Plant Molecular Biology

Title: Functional investigation of five R2R3-MYB transcription factors associated with wood development in *Eucalyptus* using DAP-seq-ML

Lazarus T. Takawira[1]*, Ines Hadj Bachir[2]*, Raphael Ployet[1], Jade Tulloch[1], Helene San Clemente[2], Nanette Christie[1], Nathalie Ladouce[2], Annabelle Dupas[2], Avanish Rai[1], Jacqueline Grima-Pettenati[2], Alexander A. Myburg[1], Eshchar Mizrachi[1], Fabien Mounet[2†], Steven G. Hussey[1‡]

*These authors contributed equally to the study

[†]Corresponding author: fabien.mounet@univ-tlse3.fr

[‡]Corresponding author: steven.hussey@fabi.up.ac.za

[1]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa

[2]Laboratoire de Recherche en Sciences Végétales, Université Toulouse, CNRS, INP, Castanet-Tolosan, France.

Supplementary information

Supplementary Tables

Table S1. Summary of clean paired-end reads and read mapping statistics for the DAP-seq libraries. E1-E4 represents four separately prepared *E. grandis* minimal adapter-ligated genomic DNA libraries.

Laboratory	Gene ID	Alias	Library	PCR cycles	Paired Clean Reads	Mapped reads	Mapping rate %	De-duplicated library size	% of deduplicated reads
1	Eucgr.G01774	EgrMYB1	E1	20	8 459 247	3 508 050	41	297 560	8
			E2	20	47 796 194	37 515 233	78	2 477 731	7
			E2	15	565 560	441 193	78	244 483	55
			E1	15	59 960	48 400	81	31 309	65
2	Eucgr.G01774	EgrMYB1	E3	20	1 424 006	1 205 706	85	998 508	83
			E3	20	2 202 568	1 854 783	84	1 166 856	63
			E3	20	1 823 308	1 534 861	84	669 680	44
1	Eucgr.G03385	EgrMYB2	E1	20	19 703 924	15 995 646	81	741 068	5
			E2	15	4 997 588	3 643 242	73	1 300 865	36
			E1	15	157 543	125 719	80	69 945	56
			E2	20	10 198 661	7 988 611	78	4 225 914	53
2	Eucgr.G03385	EgrMYB2	E3	20	942 046	759 006	81	671 512	88
					918 579	744 049	81	684 352	92
					836 500	679 824	81	608 468	90
1	Eucgr.J01601	EgrMYB122	E1	20	7 427 557	6 141 847	83	438 956	7
				15	48 711	39 217	81	32877	84
1	Eucgr.K02297	EgrMYB135	E1	20	11 332 350	9 494 243	84	505 472	5
			E1	15	97 244	78 126	80	43 284	55
			E2	15	4 246 804	3 329 494	78	1 218 500	37
			E2	20	12 609 139	10 142 791	80	2 638 144	26
2	Eucgr.K02806	EgrMYB137	E4	20	22 077 608	17 662 086	80	3 142 078	18

					30 964 771	25 081 465	81	1 776 990	7
					22 390 144	17 912 115	80	2 678 441	15
2	-	HaloTag	E3	20	3 362 527	2 413 678	72	1 495 756	62
					5 950 633	4 879 519	82	331 444	7
1	-	HaloTag	E1	20	9 377 264	7 763 437	83	474 657	6
			E1	15	103 772	82 852	80	43 103	52
			E2	20	33 700 052	26 585 971	79	2 421 002	9
			E2	15	1 840 203	1 458 177	79	663 116	45

Table S2: Descriptions of machine learning features obtained for *Arabidopsis thaliana*

Feature	Range	Description
Peak score	[102,1000]	A normalized value indicating how dark the DAP-seq peak would be displayed in a genome browser (i.e., the degree of enrichment of the peak over background)
Peak signal value	[5,6925.3]	Measurement of average enrichment for the DAP-seq peak region
Peak q-value	[2,999]	Measurement of statistical significance of the DAP-seq peak, using false discovery rate (-log10)
Distance to TSS	[-5100,5095] bp	The bp distance between the middle of the DAP-seq peak (peak summit) and the gene TSS
Absolute distance to TSS	[0,5100]	The absolute bp distance between the middle of the DAP-seq peak and the gene TSS, regardless of orientation relative to the gene
Motif score	[0,47.18]	Log-likelihood ratio score for the occurrence of a TF binding motif in a DAP-seq peak for that TF, assuming a null model in which sequences are generated at random
Motif p-value	[4.31×10^{-17},1]	Measurement of statistical significance of the motif score
Motif E-value	[8.3×10^{-11},18000]	The expected number of sequences generated at random, that would match the motif as well as the DAP-seq peak sequence does
Binding profile raw score Binding profile raw score signal value Binding profile Z score Binding profile Z score signal value Binding profile p- value score Binding profile p-value signal value	[0.01,78.22] [0.02,472.46] [-1.22,19.31] [-0.83,27.20] [1.96×10^{-83},0.89] [2.89×10^{-163},0.80]	Features relating to the regulatory "scores" or relationships between TFs and target genes, based on the binding profile of each TF.
CNS-DAP overlap (bp) CNS-DAP overlap (proportion of peak length	[0,160] bp [0,0.79]	Overlap of DAP-seq peak region with evolutionarily conserved noncoding sequences, measured in bp and as a proportion of the DAP-seq peak width
DHS-DAP overlap (bp) DHS-DAP overlap (proportion of peak length)	[0,202] bp [0,1]	Overlap of DAP-seq peak region with tissue-specific accessible chromatin regions, measured in bp and as a proportion of the DAP-seq peak width

CNS-DHS-DAP overlap (bp) CNS-DHS-DAP overlap (proportion of peak length)	[0,202] [0,1]	Overlap of DAP-seq peak region, evolutionarily conserved noncoding sequences and tissue-specific DNase-seq peak regions, measured in bp and as a proportion of the peak width
Pearson correlation coefficient Spearman correlation coefficients	[-0.71,1.0] [-0.70,1.0]	Measurements of the relationship between the TF-encoding and TF target gene expression levels. Pearson correlation measures the linear relationship between two genes' expression levels, while a monotonic relationship is evaluated by the Spearman correlation. See Dataset S1.

Table S3: List of the primers used in this study

Gene name	Accession number	sequence primer forward (5' ->3')	sequence primer reverse (5' ->3')
Cloning			
EgrMYB137	Eucgr.K02806.1	CACCTTTCCTGCGCAGTTCTCAAT	GAAGATGTTGAAATGTGCCTGTA
Expression analysis			
PtVNS9 / PtrWND2A	Potri.014G104800	ACAACTGGGCAACCCTTGATCG	GAGGTTTCAGTCTGGCCATTAAGC
PtVNS10 / PtrWND2B	Potri.002G178700	TGGCCTTAACAACTGGGTTGCC	AGAGGTTTCAGCCTGGCCATTG
PtVNS11/ PtrWND1B	Potri.001G448400	TGATGACAGCACCAACGACAC	TCCTGAACCAAGAATCTGATGAGC
PtVNS07 / PtrWND6A	Potri.013G113100	ATGGGACATCCAAGCCAAGTGC	TGTTCCGGTCGGATACTTCCTATC
PtrMYB2	Potri.001G258700	TGCCAATGAACATGGATCCGTCAC	GCTTGCATGGAGGTTTCCAACG
PtrMYB21	Potri.009G053900	ACATGTATGGCGGTGCAAGTGG	AAGGACTCCAACTTGAGCCATGC
PtrMYB3	Potri.001G267300	AGGAATGCTGGCTTGCAAAGGT	GCCACGCTTGAGGTCAGGTC
PtrMYB20	Potri.009G061500	TCCATGCTCAACAACCGCTATG	ACCCACTTGTGTCAAGCATGTAG
PtrKNAT7	Potri.001G112200	CCAACTGAAGATGACAAAGCA	CTTGGACTTCAAGGATGTCA
PtrMYB156	Potri.009G134000	CCAGCCTGATCATCACCAGCCA	GTGCTGCCACTGCTCCCAACA
PtrMYB221	Potri.004G174400	TGCTTGCAGCTTGGGGCTACA	AGCCAGTTTTGGAGCCAGCACT
PtXND1	Potri.005G064100	CTGTTCCAGGAGTTTAAA	CACATGATTCATTAACATATAAGC
PtrMYB92	Potri.001G118800	GCATGGAATTGAGCGCACTTAGC	TCATCGCTCTAGCGTACACACG
PtrMYB125	Potri.003G114100	TGTAGTACCGAGGATTCGCTCTTG	CATCAACTAGCGGAGGCTCATC
PtrMYB10	Potri.001G099800	ACAGATGGGCTCACATAGCAAGC	TCGGTTCCGGTAGTTGCATTGG
PtrMYB128	Potri.003G132000	ACAGATGGGCTCATATTGCAAGTC	TTATTGTTGCAGGTGCTGAAGGC
PtrMYB216	Potri.013G001000	TTGTGGGAAATCCAGCGACGAAG	GTCTTCAACCGTCTCGTTTGCG
PtrMYB170	Potri.005G001600	AGCAACACCAATTTCTACGAGAGC	CGCCGGATTTCCCACAATTAGC
PtrCesA4	Potri.002G257900	GGTGCATCCATGCTCCTTTT	GAACCCACCTTCTAGCAAACTCA
PtrCesA7A	Potri.006G181900	CATGTACATATTGCTTCTATGGG	GCATCTCCATTTTGTCGGTTAC

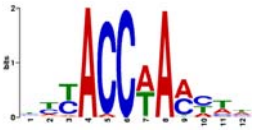
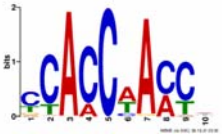
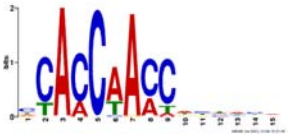
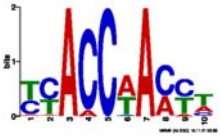
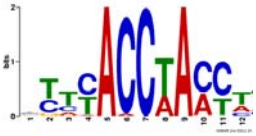
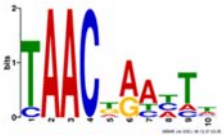
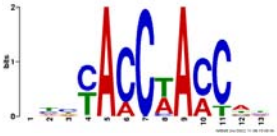
PtrCesA7B	Potri.018G103900	GATACATGTGCATCCTGCTTCTAA	CATCTCCATCTTAGTCAGTTTATAC
PtrCesA8A	Potri.011G069600.1	TTCGATCCAGATTCTACTTTCTCAT	CATCTCGCAGTTCATGTAACTC
PtrCesA8B	Potri.004G059600	TTGCTGAGCTACCTCCAATAAG	AGGGAAACTACAACGAGGATCA
PdGATL1.1/PARVUS	Potri.014G040300.1	GGAGTGGATGGAACTACAGA	GCAAGAGACTAACAGGACCA
PtrGAUT12.2	Potri.011G132600	CATTTCAATGGTCGAGCA AAGCCTTGGC	GACAGCCCGT AATGAACTTGTCAGA
PtrGXM1	Potri.004g226800	ACTGCCATATTTACAGCTGGAATG	CTACTCTGGGCAAAAGGGTCTGTC
PtrGXM2	Potri.003g003800	AGAATGACTGCCATATATACTGCC	CTACTCTGGGCAAAAAGGTCTGTC
PtrGXM3	Potri.013g102200	TGACGAGGCACCAGGGAGGATGAC	CTAAGGACAAAAAGGTCTGCCCGA
PtrGXM4	Potri.019g076300	TGATGAGGCTCCAGGGAGAATGAA	TGGACAAAAAGGTTTTCCCGAACT
PtrGT43A/IRX9	Potri.006G131000	GTCGCCCTTCATCTGTCC	TCCCTCATAGTTTTCTCCTGCT
PtrGT43B/IRX9	Potri.016G086400	GTCGCCCTTCTTCAGTCCAG	TTTTGTCTTCTTGATTTTCCTGA
PtrGT47c /FRA8	Potri.009G006500	CCACGTGGCAGGTGCTTTATG	CATGTCTCTTAGCTGGCAAAAGCG
PtrIRX10	Potri.001G068100	GGCTCGTAAGTTGCCGCAC	ACTTCTCTTACCAAGGTTTCAGGTCC
PtrRWA-A	Potri001G352300	CCCACAAAAGACAATAAGCGG	ATGGAGGTTTTTTCTTCAACTTTG
PtrPAL1	Potri.006G126800.1	CCATCCAGGTCAAATTGAGGCTGCT	ACTTCTTAGCTGCCTTCATGTAAGCT
PtrPAL2	Potri.008G038200.1	CCTAGAAGCCATCACCAAGTTGCTC	GTTTCTCCATTGGGTCCCACG
PtrPAL3	Potri.016G091100	CATCCAGGTCAAATTGAGGCTGCA	ACTTCTTAGCTGCCTTCATGTAAGC
PtrC4H1	Potri.013G157900	AGTGCGCCATAGACCATATCCTC	ATTGCAGCGACGTTGATGTTCTCA
PtrC4H2	Potri.019G130700	GAAATGTGCAATTGATCATATTTTG	ATTGCAGCAACATTGATGTTCTCC
Ptr4CL3	Potri.001G036900	ACTAGCCCATCCAGAGATATCCGA	TCATCTTCGGTGGCCTGAGACTTT
Ptr4CL5	Potri.003G188500	GTGATCATGCTCATCCTGCCAAGT	TTGGCAGCAGTAGTAATGGCACCT
PtrHCT1	Potri.003G183900	ATCAGCATGTAAGGCACGCGG	TGCCAAAGTAACCAGGTGGAAGCGT
PtrC3H3	Potri.006G033300	GTATGACCTTAGTGAAGACACAATCAT	CCCTTGGGTTCTTGATTAGCTC
CSE1	Potri.003G059200	AGCGAGCCATACAAGAACTTGCC	ATCGTTGCTAGTCCGCCCATTG
CSE2	Potri.001G175000	GGTATACGTTGCTACATGGGTGAC	TGGTAAGCCCTTGTATGGTTCGC

PtrCCoAOMT1	Potri.009G099800	CAGTAATTCAGAAAGCTGGTGTTGC	GCATCCACAAAGATGAAATCAAAAC
PtrCCoAOMT2	Potri.001G304800	CCTTCCAACGCCAGGAAAGAGAGTA	GTGGCCAACTTCTTGATGCCTTCCG
PtrCCR2	Potri.003G181400	CGGTGATTCAGAAAGCTGGTCTGGA	GCATCCACAAAGATGAAGTCATAAG
PtrCAld5H2	Potri.007G016400	AAGCCAATATAGGCAAGCCTGTGAATC	ATTTTTAGCCCCGAAAGCTGCTCTG
PtrCOMT1	Potri.015G003100	AGCACAATCGTCTCCAAGTACCCT	AACATTCTCCACACCAGGGAAAGC
PtrCOMT2	Potri.012G006400	TCTTGAAGAATTGCTATGACGCCT	GAATGCACTCAACAAGTATCACCTTG
PtrCAD1	Potri.009G095800	GGCAAGCTGATCTTGATGGGTGTT	TCCCGGTGATTGACTTTCTCCCAA
PtrCAD5	Potri.009G062800	CTCCAGGAGGTCTACAGGAGAGAAT	GTCTGAGTGACATATTCCACAAAAC
HK1	Potri.005G101200	TCAACAAGAGTGGGCGGTTCAC	ACCTGCATAGGTCCAGCAAAGAC
HK3	Potri.009G037800	CGAGTTCGTGATCCACAAGGAAGC	TCCCTGGAGAAGTGAGCATGTTG
HK11	Potri.006G275300	TAGCAGCAGCTATGTGGCTGAG	TCAGCAGTGTTGGCCATGTCAC
UBQ	Potri.001G418500	GTTGATTTTTGCTGGGAAGC	GATCTTGGCCTTCACGTTGT
CDC2/ACT1	Potri.001G309500	GGTAACATTGTGCTCAGTGG	CTCGGCCTTGGAGATCCACA

Table S4. A comparison of EgrMYB DAP-seq peak sets called using GEM and MACS2 peak callers. T denotes technical replicates analysed under identical conditions, while E1-E4 represent the four separately prepared *E. grandis* minimal adapter-ligated genomic DNA and FRiP stands for Fraction of Reads in Peaks.

TF Candidate	Library	PCR Cycles	HaloTag Control				Final Peak Set
			GEM	FRiP %	MACS2	FRiP %	GEM
EgrMYB1-T2	E3	20	1,912	6.88	1,869	8.15	1,912
EgrMYB2	E2	15	16,795	34.10	6,480	30.08	17,467
EgrMYB2	E2	20	37,138	39.16	15,137	30.37	
EgrMYB2-T1	E3	20	5,420	26.31	6,170	31.02	
EgrMYB2-T2	E3	20	6,026	25.11	6,915	29.61	
EgrMYB2-T3	E3	20	4,626	23.08	6,451	28.73	
EgrMYB122	E1	20	1,778	13.15	3,409	19.23	1,778
EgrMYB135	E2	15	5,269	16.31	2,337	16.27	13,967
EgrMYB135	E2	20	10,724	19.46	6,411	19.82	
EgrMYB137-T1	E4	20	23,443	24.03	896	9.38	27,941
EgrMYB137-T2	E4	20	6,838	10.55	447	4.28	
EgrMYB137-T3	E4	20	21,964	20.84	689	7.12	

Table S5. Enriched motifs for EgrMYB1, EgrMYB2, EgrMYB122 & EgrMYB135 DAP-seq binding sites. The EgrMYB135 motif is from the two technical replicates that passed the QC and filtering threshold and likewise, the EgrMYB2 represents the 5 technical replicates from the two labs that passed the QC and filtering thresholds. Motifs of close orthologs in *Arabidopsis* are also included in the table.

MYB TF	*E. grandis* motif	*Arabidopsis* motif[1]
EgrMYB1	 EgrMYB1-T2	 AtMYB4
EgrMYB2	 EgrMYB2	 AtMYB83
EgrMYB122	 EgrMYB122_E1_20	 AtMYB43
EgrMYB135	 EgrMYB135	 AtMYB52
EgrMYB137	 EgrMYB137	 AtMYB61

[1] (O'Malley et al. 2016)

Table S6. Pairwise comparison of peak quality matrices and enriched motifs for EgrMYB2 using GEM. Overlapping peaks are indicate on the bottom left of the diagonal, which the significance of the overlap (P-value) is indicated at the top right. *FRiP – Fraction of Reads in Peaks.

	Top enriched motif	Library E2 15-cycles	Library E2 20-cycles	Library E1 15-cycles	Library E1 20-cycles	Library E3 20-cycles R1	Library E3 20 cycles R2	Library E3 20 cycles R3
Library E2 15 cycles			$2.84x10^{-1}$	$1.07x10^{-3}$	$4.18x10^{-2}$	$2.34x10^{-1}$	$2.5x10^{-1}$	$2.04x10^{-1}$
Library E2 20 cycles		15,263		$4.85x10^{-4}$	$2x10^{-2}$	$1.2x10^{-1}$	$1.37x10^{-1}$	$1.05x10^{-1}$
Library E1 15 cycles		22	23		$8.06x10^{-3}$	$3.32x10^{-3}$	$2.99x10^{-3}$	$3.89x10^{-3}$
Library E1 20 cycles		840	907	8		$1.1x10^{-1}$	$1.02x10^{-1}$	$1.11x10^{-1}$
Library E3 20 cycles R1		5,399	6,272	23	729		$4.51x10^{-1}$	$4.51x10^{-1}$
Library E3 20 cycles R2		5,674	6,871	24	733	4,289		$4.33x10^{-1}$
Library E3 20 cycles R3		4,603	5,365	26	706	3,823	3,845	
Peaks (FRiP)		n = 16789 (FRiP = 34,1%)	n = 37121 (FRiP = 39,2%)	n = 19 (FRiP = 1,8%)	n = 856 (FRiP = 3,9%)	n = 5422 (FRiP = 26,3%)	n = 6026 (FRiP = 25,1%)	n = 4626 (FRiP = 23,1%)

Table S7: Confusion matrix statistics for each base model considered

	Base model (nearest)	Base model (5 kb binding region)
Correctly predicted functional TF-gene associations	8 776	13 620
Correctly predicted non-functional TF-gene associations	270 026	219 334
Incorrectly predicted functional TF-gene associations	74 680	125 372
Incorrectly predicted non-functional TF-gene associations	27 402	22 558

Table S8. Independent *Arabidopsis* training/testing data splits that were considered when exploring the robustness of the random forest classifier.

Independent test no.	Samples in test set	Percentage of samples in UDGs test matrix	Percentage of samples in LEGs test matrix	Percentage of samples in RANDOMs test matrix
1	NAC4, RAV1	18.36%	18.06%	20.50%
2	HHO2, HHO3, VRN1	19.52%	20.15%	19.63%
3	RAV1, NAP	21.16%	20.01%	21.40%
4	RAV1, VRN1	20.98%	19.19%	19.19%
5	HB6, MYB61	19.79%	19.65%	17.99%

Table S9: Biological process gene ontology enrichment results and enrichment analysis for EgrMYB2

EgrMYB2	**pre-ML**	**post-ML**

	Baseline model		**no cut-off**		**Pr ≥ 0.5**		**Pr ≥ 0.7**	
GO Description	corrected p-val	Enrich ment	corrected p-val	Enrich ment	corrected p-val	Enrich ment	corrected p-val	Enrich ment
phenylpropanoid metabolic process	<0.001	1.88	<0.001	2.0	<0.001	2.8	<0.001	4.2
phenylpropanoid biosynthetic process	<0.001	1.90	<0.001	2.0	<0.001	3.2	<0.001	4.7
secondary metabolic process	<0.001	1.53	<0.001	1.6	<0.026	1.7	0.001	2.4
flavonoid biosynthetic process	<0.001	2.05	<0.001	2.1	0.046	2.7	-	-
cellular aromatic compound metabolic process	0.001	1.37	<0.001	1.4	<0.001	2.0	<0.001	3.0
flavonoid metabolic process	<0.001	1.95	<0.001	2.0	-	-	-	-
cellular amino acid derivative metabolic process	0.001	1.41	0.001	1.5	0.001	2.0	<0.001	2.9
aromatic compound biosynthetic process	0.002	1.44	0.004	1.5	<0,001	2.3	<0,001	3.6
cellular amino acid derivative biosynthetic process	0.004	1.43	0.005	1.5	<0.001	2.4	<0.001	3.2
lignin metabolic process	0.005	1.81	0.005	1.9	0.012	3.1	0.003	5.1
cellular amino acid and derivative metabolic process	0.030	1.22	0.007	1.3	-	-	0.006	2.0
small molecule biosynthetic process	0.045	1.18	0.007	1.2	-	-	-	-
response to stimulus	0.007	1.09	0.014	1.1	-	-	-	-
response to biotic stimulus	0.003	1.24	0.014	1.2	-	-	-	-
response to other organism	0.004	1.24	0.028	1.2	-	-	-	-
localization	-	-	0.037	1.1	0.008	1.3	-	-
regulation of RNA metabolic process	0.001	1.26	0.040	1.2	-	-	-	-
regulation of transcription, DNA-dependent	0.001	1.25	0.050	1.2	-	-	-	-
cell wall organization or biogenesis	-	-	-	-	<0.001	2.9	<0.001	3.7
plant-type cell wall organization or biogenesis	-	-	-	-	<0.001	2.9	<0.001	4.4
plant-type cell wall biogenesis	-	-	-	-	<0.001	3.5	<0.001	4.9

cell wall biogenesis	-	-	-	-	<0.001	3.2	<0.001	4.7
cellular cell wall organization or biogenesis	-	-	-	-	<0.001	3.1	<0.001	4.3
vesicle-mediated transport	-	-	-	-	<0.001	2.3	<0.001	3.3
cellular localization	-	-	-	-	<0.001	1.8	-	-
secondary cell wall biogenesis	-	-	-	-	0.001	4.7	<0.001	8.2
establishment of localization in cell	-	-	-	-	0.001	1.7	-	-
lignin biosynthetic process	-	-	-	-	0.003	4.2	0.003	6.4
intracellular transport	-	-	-	-	0.003	1.8	-	-
regulation of actin cytoskeleton organization	-	-	-	-	0.003	8.0	-	-
regulation of actin filament-based process	-	-	-	-	0.003	8.0	-	-
cellular carbohydrate metabolic process	-	-	-	-	0.004	1.7	0.016	1.9
cellular macromolecule localization	-	-	-	-	0.008	1.8	-	-
cellular cell wall macromolecule metabolic process	-	-	-	-	0.008	5.1	0.001	9.5
carbohydrate biosynthetic process	-	-	-	-	0.008	1.9	0.035	2.2
microtubule-based process	-	-	-	-	0.009	2.3	0.003	3.4
carbohydrate metabolic process	-	-	-	-	0.010	1.5	-	-
establishment of protein localization	-	-	-	-	0.010	1.7	-	-
protein transport	-	-	-	-	0.010	1.7	-	-
establishment of localization	-	-	-	-	0.011	1.3	-	-
cell wall polysaccharide metabolic process	-	-	-	-	0.016	4.0	0.008	6.5
intracellular protein transport	-	-	-	-	0.016	1.8	-	-
protein localization	-	-	-	-	0.016	1.6	-	-
transport	-	-	-	-	0.016	1.3	-	-
cell wall organization	-	-	-	-	0.016	2.3	<0.001	4.2
plant-type cell wall organization	-	-	-	-	0.016	2.8	<0.001	5.2
cellular protein localization	-	-	-	-	0.016	1.7	-	-
cellular component macromolecule biosynthetic process	-	-	-	-	0.022	4.8	0.006	8.7

cell wall polysaccharide biosynthetic process	-	-	-	-	0.022	4.8	0.006	8.7
cell wall macromolecule biosynthetic process	-	-	-	-	0.022	4.8	0.006	8.7
cell wall macromolecule metabolic process	-	-	-	-	0.026	2.9	0.001	9.5
actin cytoskeleton organization	-	-	-	-	0.035	2.8	-	-
microtubule-based movement	-	-	-	-	0.035	2.8	0.047	3.9
cytoskeleton organization	-	-	-	-	0.035	2.1	-	-
xylan metabolic process	-	-	-	-	0.036	4.3	-	-
hemicellulose metabolic process	-	-	-	-	0.036	4.3	-	-
vacuole organization	-	-	-	-	0.036	4.3	-	-
cellular carbohydrate biosynthetic process	-	-	-	-	0.036	1.9	0.029	2.5
actin filament-based process	-	-	-	-	0.036	2.5	-	-
extracellular matrix organization	-	-	-	-	0.036	6.4	0.050	10.5
extracellular structure organization	-	-	-	-	0.036	6.4	0.050	10.5
pentose metabolic process	-	-	-	-	0.036	3.7	-	-
xylan biosynthetic process	-	-	-	-	0.036	5.0	0.023	8.7
glucuronoxylan metabolic process	-	-	-	-	0.036	5.0	0.023	8.7
glucuronoxylan biosynthetic process	-	-	-	-	0.036	5.0	0.023	8.7
regulation of cytoskeleton organization	-	-	-	-	0.036	5.0	-	-
polysaccharide metabolic process	-	-	-	-	0.036	1.8	-	-
macromolecule localization	-	-	-	-	0.044	1.5	-	-
cell wall modification	-	-	-	-	-	-	0.006	3.8
rhamnose metabolic process	-	-	-	-	-	-	0.008	17.5
rhamnose biosynthetic process	-	-	-	-	-	-	0.008	17.5
cell wall assembly	-	-	-	-	-	-	0.024	13.1
cellulose microfibril organization	-	-	-	-	-	-	0.024	13.1
plant-type cell wall assembly	-	-	-	-	-	-	0.024	13.1
nucleotide-sugar metabolic process	-	-	-	-	-	-	0.026	6.2
regulation of transcription	0.001	1.18	-	-	-	-	-	-
regulation of biosynthetic process	0.001	1.17	-	-	-	-	-	-

regulation of cellular biosynthetic process	0.001	1.17	-	-	-	-	-	-
regulation of macromolecule biosynthetic process	0.002	1.17	-	-	-	-	-	-
regulation of cellular metabolic process	0.003	1.15	-	-	-	-	-	-
regulation of primary metabolic process	0.003	1.15	-	-	-	-	-	-
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.003	1.16	-	-	-	-	-	-
response to chemical stimulus	0.004	1.13	-	-	-	-	-	-
regulation of nitrogen compound metabolic process	0.006	1.15	-	-	-	-	-	-
regulation of gene expression	0.038	1.13	-	-	-	-	-	-
response to organic substance	0.045	1.14	-	-	-	-	-	-

Table S10: Biological process gene ontology enrichment results and enrichment analysis for EgrMYB1

EgrMYB1	**pre-ML**		**post-ML**					
	Baseline model		**No cut-off**		**Pr ≥ 0.5**		**Pr ≥ 0.7**	
GO Description	Corrected p-val	Enrich ment	Corrected p-val	Enrich ment	Corrected p-val	Enrich ment	Corrected p-val	Enrich ment
secondary metabolic process	<0.001	2.9	<0.001	3.1	0.001	3.3	0.018	3.7

phenylpropanoid metabolic process	<0.001	4.0	0.001	4.0	0.001	4.7	0.001	7.2
small molecule biosynthetic process	0.001	2.0	0.002	2.1	0.003	2.2	-	-
phenylpropanoid biosynthetic process	<0.001	4.5	0.002	4.2	0.002	4.8	0.001	8.4
cellular amino acid derivative metabolic process	<0.001	2.9	0.002	3.0	0.002	3.4	0.014	4.3
aromatic compound biosynthetic process	<0.001	3.2	0.004	3.2	0.002	3.8	0.006	5.4
cellular aromatic compound metabolic process	0.002	2.4	0.006	2.5	0.001	3.1	0.012	3.9
small molecule metabolic process	0.028	1.5	0.006	1.7	0.002	1.9	-	-
cellular amino acid and derivative metabolic process	0.008	2.0	0.006	2.2	0.002	2.6	0.004	3.5
cellular amino acid derivative biosynthetic process	0.001	3.1	0.013	3.0	0.007	3.5	0.012	5.2
response to wounding	0.010	3.0	0.015	3.2	-	-	-	-
carboxylic acid biosynthetic process	0.032	2.1	0.015	2.3	0.032	2.4	-	-
organic acid biosynthetic process	0.032	2.1	0.015	2.3	0.032	2.4	-	-
indole derivative catabolic process	0.032	13.8	0.024	17.0	0.011	23.0	-	-
cellular ketone metabolic process	0.032	1.7	0.024	1.8	0.034	1.9	-	-
oxoacid metabolic process	0.037	1.7	0.030	1.8	0.040	1.9	-	-
carboxylic acid metabolic process	0.037	1.7	0.030	1.8	0.040	1.9	-	-
organic acid metabolic process	0.037	1.7	0.030	1.8	0.040	1.9	-	-
lignin metabolic process	0.050	4.2	-	-	0.029	5.9	-	-
defense response by cell wall thickening	-	-	-	-	0.029	10.2	-	-
defense response by callose deposition in cell wall	-	-	-	-	0.029	10.2	-	-
aromatic compound catabolic process	-	-	-	-	0.032	9.4	-	-
callose deposition in cell wall	-	-	-	-	0.032	9.4	-	-
cell wall thickening	-	-	-	-	0.034	8.8	-	-
defense response by callose deposition	-	-	-	-	0.034	8.8	-	-
regulation of transcription in response to stress	-	-	-	-	0.034	30.6	-	-
glycosinolate catabolic process	-	-	-	-	0.034	30.6	-	-

glucosinolate catabolic process	-	-	-	-	0.034	30.6	-	-
indole glucosinolate catabolic process	-	-	-	-	0.034	30.6	-	-
coumarin metabolic process	-	-	-	-	0.034	30.6	0.012	73.2
coumarin biosynthetic process	-	-	-	-	0.034	30.6	0.012	73.2
S-glycoside catabolic process	-	-	-	-	0.034	30.6	-	-
callose localization	-	-	-	-	0.036	8.2	-	-
polysaccharide localization	-	-	-	-	0.040	7.7	-	-
defense response to virus	-	-	-	-	0.040	7.7	-	-
response to UV-B	-	-	-	-	0.049	5.5	-	-
vesicle-mediated transport	-	-	-	-	-	-	0.022	4.4
monocarboxylic acid metabolic process	0.031	2.1	-	-	-	-	-	-
fatty acid metabolic process	0.032	2.4	-	-	-	-	-	-
regulation of transcription, DNA-dependent	0.006	1.9	-	-	-	-	-	-
regulation of RNA metabolic process	0.006	1.9	-	-	-	-	-	-
regulation of transcription	0.032	1.5	-	-	-	-	-	-

Table S11: Biological process gene ontology enrichment results and enrichment analysis for EgrMYB137

EgrMYB137	**pre-ML**		**post-ML**					
	Baseline model		**no cut-off**		**Pr ≥ 0.5**		**Pr ≥ 0.7**	
GO Description	corrected p-val	Enrichment	corrected p-val	Enrichment	corrected p-val	Enrichment	corrected p-val	Enrichment
response to nematode	0.023	1.5	0.008	1.6	-	-	-	-

flavonoid biosynthetic process	0.034	1.5	0.025	1.6	-	-	-	-
phenylpropanoid biosynthetic process	0.003	1.4	0.001	1.5	0.003	2.3	0.047	2.8
phenylpropanoid metabolic process	0.001	1.4	<0.001	1.5	0.021	1.9	-	-
cellular amino acid derivative biosynthetic process	0.027	1.3	0.009	1.3	-	-	-	-
secondary metabolic process	0.003	1.2	0.001	1.3	-	-	-	-
cellular amino acid derivative metabolic process	0.027	1.2	0.008	1.3	-	-	-	-
cellular aromatic compound metabolic process	-	-	0.030	1.2	0.011	1.6	-	-
monocarboxylic acid metabolic process	-	-	0.040	1.2	-	-	-	-
response to osmotic stress	-	-	0.037	1.2	-	-	-	-
response to biotic stimulus	0.023	1.2	0.007	1.2	-	-	-	-
response to other organism	0.027	1.2	0.008	1.2	-	-	-	-
cellular amino acid and derivative metabolic process	-	-	0.037	1.2	-	-	-	-
regulation of biological quality	-	-	0.026	1.2	0.014	1.4	-	-
small molecule biosynthetic process	-	-	0.019	1.2	-	-	-	-
multi-organism process		-	0.025	1.2	-	-	-	-
response to chemical stimulus	0.036	1.1	0.013	1.1	-	-	-	-
cell wall biogenesis	-	-	-	-	<0.001	3.3	0.002	4.1
plant-type cell wall biogenesis	-	-	-	-	<0.001	3.5	0.002	4.3
vesicle-mediated transport	-	-	-	-	<0.001	2.3	0.002	2.8
cellular cell wall organization or biogenesis	-	-	-	-	<0.001	3.0	0.004	3.8
cell wall organization or biogenesis	-	-	-	-	<0.001	2.3	0.004	2.7
plant-type cell wall organization or biogenesis	-	-	-	-	<0.001	2.7	0.041	2.8

microtubule-based process	-	-	-	-	<0.001	2.5	0.028	3.0
cellular localization	-	-	-	-	<0.001	1.7	-	-
intracellular transport	-	-	-	-	<0.001	1.8	-	-
establishment of localization in cell	-	-	-	-	0.001	1.7	-	-
cellular macromolecule localization	-	-	-	-	0.001	1.8	-	-
cellular protein localization	-	-	-	-	0.001	1.9	-	-
intracellular protein transport	-	-	-	-	0.001	1.9	-	-
establishment of protein localization	-	-	-	-	0.001	1.7	-	-
protein transport	-	-	-	-	0.001	1.7	-	-
protein localization	-	-	-	-	.002	1.7	-	-
regulation of cellular component size	-	-	-	-	0.003	1.8	-	-
regulation of anatomical structure size	-	-	-	-	0.003	.8	-	-
cytoskeleton organization	-	-	-	-	0.003	2.3	-	-
secondary cell wall biogenesis	-	-	-	-	0.004	3.8	0.047	5.4
regulation of cell size	-	-	-	-	0.004	1.8	-	-
localization	-	-	-	-	0.006	1.3	-	-
microtubule-based movement	-	-	-	-	0.006	2.9	-	-
regulation of actin cytoskeleton organization	-	-	-	-	0.006	6.4	-	-
regulation of actin filament-based process	-	-	-	-	0.006	.4	-	-
establishment of localization	-	-	-	-	0.006	1.3	-	-
transport	-	-	-		0.006	1.3	-	

				-				-
polysaccharide metabolic process	-	-	-	-	0.008	1.9	-	-
carbohydrate biosynthetic process	-	-	-	-	0.009	1.8	-	-
cell growth	-	-	-	-	0.010	1.7	-	-
cellular carbohydrate biosynthetic process	-	-	-	-	0.011	1.9	-	-
growth	-	-	-	-	0.013	1.6	-	-
carbohydrate metabolic process	-	-	-	-	0.014	1.4	-	-
pectin metabolic process	-	-	-	-	0.018	4.1	-	-
cellular cell wall macromolecule metabolic process	-	-	-	-	0.018	4.1	.041	7.0
multidimensional cell growth	-	-	-	-	0.022	2.8	-	-
S-adenosylmethionine biosynthetic process	-	-	-	-	0.025	6.4	-	-
macromolecule localization	-	-	-	-	0.027	1.4	-	-
polysaccharide biosynthetic process	-	-	-	-	0.027	2.0	-	-
aromatic compound biosynthetic process	-	-	-	-	0.030	1.7	-	-
actin filament-based process	-	-	-	-	0.032	2.3	-	-
cellular carbohydrate metabolic process	-		-		0.033	1.5	0.041	1.9

		-		-				
cellular polysaccharide biosynthetic process	-	-	-	-	0.033	2.0	-	-
phosphate metabolic process	-	-	-	-	0.036	1.3	-	-
phosphorus metabolic process	-	-	-	-	0.037	1.3	-	-
cell wall macromolecule metabolic process	-	-	-	-	0.037	2.6	-	-
lignin biosynthetic process	-	-	-	-	0.037	3.0	-	-
cell wall polysaccharide metabolic process	-	-	-	-	0.045	3.2	-	-
cellular polysaccharide metabolic process	-	-	-	-	0.046	1.8	-	-
L-ascorbic acid metabolic process	-	-	-	-	0.046	4.6	-	-
L-ascorbic acid biosynthetic process	-	-	-	-	0.046	4.6	-	-
actin cytoskeleton organization	-	-	-	-	0.047	2.5	-	-
cellular component macromolecule biosynthetic process	-	-	-	-	0.049	3.9	-	-
cell wall polysaccharide biosynthetic process	-	-	-	-	0.049	3.9	-	-
cell wall macromolecule biosynthetic process	-	-	-	-	0.049	3.9	-	-
extracellular structure organization	-	-	-	-	-	-	0.014	12.3
regulation of transcription	0.027	1.1	-	-	-	-	-	-
regulation of macromolecule biosynthetic process	0.027	1.1	-	-	-	-	-	-

regulation of biosynthetic process	0.027	1.1	-	-	-	-	-	-
regulation of cellular biosynthetic process	0.027	1.1	-	-	-	-	-	-
regulation of RNA metabolic process	0.034	1.1	-	-	-	-	-	-
regulation of transcription, DNA-dependent	0.034	1.1	-	-	-	-	-	-
response to organic substance	0.039	1.1	-	-	-	-	-	-
regulation of primary metabolic process	0..044	1.1	-	-	-	-	-	-

Table S12: Biological process gene ontology enrichment results and enrichment analysis for EgrMYB122

EgrMYB122	pre-ML		post-ML					
	Baseline model		no cut-off		Pr ≥ 0.5		Pr ≥ 0.7	
GO Description	corrected p-val	Enrichment	corrected p-val	Enrichment	corrected p-val	Enrichment	corrected p-val	Enrichment
secondary metabolic process	-	-	0.013	2.9	0.009	3.3	0.047	3.8
phenylpropanoid metabolic process	-	-	-	-	0.015	4.3	0.008	7.2
response to wounding	-	-	-	-	0.015	4.3	-	-
phenylpropanoid biosynthetic process	-	-	-	-	0.030	4.6	0.008	8.0
cellular aromatic compound metabolic process	-	-	-	-	0.039	2.8	0.006	5.0
cellular amino acid and derivative metabolic process	-	-	-	-	-	-	0.008	3.6
cellular amino acid derivative biosynthetic process	-	-	-	-	-	-	0.012	5.8
cellular amino acid derivative metabolic process	-	-	-	-	-	-	0.012	4.9
aromatic compound biosynthetic process	-	-	-	-	-	-	0.015	5.3
small molecule biosynthetic process	-	-	-	-	-	-	0.032	2.8

Table S13: Biological process gene ontology enrichment results and enrichment analysis for EgrMYB135

EgrMYB135	**pre-ML**		**post-ML**					
	Baseline model		**No cut-off**		**Pr ≥ 0.5**		**Pr ≥ 0.7**	
GO Description	corrected p-val	Enrichment	corrected p-val	Enrichment	corrected p-val	Enrichment	corrected p-val	Enrichment
response to other organism	<0.001	1.4	0.001	1.4	-	-	-	-
response to biotic stimulus	<0.001	1.4	0.001	1.3	-	-	-	-
multi-organism process	0.001	1.3	0.004	1.3	-	-	-	-
response to salicylic acid stimulus	0.001	1.6	0.011	1.6	-	-	-	-
response to stimulus	0.006	1.1	0.018	1.1	-	-	-	-
phenylpropanoid metabolic process	0.001	1.6	0.031	1.5	-	-	-	-
phenylpropanoid biosynthetic process	0.001	1.7	0.031	1.6	-	-	-	-
carbohydrate metabolic process	-	-	0.040	1.2	-	-	-	-
response to salt stress	-	-	0.042	1.3	-	-	-	-
response to chemical stimulus	0.026	1.1	0.043	1.1	-	-	-	-
response to osmotic stress	-	-	0.048	1.3	-	-	-	-
post-translational protein modification	-	-	0.048	1.2	-	-	-	-
response to abscisic acid stimulus	-	-	0.048	1.3	-	-	-	-
defense response	0.003	1.3	0.048	1.2	-	-	-	-
cellular response to endogenous stimulus	0.012	1.3	0.048	1.3	-	-	-	-
response to jasmonic acid stimulus	0.006	1.5	0.048	1.5	-	-	-	-
protein amino acid phosphorylation	0.038	1.2	0.048	1.2	-	-	-	-
response to arsenic	-	-	0.050	2.7	-	-	-	-
regulation of biosynthetic process	0.001	1.2	-	-	-	-	-	-
regulation of cellular biosynthetic process	0.001	1.2	-	-	-	-	-	-
regulation of nitrogen compound metabolic process	0.001	1.2	-	-	-	-	-	-
regulation of primary metabolic process	0.001	1.2	-	-	-	-	-	-

regulation of macromolecule biosynthetic process	0.001	1.2	-	-	-	-	-	-
regulation of cellular metabolic process	0.002	1.2	-	-	-	-	-	-
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.002	1.2	-	-	-	-	-	-
regulation of transcription	0.002	1.2	-	-	-	-	-	-
cellular amino acid derivative biosynthetic process	0.003	1.5	-	-	-	-	-	-
response to bacterium	0.006	1.4	-	-	-	-	-	-
regulation of metabolic process	0.006	1.1	-	-	-	-	-	-
regulation of macromolecule metabolic process	0.006	1.2	-	-	-	-	-	-
secondary metabolic process	0.006	1.3	-	-	-	-	-	-
aromatic compound biosynthetic process	0.007	1.4	-	-	-	-	-	-
cellular amino acid derivative metabolic process	0.012	1.4	-	-	-	-	-	-
regulation of gene expression	0.017	1.1	-	-	-	-	-	-
response to endogenous stimulus	0.027	1.2	-	-	-	-	-	-
response to hormone stimulus	0.032	1.2	-	-	-	-	-	-
cellular response to hormone stimulus	0.034	1.3	-	-	-	-	-	-
regulation of flavonoid biosynthetic process	0.034	2.4	-	-	-	-	-	-
defense response to bacterium	0.036	1.3	-	-	-	-	-	-
regulation of cellular process	0.048	1.1	-	-	-	-	-	-

Table S14. Most discriminant FT-IR wavenumbers associated with CW compounds from literature data. 13 wavenumbers related to CW compounds were identified from Sparse-PLS-DA loadings contribution to PC1 axis and PC2 explaining samples separation (see Figure 7)

Peak number	Contribution on	Associated WN in the literature	Associated CW compound	References
1	PC2	1948	lignin	(Salim et al. 2021)
2	PC2	1730-1742	lignin/hemicelluloses	(Faix 1991; Chang et al. 2014)
3	PC1	1596-1600	lignin	(Stark et al. 2016)
4	PC1	1510	lignin	(Faix 1991; Sammons et al. 2013; Largo-Gosens et al. 2014; Stark et al. 2016; Salim et al. 2021)
5	PC1	1447	lignin	(Salim et al. 2021)
6	PC1	1365-1372	lignin/cellulose/ hemicelluloses	(Chang et al. 2014)
7	PC1	1330	lignin	(Largo-Gosens et al. 2014; Stark et al. 2016)
8	PC1	1265-1275	lignin	(Sammons et al. 2013)
9	PC2	1047	pectin	(Largo-Gosens et al. 2014)
10	PC1	1030-1035	lignin	(Sammons et al. 2013)
11	PC2	966-990	lignin	(Sammons et al. 2013)
12	PC2	915-930	lignin	(Sammons et al. 2013)

13	PC2	863	lignin	(Stark et al. 2016)

Supplementary Figures

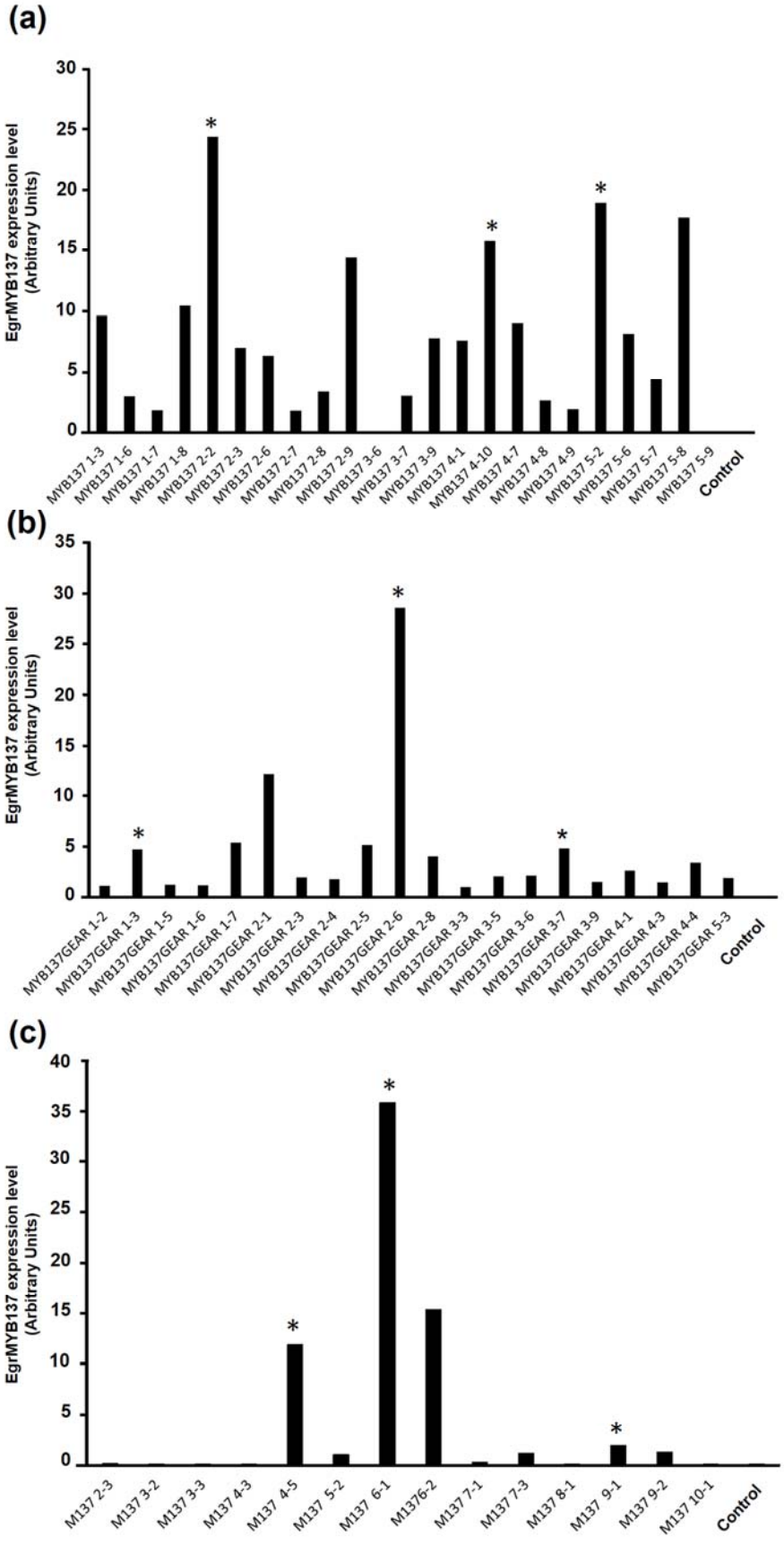

(a)
EgrMYB137 expression level (Arbitrary Units)
0
5
10
15
20
25
30
*
*
*
MYB137 1-3
MYB137 1-6
MYB137 1-7
MYB137 1-8
MYB137 2-2
MYB137 2-3
MYB137 2-6
MYB137 2-7
MYB137 2-8
MYB137 2-9
MYB137 3-6
MYB137 3-7
MYB137 3-9
MYB137 4-1
MYB137 4-10
MYB137 4-7
MYB137 4-8
MYB137 4-9
MYB137 5-2
MYB137 5-6
MYB137 5-7
MYB137 5-8
MYB137 5-9
Control
(b)
EgrMYB137 expression level (Arbitrary Units)
0
5
10
15
20
25
30
35
*
*
*
MYB137GEAR 1-2
MYB137GEAR 1-3
MYB137GEAR 1-5
MYB137GEAR 1-6
MYB137GEAR 1-7
MYB137GEAR 2-1
MYB137GEAR 2-3
MYB137GEAR 2-4
MYB137GEAR 2-5
MYB137GEAR 2-6
MYB137GEAR 2-8
MYB137GEAR 3-3
MYB137GEAR 3-5
MYB137GEAR 3-6
MYB137GEAR 3-7
MYB137GEAR 3-9
MYB137GEAR 4-1
MYB137GEAR 4-3
MYB137GEAR 4-4
MYB137GEAR 5-3
Control
(c)
EgrMYB137 expression level (Arbitrary Units)
0
5
10
15
20
25
30
35
40
*
*
*
M137 2-3
M137 3-2
M137 3-3
M137 4-3
M137 4-5
M137 5-2
M137 6-1
M1376-2
M137 7-1
M137 7-3
M137 8-1
M137 9-1
M137 9-2
M137 10-1
Control

Figure S1: Transgene expression level in *EgrMYB137* transgenic lines. A and B *EgrMYB137* expression level in *Arabidopsis* Pro35S:EgrMYB137-EAR dominant-repressive lines and Pro35S:EgrMYB137 overexpressing lines, respectively. C *EgrMYB137* expression level in *Populus* Pro35S:EgrMYB137-EAR dominant-repressive lines. *EgrMYB137* expression was measured in control lines (WT Col-0 *Arabidopsis* and *Populus* transformed with empty vector). Stars represent the 3 independent lines per construct chosen for functional characterization.

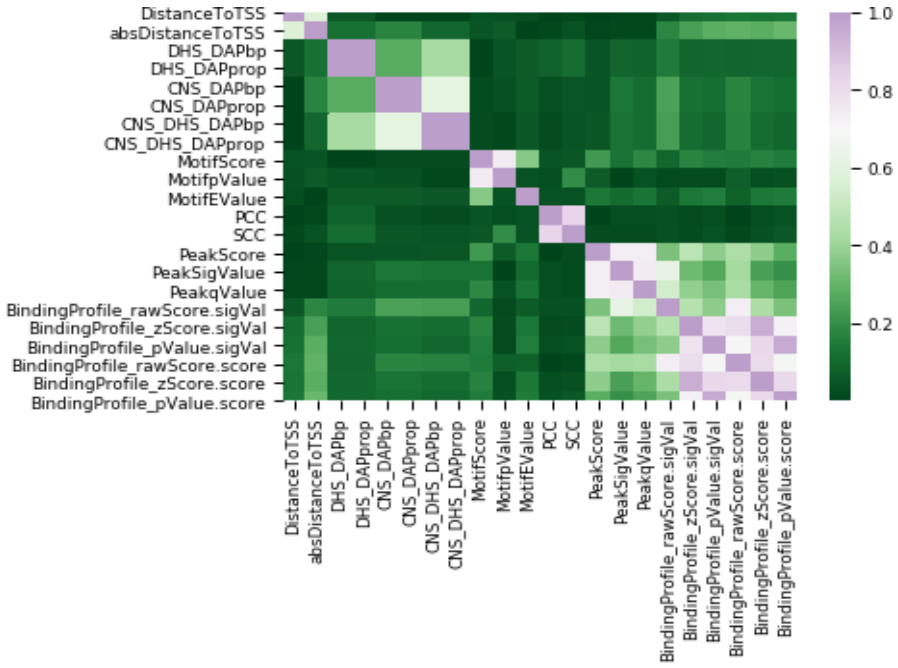


Figure S2. Evaluation of the absolute pairwise correlations between 22 machine learning features. The heatmap represents the absolute values of the Pearson correlation coefficients between each feature pair for the positive sample data. Each of the feature names are represented on the x- and y-axes. Highly correlated features were identified as those with an absolute pairwise correlation ≥ 0.7.

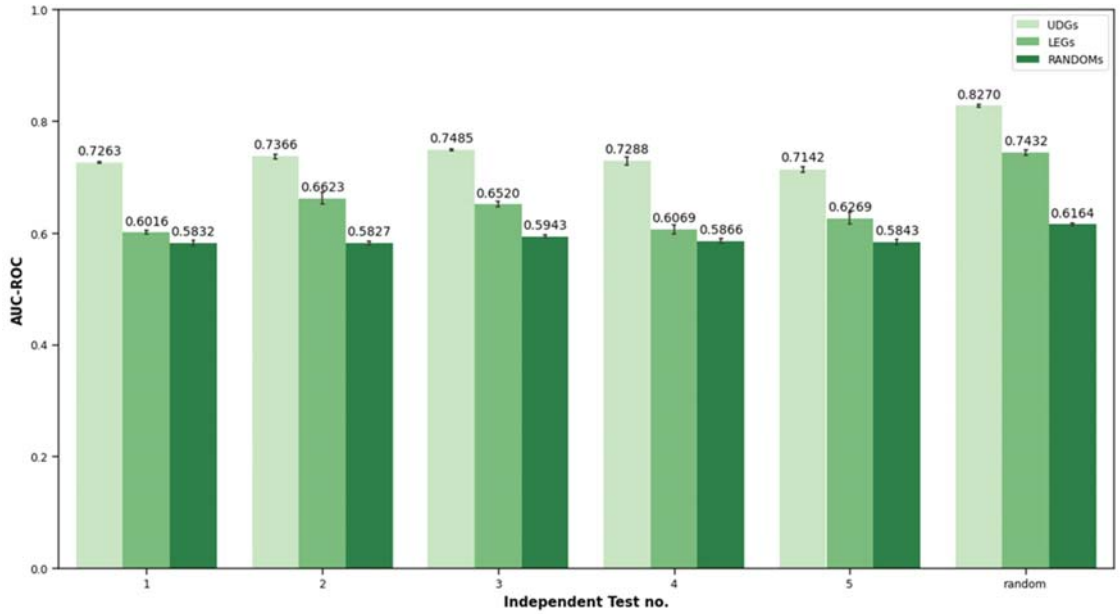


Figure S3. Evaluation of the random forest classifier performance (AUC-ROC) on independent training and testing subsets of *Arabidopsis* data. UDGs: undetected genes, LEGs: lowly expressed genes, RANDOMs: randomly selected genes. Lower case letters on each bar represent AUC-ROC scores that are statistically significantly different from each other for models trained on each of the different negative sample sets, based on post-hoc testing (Tukey HSD test, p-value < 0.05). Error bars represent one standard deviation and show variability in the AUC-ROC across the best-performing models for each of 5 cross-validation folds. For each of the feature matrices differing in negative samples, AUC-ROC was statistically significantly lower when the model is trained and tested on independent subsets of data, rather than randomly split training and testing sets ('random split').

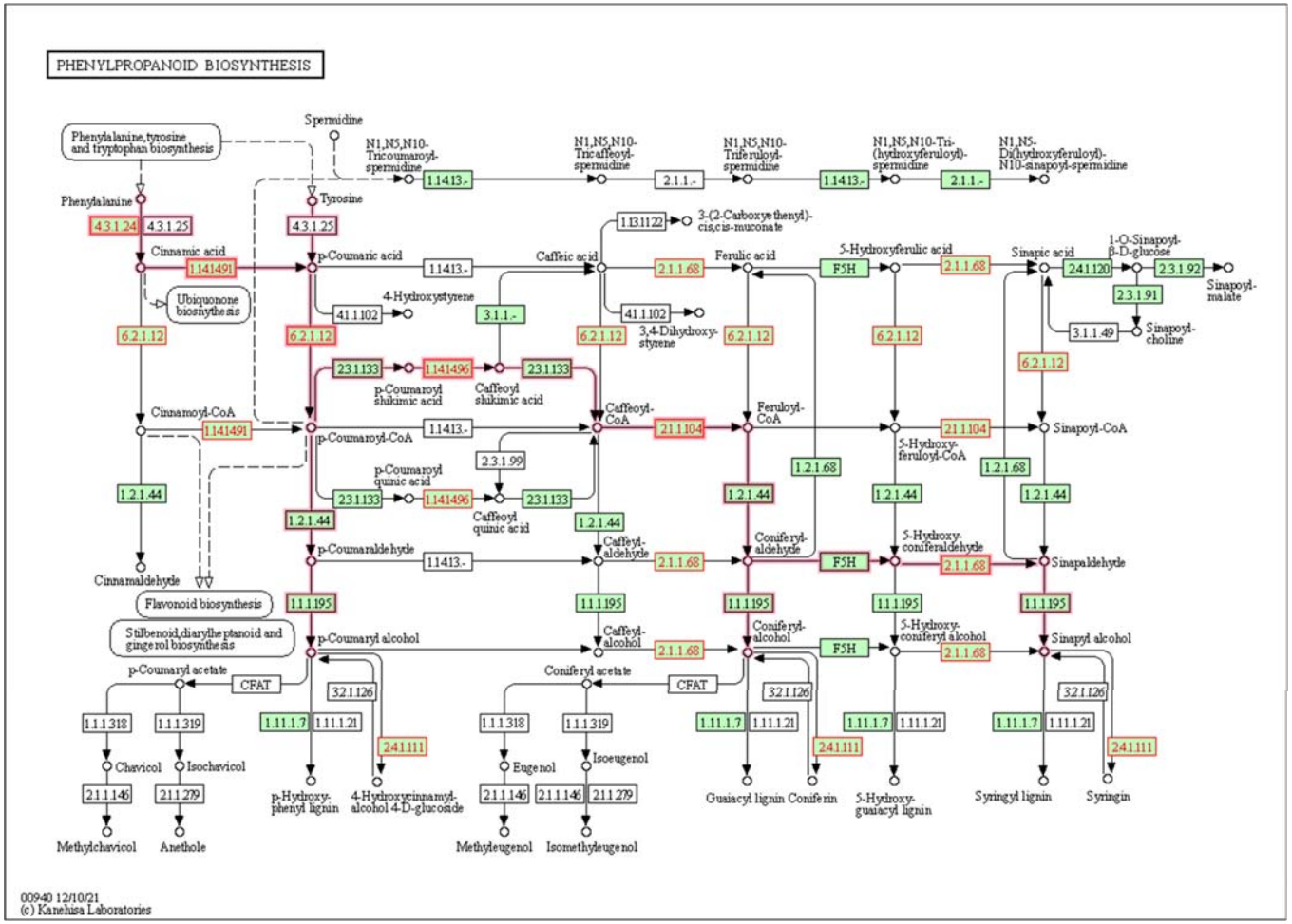


Figure S4: KEGG pathway analysis of the phenylpropanoid biosynthesis pathway showing predicted targets of EgrMYB122. EgrMYB122 gene targets selected at Pr ≥ 0.5 cut-off using the machine learning classifier.

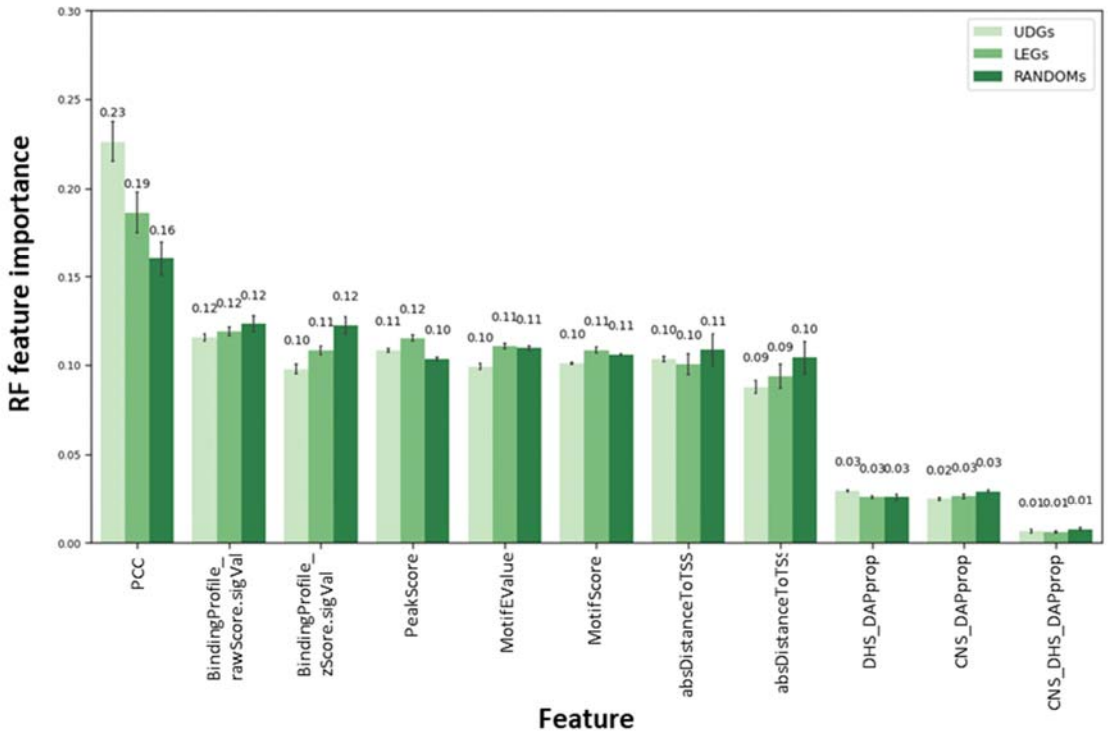


Figure S5. Evaluation of non-redundant feature importance for the random forest classifier and different negative training sample sets. UDGs: undetected genes, LEGs: lowly expressed genes, RANDOMs: randomly selected non-differentially expressed genes, RF: random forest. Error bars represent one standard deviation for cross-validation iterations (n = 5). Feature labels are described in Supplementary Table S2.

References

Chang S-S, Salmén L, Olsson A-M, Clair B (2014) Deposition and organisation of cell wall polymers during maturation of poplar tension wood by FTIR microspectroscopy. Planta 239:243–254. https://doi.org/10.1007/s00425-013-1980-3

Faix O (1991) Classification of Lignins from Different Botanical Origins by FT-IR Spectroscopy. Holzforschung 45:21–28. https://doi.org/10.1515/hfsg.1991.45.s1.21

Largo-Gosens A, Hernández-Altamirano M, García-Calvo L, Alonso-Simón A, Álvarez J, Acebes JL (2014) Fourier transform mid infrared spectroscopy applications for monitoring the structural plasticity of plant cell walls. Front Plant Sci 5. https://doi.org/10.3389/fpls.2014.00303

O'Malley RC, Huang SSC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR (2016) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165:1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Salim RM, Asik J, Sarjadi MS (2021) Chemical functional groups of extractives, cellulose and lignin extracted from native Leucaena leucocephala bark. Wood Sci Technol 55:295–313. https://doi.org/10.1007/s00226-020-01258-2

Sammons RJ, Harper DP, Labbé N, Bozell JJ, Elder T, Rials TG (2013) Characterization of Organosolv Lignins using Thermal and FT-IR Spectroscopic Analysis. Bioresources 8. https://doi.org/10.15376/biores.8.2.2752-2767

Stark NM, Yelle DJ, Agarwal UP (2016) Techniques for Characterizing Lignin. In: Lignin in Polymer Composites. Elsevier, pp 49–66