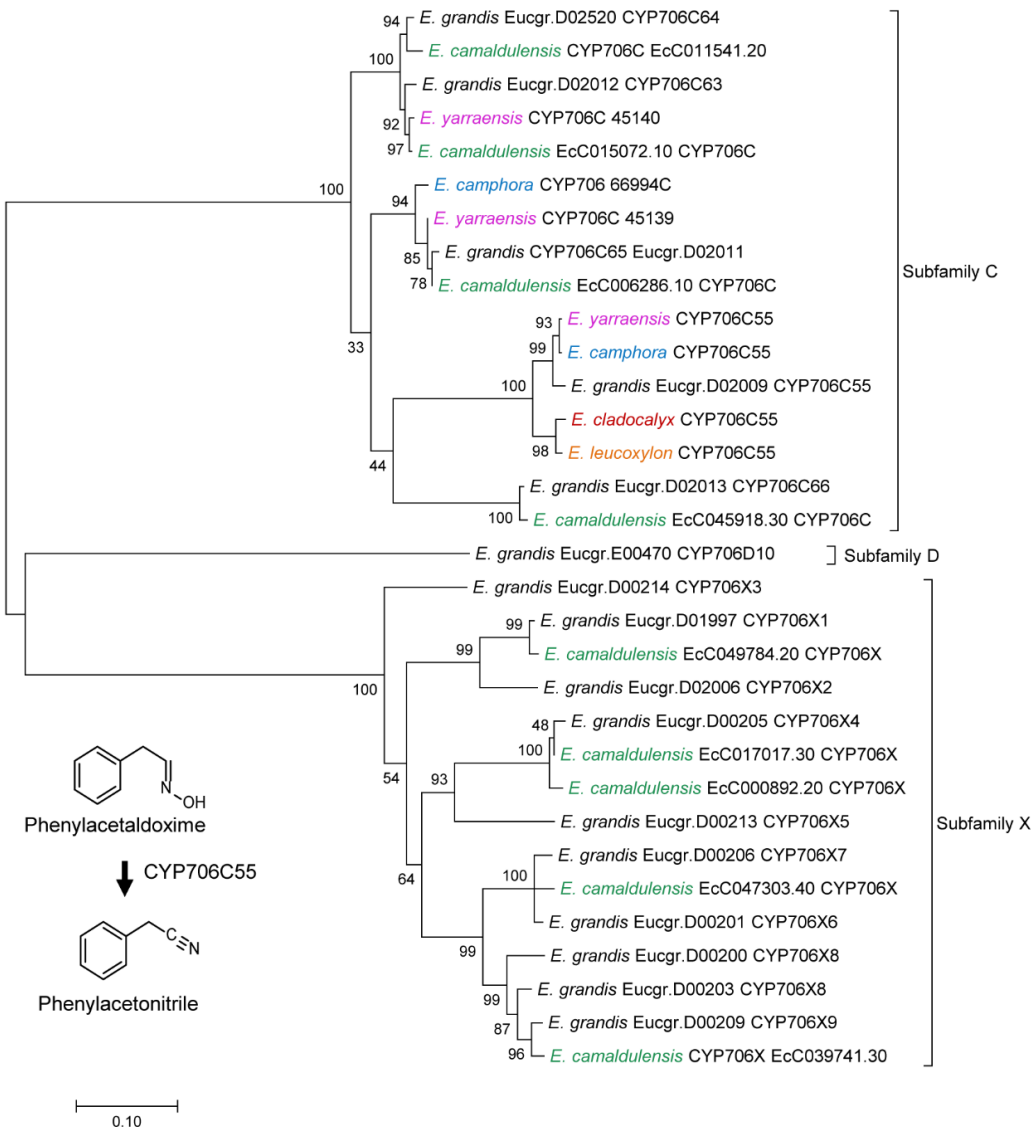

Fig. S8. *Eucalyptus* CYP706 phylogeny. The second step in prunasin biosynthesis catalyzed by CYP706C55 enzymes is shown next to the phylogenetic tree. Full-length amino acid sequences were aligned using MUSCLE and phylogeny was inferred using the Maximum Likelihood method with n=1000 bootstrap replicates. The branch lengths are measured in number of substitutions per site.

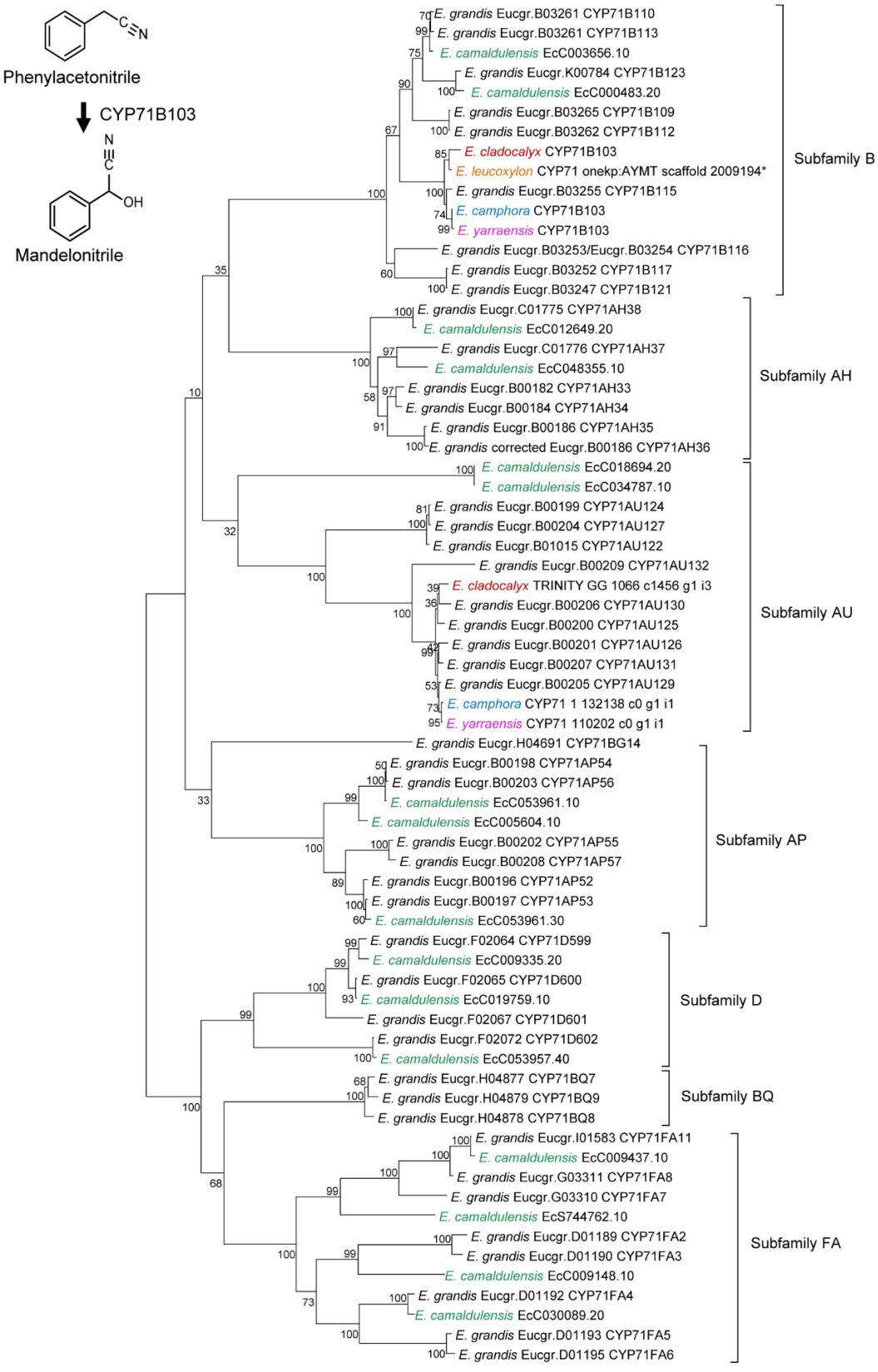
Phenylacetonitrile
CYP71B103
Mandelonitrile
E. grandis Eucgr.B03261 CYP71B110
E. grandis Eucgr.B03261 CYP71B113
E. camaldulensis EcC003656.10
E. grandis Eucgr.K00784 CYP71B123
E. camaldulensis EcC000483.20
E. grandis Eucgr.B03265 CYP71B109
E. grandis Eucgr.B03262 CYP71B112
E. cladocalyx CYP71B103
E. leucoxylon CYP71 onekp:AYMT scaffold 2009194*
E. grandis Eucgr.B03255 CYP71B115
E. camphora CYP71B103
E. yarraensis CYP71B103
E. grandis Eucgr.B03253/Eucgr.B03254 CYP71B116
E. grandis Eucgr.B03252 CYP71B117
E. grandis Eucgr.B03247 CYP71B121
Subfamily B
E. grandis Eucgr.C01775 CYP71AH38
E. camaldulensis EcC012649.20
E. grandis Eucgr.C01776 CYP71AH37
E. camaldulensis EcC048355.10
E. grandis Eucgr.B00182 CYP71AH33
E. grandis Eucgr.B00184 CYP71AH34
E. grandis Eucgr.B00186 CYP71AH35
E. grandis corrected Eucgr.B00186 CYP71AH36
Subfamily AH
E. camaldulensis EcC018694.20
E. camaldulensis EcC034787.10
E. grandis Eucgr.B00199 CYP71AU124
E. grandis Eucgr.B00204 CYP71AU127
E. grandis Eucgr.B01015 CYP71AU122
E. grandis Eucgr.B00209 CYP71AU132
E. cladocalyx TRINITY GG 1066 c1456 g1 i3
E. grandis Eucgr.B00206 CYP71AU130
E. grandis Eucgr.B00200 CYP71AU125
E. grandis Eucgr.B00201 CYP71AU126
E. grandis Eucgr.B00207 CYP71AU131
E. grandis Eucgr.B00205 CYP71AU129
E. camphora CYP71 1 132138 c0 g1 i1
E. yarraensis CYP71 110202 c0 g1 i1
Subfamily AU
E. grandis Eucgr.H04691 CYP71BG14
E. grandis Eucgr.B00198 CYP71AP54
E. grandis Eucgr.B00203 CYP71AP56
E. camaldulensis EcC053961.10
E. camaldulensis EcC005604.10
E. grandis Eucgr.B00202 CYP71AP55
E. grandis Eucgr.B00208 CYP71AP57
E. grandis Eucgr.B00196 CYP71AP52
E. grandis Eucgr.B00197 CYP71AP53
E. camaldulensis EcC053961.30
Subfamily AP
E. grandis Eucgr.F02064 CYP71D599
E. camaldulensis EcC009335.20
E. grandis Eucgr.F02065 CYP71D600
E. camaldulensis EcC019759.10
E. grandis Eucgr.F02067 CYP71D601
E. grandis Eucgr.F02072 CYP71D602
E. camaldulensis EcC053957.40
Subfamily D
E. grandis Eucgr.H04877 CYP71BQ7
E. grandis Eucgr.H04879 CYP71BQ9
E. grandis Eucgr.H04878 CYP71BQ8
Subfamily BQ
E. grandis Eucgr.I01583 CYP71FA11
E. camaldulensis EcC009437.10
E. grandis Eucgr.G03311 CYP71FA8
E. grandis Eucgr.G03310 CYP71FA7
E. camaldulensis EcS744762.10
E. grandis Eucgr.D01189 CYP71FA2
E. grandis Eucgr.D01190 CYP71FA3
E. camaldulensis EcC009148.10
E. grandis Eucgr.D01192 CYP71FA4
E. camaldulensis EcC030089.20
E. grandis Eucgr.D01193 CYP71FA5
E. grandis Eucgr.D01195 CYP71FA6
Subfamily FA

Fig. S9. *Eucalyptus* CYP71 phylogeny. The third step in prunasin biosynthesis catalyzed by CYP71B enzymes is shown next to the phylogenetic tree. Full-length amino acid sequences were aligned using MUSCLE and phylogeny was inferred using the Maximum Likelihood method with n=1000 bootstrap replicates. The branch lengths are measured in number of substitutions per site. *Approximately 16 amino acids are missing from the N-terminal of CYP71 from *E. leucoxylon*.

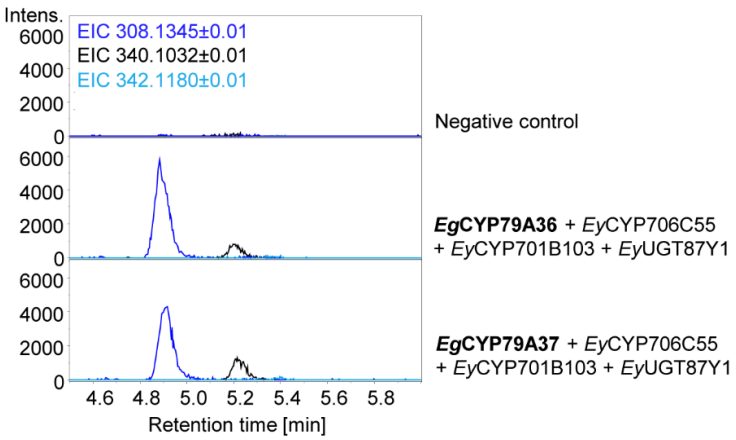

Fig. S10. Transient expression of *Eg*CYP79A36 and *Eg*CYP79A37 from the acyanogenic *E. grandis*. Extracted ion chromatograms (EIC) showing small levels of *m/z* 308.1345. This compound could potentially be 3-methylbutyraldoxime (leucine-derived) or 3-methylbutyraldoxime glucoside (isoleucine-derived). Accumulation of *m/z* 308.1345 was not detected upon co-expression of *Ecl*CYP79A125, *Ey*CYP79A34 or *Ecam*CYP79A34. The EIC of prunasin (*m/z* 340.1032; $[M+FA]^-$) is shown for comparison. Phenylacetaldoxime-glucoside (m/z 342.1180; $[M+FA]^-$) does not accumulate to detectable levels due to conversion of phenylacetaldoxime by downstream co-expressed prunasin pathway genes.

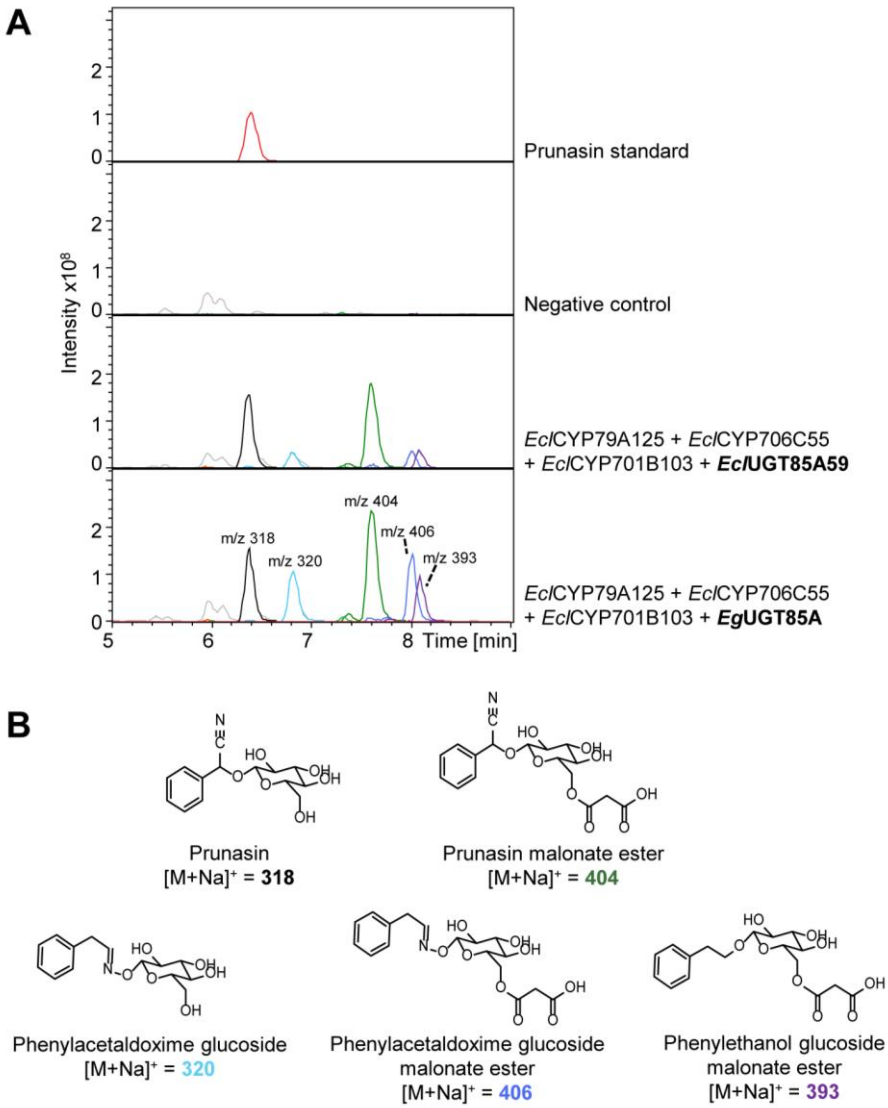

Fig. S11. Comparison of product profiles upon transient co-expression of *Ecl*UGT85A59 or *Eg*UGT85A with the prunasin CYPs from *E. cladocalyx*. **A.** LC-ion trap chromatograms of extracts from *N. benthamiana* leaves. The base peak chromatogram is shown in light grey and specific extracted ion chromatograms (based on $[M+Na]^+$) are colored according to *m/z* value. **B.** Structures of putatively annotated compounds of the major peaks detected in the chromatograms. Presence of prunasin in the extracts was confirmed by comparison to an authentic standard. Additional compounds are predicted based on their observed *m/z* value for $[M+Na]^+$ adducts, MS2 fragmentation pattern and predictions from literature (Hansen *et al.*, 2018).

Table S1. Primers used to isolate target open reading frames from cDNA. *att*B overhangs used for subsequent Gateway recombination into pDONR207 are underlined. Abbreviations: F, forward primer; R, reverse primer.

Gene	Primer sequences (5' -> 3')	F/R
Ecam/EyCYP79A34	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGAGCTCCGCCACCC	F
Ecam/EyCYP79A34	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAAGCAAGGTAGAGATGACTTGG	R
Ecam/EyCYP706C55	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGTCTCCGTCCATCTCCCT	F
Ecam/EyCYP706C55	GGGGACCACTTTGTACAAGAAAGCTGGGTTCATTCATAAAGCCGAGAATCAGAGAAC	R
Ecam/EyCYP71B103	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGCTCTCGACATCCTCTTC	F
Ecam/EyCYP71B103	GGGGACCACTTTGTACAAGAAAGCTGGGTTCATACATTGGTTTCATGCGATGGC	R
Ecam/EyUGT87Y1	ATGGAGTCGAAACCCACCC	F
Ecam/EyUGT87Y1	TCATTCGGCCCTAGATATGTCAC	R

SI References

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 Physiol* **123**(4): 1437-1448.

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 Physiol* **178**(3): 1081-1095.

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 of the United States of America* **102**(5): 1779-1784.

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 Physiol* **178**(3): 1096-1111.

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**(34): 11692-11700.