

A systems genetics analysis in *Eucalyptus* reveals coordination of metabolic pathways associated with xylan modification in wood forming tissues

Martin P. Wierzbicki¹, Nanette Christie¹, Desré Pinard¹, Shawn D. Mansfield², Eshchar Mizrachi¹ and Alexander A. Myburg¹

¹Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Private bag X20, Pretoria, 0028, South Africa

²Department of Wood Science, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada.

Author contributions:

M.P.W., E.M., S.D.M. and A.A.M. conceived and planned the research; M.P.W., N.C. and D.P. performed the research; M.P.W. analysed the data and drafted the manuscript; N.C., D.P., S.D.M., E.M. and A.A.M. helped draft and edit the manuscript.

Section	Word count
Main body	6492
Introduction	999
Materials and Methods	891
Results	2543
Discussion	1948
Acknowledgements	111

Authors for correspondence:

Prof. Eshchar Mizrachi and Alexander A. Myburg

Department of Biochemistry, Genetics and Microbiology

University of Pretoria,

Private bag X20, Pretoria, 0028, SOUTH AFRICA

Tel: AAM: (+27)-12-420-4945; EM: (+27)-83-789-0404

Email: AAM: zander.myburg@fabi.up.ac.za ; EM: eshchar.mizrachi@fabi.up.ac.za

Social Media

Twitter: @FMG_UP @Fabiteam1 @ResearchUP @zandermyburg @Eshmiz

ORCID IDs: AAM: [0000-0003-0644-5003](#), EM: [0000-0002-3126-9593](#), SDM: [0000-0002-0175-554X](#), NC: [0000-0003-4331-2103](#) DP: [0000-0001-5600-0139](#), MPW: [0000-0002-8614-1781](#)

Short heading: Systems genetics of xylan modification in *Eucalyptus*

Summary (200 words)

Acetyl and (methyl)glucuronic acid decorations of xylan, the dominant hemicellulose in secondary cell walls (SCWs) of woody dicots, affects its interaction with cellulose and lignin to determine SCW structure and extractability. Genes and pathways involved in these modifications may be targets for genetic engineering, however, little is known about the regulation of xylan modifications in woody plants.

To address this, we assessed genetic and gene expression variation associated with xylan modification in developing xylem of *Eucalyptus grandis* x *E. urophylla* interspecific hybrids. Expression QTL (eQTL) mapping identified potential regulatory polymorphisms affecting gene expression modules associated with xylan modification.

We identified 14 putative xylan modification genes that are members of five expression modules sharing seven trans-eQTL hotspots. The xylan modification genes are prevalent in two expression modules. The first comprises nucleotide sugar interconversion pathways supplying the essential precursors for cellulose and xylan biosynthesis. The second contains genes responsible for phenylalanine biosynthesis and S-adenosylmethionine biosynthesis required for glucuronic acid and monolignol methylation. Co-expression and co-regulation analyses also identified four metabolic sources of acetyl-CoA that appear to be transcriptionally coordinated with xylan modification.

Our systems genetics analysis may provide new avenues for metabolic engineering to alter wood SCW biology for enhanced biomass processability.

Key-words: secondary cell wall, xylan, cellulose, lignin, metabolism, systems genetics, dissolving pulp, *Eucalyptus*

Introduction

Plant secondary cell walls (SCW) are largely comprised of cellulose, hemicellulose and the heterogeneous phenolic biopolymer lignin. Lignocellulosic biomass from fast growing trees is currently primarily used for pulp and paper (Myburg *et al.*, 2014), but in recent years this biomass represents a putative feedstock for textiles, bioplastics, nanocellulose and second generation biofuels (Suzuki & Suzuki, 2014; Ragauskas *et al.*, 2014; Wierzbicki *et al.*, 2019). Approximately 70% of woody biomass is present in the polysaccharide fraction of cellulose and various hemicelluloses (Plomion *et al.*, 2001). In dicot plant SCWs, the dominant hemicellulose is glucuronoxylan (xylan); a linear molecule composed of repeating β 1-4 linked xylose residues, which are modified with side groups of α 1,2 linked D-glucuronic acid (GlucA) and 4-O-methylglucuronic acid (MeGlucA) units, as well as 2-O and/or 3-O linked acetyl units (Reviewed by Scheller & Ulvskov, 2010; Rennie & Scheller, 2014; Smith *et al.*, 2017). Xylan has been shown to be a major contributor to biomass recalcitrance (Petersen *et al.*, 2012; Pereira *et al.*, 2017), and an obstacle in the extraction of high purity chemical cellulose (in dissolving pulp). Additionally, the release of acetyl groups alters the pH of the reaction making subsequent fermentation of free glucose into ethanol for biofuel application unfavourable (Biely *et al.*, 1986; Zhang *et al.*, 2011). The importance of xylan in lignocellulosic biomass processing has stimulated numerous studies detailing its biosynthesis and function, however many unknowns persist such as how xylan biosynthesis and modification is co-regulated with central and

secondary metabolism (Persson *et al.*, 2007; Brown *et al.*, 2007; Mortimer *et al.*, 2010; Xiong *et al.*, 2013; Jensen *et al.*, 2014).

Xylan is synthesised in the Golgi apparatus by membrane bound proteins and upon completion is transported to the cell wall by vesicles (Brown *et al.*, 2011; Busse-Wicher *et al.*, 2014). When xylan arrives at the SCW, its conformation changes to facilitate association with cellulose microfibrils, stabilising and orientating them by crosslinking adjacent microfibrils (Simmons *et al.*, 2016), and creates hydrophobic pockets which facilitate the polymerisation of monolignols into lignin (Johnson *et al.*, 2017). The xylan backbone can adopt two conformations: a twofold screw known as the major domain, or a threefold screw known as the minor domain. The major domain is structurally similar to the twofold screw conformation of the cellulose microfibril (Simmons *et al.*, 2016). However, major domain xylan chains do not intertwine with each other, but rather enter grooves on the hydrophilic face of microfibrils and physically interact with them through hydrogen bonding, allowing the xylan to form a “coat” on the outer face of the cellulose (Simmons *et al.*, 2016; Pereira *et al.*, 2017). Modifications of the major domain are evenly substituted on alternate residues, resulting in these groups uniformly facing away from cellulose in a single direction (Bromley *et al.*, 2013). The substitutions at the O-2 position of xylose residues are vital for the formation of the twofold screw conformation, whereas substitutions at O-3 positions aid in stabilising the cellulose-xylan interaction (Pereira *et al.*, 2017). The minor domain may interact with the hydrophobic face of cellulose microfibrils, but may also potentially play a role in establishing the spacing between microfibrils and/or the angle at which they are orientated (Busse-Wicher *et al.*, 2014). The substitutions on this domain are unevenly spaced and closer together, which produce a threefold screw conformation where modifications face several directions (Bromley *et al.*, 2013). The functions of the substitutions at O-2 and O-3 positions in the minor domain are currently unclear, but the hydrophobic nature of the acetyl and MeGlucA may aid in the establishment of hydrophobic pockets in the spaces between microfibrils (Hao & Mohnen, 2014; Johnson *et al.*, 2017).

How these two distinct domains are synthesised is not fully understood, but genes associated with xylan backbone elongation (*IRX9*, *IRX10* and *IRX14*), reducing end addition (*IRX7*, *IRX7L*, *IRX8* and *PARVUS*) and backbone modification (members of *GUX*, *GXM*, *RWA* and *TBL*) have been identified (Reviewed by Rennie & Scheller, 2014; Marriott *et al.*, 2016; Smith *et al.*, 2017). Additionally, the metabolic precursors for xylan biosynthesis are known, namely: nucleotide sugars (needed for backbone, reducing end and glucuronic acid modifications),

acetyl-CoA (xylan acetylation) and S-adenosylmethionine (SAM; methylation of GlucA); (Rennie *et al.*, 2012; Urbanowicz *et al.*, 2012, 2014). The proteins involved in xylan backbone synthesis are thought to interact with each other creating the xylan synthase complex (XSC); (Zeng *et al.*, 2010, 2016; Jiang *et al.*, 2016). Whether proteins involved in xylan modification or reducing end synthesis also interact with each other, or are part of the XSC is currently unknown. The effect of metabolic precursor supply on xylan biosynthesis and the modification patterns on the backbone, or how the supply is regulated, has not been addressed. The rapid advancement of next generation “omics” approaches may be the key to understanding the regulation of major and minor domain formation by identifying the composition and stoichiometry of xylan biosynthetic machinery and the effect of metabolism on xylan modification patterns, which can be supplemented and validated using *in vitro* or *in planta* approaches (Mutwil *et al.*, 2009; Ruprecht *et al.*, 2011; Vanholme *et al.*, 2012; Urbanowicz *et al.*, 2014; Jiang *et al.*, 2016; Zeng *et al.*, 2016). Although gene co-expression studies have been performed to identify novel genes associated with xylan biosynthesis (eg. Oikawa *et al.*, 2010), co-regulation of xylan biosynthesis with the metabolic pathways supplying its precursors still needs to be further investigated (Wierzbicki *et al.*, 2019). One approach may be to investigate gene and trait co-expression and correlation across many time points and experiments (systems biology) or in segregating populations (systems genetics). This has been performed for lignin (Vanholme *et al.*, 2012) and cellulose (Mizrachi *et al.*, 2013; Yen *et al.*, 2013), and has yielded valuable knowledge about the regulation of the biosynthesis of these biopolymers.

Systems analysis involves incorporating multiple “omics” datasets such as genomic, transcriptomic, metabolomic and proteomic data to study the process or trait of interest (Vanholme *et al.*, 2012; Guerriero *et al.*, 2013; Li *et al.*, 2016; Christie *et al.*, 2017). A powerful approach related to systems biology is systems genetics, where genetic variation leads to coordinated expression variation in a structured or unstructured population (Civelek & Lusis, 2014; Mizrachi & Myburg, 2016; Zinkgraf *et al.*, 2017). This population level variation and covariation in gene expression is affected by genetic perturbations leading to altered protein or metabolite abundance, which often results in quantifiable differences in phenotypic traits of interest (Mizrachi & Myburg, 2016; Mizrachi *et al.*, 2017). An expression quantitative trait locus (eQTL) mapping approach can identify segregating genetic loci affecting gene expression. eQTL co-localisation can be used to identify shared regulatory factors often leading to “hotspot” loci affecting the expression of hundreds of genes (Breitling *et al.*, 2008). By using population wide gene co-expression clustering in combination with eQTL data, the systems

genetics approach can be used to better understand the genetic correlation of pathways involved in traits of interest (Ayroles *et al.*, 2009; Civelek & Lusi, 2014; Christie *et al.*, 2017; Wierzbicki *et al.*, 2019).

Here, using an integrated analysis of population wide developing xylem gene co-expression and eQTLs, we identified putative xylan modification genes in *Eucalyptus*, and describe five expression modules as well as seven regulatory loci associated with this process. This revealed new insights into the co-ordinated regulation of metabolic pathways such as nucleotide sugars interconversion, SAM production from amino acids, and metabolic sources of acetyl-CoA involved in xylan synthesis and modification. The understanding gained in the genetic regulation of xylan modification provides novel opportunities for metabolic rewiring for biotechnological improvement of woody biomass crop species.

Materials and methods

Identification of xylan modifying genes in *Arabidopsis thaliana*, *Eucalyptus grandis* and *Populus trichocarpa*

Using the online genomic database PLAZA 3.0 (Proost *et al.*, 2015), the terms “TBL”, “RWA”, “GUX” and “DUF579” in *Arabidopsis thaliana* were queried, and PLAZA’s homology tool was used to obtain the full predicted genomic set of putative xylan modification genes in *Arabidopsis*, *Populus* and *Eucalyptus*. Protein sequences from all four xylan modification associated families (**Supplementary file 1**) were then subjected to domain prediction tools to identify protein domain arrangement using the online service InterProScan (Jones *et al.*, 2014). Transcript abundance, measured in FPKM, for all identified genes from stem/xylem and leaf were obtained for *Arabidopsis* (Davin *et al.*, 2016), *Populus* (Hefer *et al.*, 2015) and *Eucalyptus* (Vining *et al.*, 2015). For *Arabidopsis* expression data, wild type cambium stem sections (Davin *et al.*, 2016) were used as a proxy for xylem. Due to the fact that the RNA-seq data was obtained from three separate studies, where sample preparation, sequencing depth and normalisation was not done in the same manner, the FPKM counts could not be directly compared between the three species. Therefore, two different comparisons were performed within each species, the first being xylem to leaf ratio: Log2 (Xylem FPKM/ Leaf FPKM), and secondly, the expression of each gene was converted to a percentile relative to that of the highest expressed gene (which was set to 100%) in the tissue after removing genes with no detectable expression (0 FPKM). Genes extracted from PLAZA that had no detectable

expression in the RNA-seq dataset, or on Phytozome V12.1 (Goodstein *et al.*, 2012), were removed from further analysis.

Maximum likelihood phylogenetic reconstruction

Sequence alignment was performed using the software package Simultaneous Alignment and Tree Estimator –II (SATE-II); (Liu *et al.*, 2012). Default settings were used for protein sequence (**Supplementary file 1**) alignment using MAFFT as the aligner, MUSCLE for merging, FASTTREE as the estimator, and the JTT+G20 modelling algorithm for 100 iterations using the SATé-II-ML quick set. Alignments produced by SATé-II were trimmed using trimAL v.1.2 (Capella-Gutiérrez *et al.*, 2009) making use of the automated1 function provided by the software. Phylogenetic reconstruction on this trimmed sequence file was performed using the phylogenetic software FastTree2 (Price *et al.*, 2010) using 1000 bootstraps, making use of the –spr 4, –gamma, –mlacc 2 and –pseudo options to obtain the best possible tree, as recommended by the authors (Price *et al.*, 2010). Phylogenetic reconstruction was combined with RNA-seq derived expression data and predicted protein domain structure to improve candidate selection due to a large number of tandem duplicates present in the *E. grandis* genome. Additionally, synteny analysis was performed using the “local gene organisation for homologous genes” function in PLAZA to aid in assigning names to the xylan modification candidates. This information was combined and visualised using the Interactive Tree Of Life (iTOL) v3 (Letunic & Bork, 2016). In the study, we employed the naming convention suggested by Kumar *et al.* for the *CELLULOSE SYNTHASE A* (*CesA*) genes in *Populus* (Kumar *et al.*, 2009). This convention entails using pre-existing names from *A. thaliana* to name the genes found in other plant species that match as probable orthologs with a suffix denoting the species name (*E. grandis*: Egr). Ascending letters were used to denote duplicates which are not present in *A. thaliana*.

Co-expression analysis and expression module detection

Population wide transcriptome data was obtained from developing xylem tissue of 156 individuals of a (*E.grandis* x *E.urophylla*) x *E.urophylla* (GU x U) F2 backcross family (Kullan *et al.*, 2012) previously employed in a systems genetics analysis (Mizrachi *et al.*, 2017; NCBI Sequence Read Archive accession no. SUB2087452). Co-expression analysis was performed

using Pearson correlation combined with joint likelihood scores to identify genes significantly correlated with xylan modification genes identified from PLAZA, that were equally significantly correlated in a reciprocal manner (Djidjev, 2008; Pinard *et al.*, 2019; **Methods S1**). Clustering was performed using a multi-level community detection algorithm to identify expression modules (EM). Using a custom enrichment R script (Pinard *et al.*, 2019; **Methods S1**), functional annotation was performed for genes co-expressed with xylan modification query genes by using the Gene Ontology (GO) Molecular Function, GO Biological Process, MapMan and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment for *E. grandis* (**Methods S1**).

***Trans*-eQTL analysis and hotspot detection**

Using eQTL data available for *Eucalyptus* GU x U (**Methods S2**), all *trans*-eQTLs for genes co-expressed with xylan modification were extracted. *Trans*-eQTL hotspot loci identification was performed on all co-expressed genes which have their expression affected by at least one *trans*-eQTL. By using sliding window bin sizes of 4-6 cM by calculating the genome-wide frequency of eQTLs and then normalising for local gene density (Christie *et al.*, 2017), loci statistically enriched in *trans*-eQTL were identified. The dataset was then separated into the previously identified EMs using custom R scripts, to dissect the genetic architecture by identifying the main loci affecting the expression of an EM. Using the aforementioned method for GO, KEGG and MapMan enrichment, enriched terms were identified for the genes which were present in an EM and had a *trans*-eQTL co-locating to a hotspot locus. The genes associated with enrichment terms and the terms themselves were manually curated to categories pertaining to transcriptional regulation, SCW associated processes and primary metabolic pathways. The EM membership was subsequently used to separate the genes and their associated processes into the previously identified EMs using custom R scripts.

Module eigenQTL construction

Using outputs of the co-expression analysis, the “moduleEigengenes” function of the Weighted Gene Co-expression Network Analysis (Langfelder & Horvath, 2008) R package was used to

obtain the first principle component of variation for the EMs (Jansen & Nap, 2001; Langfelder & Horvath, 2008; Christie *et al.*, 2017). These module eigengenes were then used as traits for QTL mapping, using Composite Interval Mapping in QTL Cartographer (Wang *et al.*, 2006).

Pathway analysis

Using the closest *Arabidopsis* BLAST hit for KEGG mapper (Kanehisa *et al.*, 2012), Reactome (Croft *et al.*, 2011), PlantCyc within Plant Metabolic network (Dreher, 2014) and AraCyc (Zhang *et al.*, 2005), metabolic pathway analysis was performed. MapMan (Thimm *et al.*, 2004) was also used in this analysis, but in this case a MapMan database for the *Eucalyptus* genome was queried using *Eucalyptus* gene IDs. This analysis identified the most enriched pathways that were further investigated and manually curated to display the most relevant pathways represented in the dataset in accordance with literature.

Results:

Identification of 13 putative xylan modification genes in *Eucalyptus*

We identified 72 genes from the four gene families (55 TBL genes, 2 RWA genes, 4 GUX genes and 11 DUF579 genes) obtained from the PLAZA 3.0 query of the *E. grandis* genome (Myburg *et al.*, 2014). To determine which of these genes correspond to the putative homologs of SCW xylan modification genes in *Eucalyptus*, we combined phylogenetic reconstruction with co-expression and network reconstruction. The overlapping genes from each of the experiments were regarded as putative xylan modification genes in *Eucalyptus*.

We aimed to identify putative homologs in *Eucalyptus grandis*, *Arabidopsis thaliana* and *Populus trichocarpa* using maximum likelihood phylogenetic analysis, expression profiling, synteny, and protein domain prediction (**Fig. S1, Fig. S2, Fig. S3, Fig. S4, Fig. S5, Table S1**). Twenty nine putative orthologs in *Eucalyptus* could be identified for experimentally verified xylan modification genes. However, due to several tandem duplicates of genes being present in the *Eucalyptus*, the supplementation of the phylogenetic analysis with expression data and domain structure allowed us to narrow down the number of genes to two RWA, three GUX, four DUF579 and ten TBL candidates (**Table S2**). These 19 genes were found to be the most probable homologs of *Arabidopsis* and *Populus* genes known to be involved in xylan modification, and will be referred to as phylogenetic candidates.

In addition to phylogenetic analysis, all initial 72 genes were subjected to network reconstruction and co-expression analysis to obtain co-expressed gene networks and clusters (Pinard *et al.*, 2019; **Methods S1**). This analysis should aid in the identification of probable homologs of known- and novel xylan modification genes co-expressed with other SCW related genes. For network construction, we used a correlation cut-off of 0.6 and an F-score of 3.8 (**Fig. S6**) to identify co-expressed genes in the developing xylem transcriptomes of 156 GUxU back-cross hybrid individuals (Pinard *et al.*, 2019; **Methods S1**). Of the 72 potential xylan modification genes, 24 query genes passed the threshold (**Table S3**). In total, 1,136 *E. grandis* genes were co-expressed with the 24 query genes that were clustered into five non-redundant expression modules (EM1 to EM5); (**Fig. S7, Table S4**). EM1 and EM5 were the largest modules, contained eight query genes each and were highly enriched for SCW GO biological process terms (**Tables S5, S6**). 15 query genes (one RWA, two GUX-, three DUF579- and nine TBL genes) were found to have their expression correlated (>0.6) with genes associated with the biosynthesis of the SCW biopolymers cellulose (Ranik & Myburg, 2006), lignin (Carocha *et al.*, 2015) and xylan (Myburg *et al.*, 2014) in *E. grandis* (**Tables S3, S7**).

An overlap of 13 genes was found between the phylogenetic candidates and SCW co-expressed query genes. These *Eucalyptus* genes are candidate SCW xylan modification genes (**Fig. 1**). SCW specific *EgrGUX1* and *EgrGUX2* were correlated with SCW associated genes, while *EgrGUX3-A* was not correlated with SCW genes. This is in line with what has been shown in *A. thaliana*, where *GUX1* and *GUX2* are SCW specific (Mortimer *et al.*, 2010) and *GUX3* is primary cell wall specific (Mortimer *et al.*, 2015). Of the two GXM genes present in *Eucalyptus* only *EgrGXM1* was co-expressed with SCW genes (**Table S7**), whereas *EgrGXM2* may have a function similar to *PtrDUF579-1* (Song *et al.*, 2014). It is therefore possible that *EgrGXM1* alone methylates GlucA residues as opposed to three GXM proteins in *Arabidopsis* (Urbanowicz *et al.*, 2012; Lee *et al.*, 2012).

In *Arabidopsis*: RWA1, RWA3 and RWA4 transport acetyl-CoA preferentially for SCW polysaccharide acetylation, whereas RWA2 preferentially transports acetyl-CoA for xyloglucan acetylation (Manabe *et al.*, 2013). In this study, we found that *EgrRWA1* is co-expressed with all SCW genes in the analysis, whereas *EgrRWA2* is correlated only with secondary cell wall *CeSA* genes; which suggest that *EgrRWA1* is the predominant RWA protein involved in SCW xylan modification, as opposed to several RWAs involved in *Arabidopsis* (Lee *et al.*, 2011; Manabe *et al.*, 2013). All clade IV *E. grandis* TBL genes (**Fig. S5**) that show strong stem preferential expression (**Fig. S3**) were correlated with SCW (**Table**

S7) with exception to the highly and stem preferentially expressed *EgrTBL36-A*; which was not among the co-expression query genes (**Table S3**). Even though *EgrTBL36-A* was not co-expressed with xylan modification or SCW biosynthesis genes, functional validation is needed to determine its biological function. Two additional SCW correlated TBL genes were identified: *EgrTBL25* (Eucgr.E00070) which is included in clade I, and *EgrTBL16* (Eucgr.C03303) which is found in clade IIa, as opposed to all other experimentally validated xylan O-acetyltransferases in clade IV (**Fig. S1**); (Xiong *et al.*, 2013; Yuan *et al.*, 2016a,b,c). This implies that *EgrTBL16* may be novel SCW xylan modification gene or it may be involved in the acetylation of SCW glucomannan like TBL25 (Gille *et al.*, 2011; Haghighat *et al.*, 2016; Roka *et al.*, 2018; Zhong *et al.*, 2018). However, further functional validation needs to be completed to validate the association of this TBL with the biosynthesis or modification of a SCW polysaccharide. Combining all of the evidence from phylogenetic and co-expression analysis, 13 genes were selected as putative xylan modification candidates (**Fig. 1**), as well as *EgrTBL16* and *EgrTBL25*, as they may aid in understanding transcriptional regulation of SCW associated acetyltransferases (**Table S3**).

Genetic architecture and systems genetics of xylan modification gene expression

To understand the genetic orchestration responsible for these five expression modules, QTL mapping was performed to identify loci where one or more closely linked polymorphisms affect the expression of genes in the co-expression modules. We queried all 1,136 genes for previously mapped *trans*-eQTLs (**Methods S2**), and found that 793 genes (69.81%) had at least one *trans*-eQTL affecting their expression and a total of 1,273 *trans*-eQTLs (1.6 per gene) were mapped for these genes (**Fig. S8**). It was further investigated whether any *trans*-eQTLs were statistically more likely to occur at specific loci in the genome than others (in the context of local gene density). By dividing the genome into 1 cM bins and using 4-6 cM sliding window intervals (Christie *et al.*, 2017) we were able to identify seven statistically enriched loci (**Fig. 2a, Fig. S9, Tables S8, S9**). These loci represent potential regulatory polymorphisms that affect many genes co-expressed with xylan modification, and are henceforth referred to as hotspot loci. 26 genome wide hotspot loci have been identified in the GUxU population (**Methods S2**), and the seven identified in this study were named according to the hotspots they overlapped, namely: HS_3.1, HS_3.3, HS_6.1, HS_9.1, HS_10.2, HS_10.3 and HS_10.4 (**Table S10**). These seven hotspots contain 666 *trans*-eQTLs (539 genes) which represent 53.42% of all

trans-eQTL loci (**Fig. 2b**). Of the seven loci, it was seen that five (HS_3.3, HS_9.1, HS_10.2, HS_10.3 and HS_10.4) affect the probable xylan modification genes, which shows that genes affected by these hotspots are coordinated together with xylan modification and SCW formation.

To understand how genetic variation affects the structure of the expression profiles observed in the five expression modules, module eigengene analysis was performed. We obtained a module eigengene for each EM, and each EM was found to separate from the others in a principle component analysis (**Fig. S10**). Subsequently, to identify loci where genetic variation is linked to the EMs, we mapped eQTL for each EM by using the representative module eigengene profile as a trait; eight eigen-eQTLs were detected in total, and genetic variation was found to contribute between 6-8% of trait variation (**Table 1, Fig. S11a**; Jansen & Nap, 2001; Christie *et al.*, 2017). Four of the seven hotspot loci, namely HS_3.3, HS_6.1, HS_9.1 and HS_10.3, were found to co-locate with the eigen-eQTLs (**Table 1, Fig. 2c, Fig. S11a, Fig. S11b**). The enrichment of these four hotspots over the other three hotspots (HS_3.1, HS_10.2 and HS_10.4) may potentially indicate that there is a hierarchy in the regulation of xylan modification and SCW associated genes.

To identify the cellular processes and pathways co-regulated with xylan modification genes, the 539 genes from the five EMs affected by the seven hotspot loci were extracted and the total set was analysed for enriched processes (**Table S9**). Functional enrichment analysis was performed on 531 genes using Gene Ontology (GO), Kyoto Encyclopaedia of Genes and Genomes (KEGG) and MapMan annotation (**Tables 2, S11**). The genes corresponding to all enriched terms were extracted and manually curated into the eight most enriched processes associated with xylan modification namely: three SCW biopolymer terms (cellulose, xylan and lignin), four metabolic terms (sugars, acetyl-CoA, SAM and phenylalanine) and transcriptional regulation (**Tables 2, S12**). Five gene co-expression modules (EMs), seven *trans*-eQTL hotspots which affect gene expression in these EMs, and eight biological processes were collated into a three tier systems genetics model for xylan modification (**Fig. 2d**). Two hotspot loci; HS_9.1 and HS_10.3 represent loci which contain master regulators of SCW as they affect the most genes (187 and 227 respectively), and co-locate with eigen eQTLs for EM1 and EM5 which are highly enriched in SCW associated terms. HS_9.1 was found to prominently affect TF genes regulating SCW formation along with several genes related to cellulose, xylan and lignin biosynthesis. HS_10.3 was shown to affect SCW biosynthesis in general, including the majority of lignin and xylan biosynthesis genes, as well as many cellulose biosynthetic genes.

HS_9.1 influences the expression of more genes in EM1, whereas HS_10.3 influences more genes in EM5 (**Fig. 2c, Fig. S12a, Fig. S12b**).

SCW biosynthetic genes and their associated metabolic pathways are transcriptionally co-regulated

It was noted that EM5 contained genes associated with the biosynthesis of cellulose and the sugar fraction of xylan (backbone, reducing end and GlucA modifications); (**Fig. S13, Table S5**), the expression of which was primarily affected by HS_9.1 and HS_10.3 (**Fig. S13, Tables S13, S14**). Genes related to the interconversion of nucleotide sugars were also found primarily in EM5 (**Table S5, Fig. 2d**), but have their expression affected by HS_10.3 and HS_10.4 (**Fig. S13**). The presence of the *UDP-GLUCOSE 6-DEHYDROGENASE* (*UGD*: Eucgr.J01372) in EM5 further supports the findings that UDP-GlucA is derived from UDP-glucose (UDP-Gluc) rather than glucose-6-phosphate (Sharples & Fry, 2007). Genes involved in the phosphorylated pathway of serine biosynthesis (Ho *et al.*, 1999; Ho & Saito, 2001) were present in EM5 (**Fig. 3**), as well as four *METHIONINE SYNTHASE* (*MS*) genes (Ravanel *et al.*, 2004), and two *S-ADENOSYLHOMOCYSTEINE HYDROLASE1* (*SAHH1*) genes (**Table S4**; Ouyang *et al.*, 2012).

Genes involved in phenylalanine biosynthesis were found primarily in EM1 as well as the majority of genes involved the phenylpropanoid biosynthesis (**Table S5, Fig. 3**). Phenylalanine is the precursor to phenylpropanoid biosynthesis (Maeda *et al.*, 2011), which produces among other things, monolignols for lignin polymerisation. The co-regulation of these two pathways is consistent with previous results (**Fig. 3, Tables S5, S13, S15**; Mizrachi *et al.*, 2013; Pinard *et al.*, 2019), where both the shikimate and phenylpropanoid pathway genes were affected by HS_9.1, HS_10.2, HS_10.3 and HS_10.4, but the shikimate pathway genes *CHORISMATE SYNTHASE* and *CHORISMATE MUTASE* are also affected by HS_3.3 (**Fig. 3**). Genes related to the glycolate pathway responsible for the formation of serine (Douce *et al.*, 2001; Voll *et al.*, 2006) were found in EM1, along with one *MS* gene and two *SAM SYNTHASE* (*SAMS*) genes (**Fig. 3, Tables S13, S15**) and affected by HS_9.1 and HS_10.3.

With respect to xylan modification, there is also a division in the EM membership of the genes; specifically *EgrGUX1*, *EgrGUX2*, *EgrTBL29*, *EgrTBL3-B* and *EgrRWA1* were found in EM5, whereas *EgrTBL34-A*, *EgrTBL33-A*, *EgrGXM1*, *EgrIRX15* and *EgrIRX15L* were in EM1 (**Table S2**). All of the genes encoding proteins involved in the modification of the O-2 position

required for cellulose interaction (Pereira *et al.*, 2017) were included in EM5. In contrast, genes involved in SAM dependant methylation of GlucA, and the formation of the 3-O acetyl, α 1,2 (Me)GlucA structural fragment, were included in EM1 (**Fig. 4**). EM1 and EM5 also contained genes (e.g. *EgrTBL29* and *EgrTBL3-B* in EM5, and *EgrTBL34* in EM1) involved in both 3-O acetylation and 2-O; 3-O diacetylation (**Fig. 4**; Zhong *et al.*, 2017). The functions of these two modifications are currently unclear, but they have been proposed to be necessary for stabilising the xylan-cellulose interaction (Pereira *et al.*, 2017). The functions of IRX15 and IRX15L are still unclear, but double knockout mutants had SCWs that were corrugated in appearance and segments of the SCW appeared to be breaking away (Brown *et al.*, 2011; Jensen *et al.*, 2011) suggesting that the proteins are required for the integrity of the SCW. The presence of *EgrIRX15* in EM1 may additionally hint at a role in either lignin's interaction with xylan, or another late-stage SCW process.

Four acetyl-CoA producing pathways co-regulated with xylan modification genes

The metabolic source of cytosolic acetyl-CoA for xylan acetylation (Pauly & Scheller, 2000; Urbanowicz *et al.*, 2014) is currently unknown, but could originate from several sources (Oliver *et al.*, 2009). Possible sources include (i) pyruvate derived from glycolysis, (ii) acetate obtained from fermentation products acetaldehyde and ethanol, (iii) peroxisomal β -oxidation of fatty acids (FA) or through the degradation of the amino acids, (iv) leucine, isoleucine, valine, and (v) possibly lysine (Fatland *et al.*, 2002; Mentzen *et al.*, 2008; Hildebrandt *et al.*, 2015; Wierzbicki *et al.*, 2019). Cytosolic acetyl-CoA is derived from citrate from the tricarboxylic acid (TCA) pathway, to which (i) pyruvate, (iv) leucine, valine and (v) lysine contribute carbon skeletons. Acetyl-CoA produced in the peroxisomes can either exit the peroxisomes and enter the TCA cycle as citrate, or enter the glyoxylate pathway and leading to incorporation into TCA cycle through succinate or oxaloacetate. Possible sources of peroxisomal acetyl-CoA include (ii) acetate, (iii) β -oxidation products, and (iv) isoleucine.

Genes involved in (i) glycolysis were co-regulated with xylan modification, however in plants, glycolysis can occur in both the cytosol and plastid (Ravanel *et al.*, 2004). Cytosolic glycolysis genes included *PHOSPHOFRUCTOKINASE* (*PFK*), *ALDOLASE* and *PHOSPHOGLYCEROMUTASE*, whereas the plastidial glycolysis genes were *TRIOSE PHOSPHATE ISOMERASE*, *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* and *PHOSPHOGLYCERATE KINASE* (**Fig. 5**, **Tables S13**, **S16**). Neither the cytosolic nor

plastidial copies of *ENOLASE* or *PYRUVATE KINASE* were present in any EMs, nor affected by any hotspot loci. These genes were found to be highly and constitutively expressed so regulation of these proteins may be on a post-transcriptional level (**Fig. 5**; Guo *et al.*, 2003; Zhang & Liu, 2013; Vining *et al.*, 2015). Cytosolic *PHOSPHOENOLPYRUVATE CARBOXYLASE (PEPC) 4* expression is effected by HS_3.1; the enzyme encoded by this gene produces oxaloacetate from phosphoenolpyruvate (PEP), which may enter the TCA pathway (**Fig. 5**). An almost complete glycolytic pathway was identified, however, this pathway contains genes that code for cytosolic glycolysis enzymes as well as those that code for plastidial glycolysis enzymes. It is not clear whether these two compartmentally separated pathways contribute to the same end point metabolites. The former predominantly provides carbon skeletons for the TCA cycle, while the latter contributes to amino acid and fatty acid biosynthesis, but recent studies have pointed out that cytosolic glycolysis may also contribute to the latter processes (Schwender *et al.*, 2003, 2006; Andriotis *et al.*, 2010; Prabhakar *et al.*, 2010).

Of the genes encoding TCA pathway enzymes; *PYRUVATE DEHYDROGENASE (PDH)*, *SUCCINATE DEHYDROGENASE (SDH)* and *ADENOSINE 5'-TRIPHOSPHATASE (ATPase)* were present in EM2 (**Fig. 5, Table S16**). Two potential entry points into the TCA cycle are co-regulated with xylan modification genes; either through acetyl-CoA by the action of *PDH* (Ravanel *et al.*, 2004) or through cytosolic oxaloacetate by *PEPC4*. Both these routes require *CITRATE SYNTHASE (CSY)*, which was not co-regulated with xylan modification in our study (not in any EMs or affected by any of the hotspot loci), this may be due to *CSY* being part of a “metabolon” required for channelling of TCA intermediates (Zhang *et al.*, 2017). *PDH* is present in EM2 and its expression is affected by HS_6.1, this suggests that pyruvate may be a likely route that carbon follows to enter the TCA cycle during SCW formation. The fate of mitochondrial citrate involves either progression through TCA cycle, or exiting the mitochondrion in exchange for a cytosolic dicarboxylate or tricarboxylate facilitated by the *DICARBOXYLATE-TRICARBOXYLATE CARRIER (DTC)*; (Picault *et al.*, 2004). *PEPC4* derived oxaloacetate may act as the dicarboxylate used in exchange for citrate (Hanning *et al.*, 1999; Picault *et al.*, 2002), and may perhaps be the reason why *PEPC4* and *PEPC KINASE (PEPCK) 1* have their expression affected by hotspots HS_3.1 and HS_3.3 respectively (**Tables S13, S16**). The support for this hypothesis is less conclusive when compared to *PDH*, and needs to be further validated. Citrate transported into the cytosol can contribute to the synthesis of glutamate and acetyl-CoA (Fatland *et al.*, 2005). Genes involved in glutamate

synthesis (*ISOCITRATE DEHYDROGENASE (ICDH)* and *GLUTAMATE SYNTHASE (GS)*) and acetyl-CoA formation (*ATP CITRATE LYASE*); (*ACL*) were included in EM5 (**Fig. 5**). Three *ACL* genes were included in the co-expression or co-regulation data (**Table S16**), with two being present in EM5 (**Fig. 5, Table S16**) and one being affected by both HS_6.1 and HS_10.3 (**Fig. 5, Table S13**). The transcriptional coordination of these three genes suggest that they are most likely responsible for producing the cytosolic acetyl-CoA pool which is used for xylan acetylation (Faria-Blanc *et al.*, 2018).

Genes associated with (ii) ethanol degradation, (iii) β -oxidation, and (iv) the catabolism of branched chain amino acid (BCAA) leucine and isoleucine (Mentzen *et al.*, 2008) were found within EM1 and effected primarily by HS_6.1, HS_9.1 and HS_10.3 (**Fig. 5, Table S16**). *ISOCITRATE LYASE (ICL)* of the glyoxylate pathway was affected by HS_6.1 and HS_9.1, which may be an indication that peroxisomal citrate enters into the glyoxylate pathway, is converted to succinate and incorporated into the TCA cycle (Eastmond & Graham, 2001). Valine and isoleucine degradation (iv) results in the formation propionyl-CoA which can be converted to acetyl-CoA, but no genes involved in this process were co-regulated with xylan modification (Hildebrandt *et al.*, 2015). Lysine degradation is not fully understood, but only one gene from the proposed degradation pathway (Hildebrandt *et al.*, 2015), *LYSINE-KETOGLUTARATE REDUCTASE/SACCHAROPINE DEHYDROGENASE* was found in EM1, and is most probably involved in glutamate production from oxoglutarate (Arruda *et al.*, 2000). From this analysis we cannot definitely implicate a source of acetyl-CoA for xylan acetylation, but (iv) valine and (v) lysine can largely be eliminated as possible sources.

Discussion

Eucalyptus grandis has 13 candidate SCW xylan modification genes

By using a combined approach of phylogenetics, expression profiling and population wide co-expression, we were able to identify putative candidates associated with SCW xylan modification (**Table S3**). One RWA (**Fig. S1**), one GXM (**Fig. S3**), two GUX (**Fig. S2**), and eight predicted TBL proteins along with the AXY9 singleton gene in the *E. grandis* genome (**Fig. S4, Fig. S5**) were found to be associated with the process. In contrast to *Arabidopsis* and *Populus*, *Eucalyptus* has mostly single genes associated with xylan modification (single SCW associated RWA and GXM), likely due to the lack of a recent genome duplication (Myburg *et al.*, 2014). This might indicate that a core set of xylan modification genes are responsible for

modifying the backbone polymer, and duplicates observed in model species may act redundantly or contribute to species specific differences in xylan function. Importantly, fewer genes functionally involved in xylan modification in *Eucalyptus* may be beneficial for genetic modification approaches to alter wood traits, as it would simplify CRISPR guide RNA and siRNA design for use in knock-out and downregulation experiments. In conclusion, we were able to identify 13 genes putatively associated with SCW xylan modification in *Eucalyptus*.

Xylan modification is transcriptionally co-regulated with major SCW regulators and metabolic pathways

By using population transcriptomics data, we were able to identify the processes that are transcriptionally coordinated with xylan modification processes and to unravel possible metabolic dependencies necessary for this process (**Fig. 2d, Fig. 3, Fig. 5, Fig. S7, Fig. S13**). It is important to note that, in the absence of actual metabolomics data, our inference is limited to metabolic steps that are encoded by genes that share genetic architecture (co-regulation and co-expression) with xylan modification genes. Nevertheless, we can begin to infer that these metabolic pathways are co-regulated with xylan modification due to the fact that SCW formation is such a strong metabolic sink and that it is crucial to coordinate precursor supply and allocation to avoid disrupting other cellular processes using the same precursors. In the remainder of this article, our metabolic inferences are based on transcriptional evidence with the above stated limitations implied. We do highlight cases where our metabolic inferences are supported by published phenotypes and observations.

Systems genetics approaches have yielded valuable results before in a range of organisms, from fruit flies to humans, *Arabidopsis*, maize and forest trees (Atwell *et al.*, 2010; McDermott-Roe *et al.*, 2011; Stančáková *et al.*, 2012 ; Ayroles *et al.*, 2009; Drost *et al.*, 2010; Christie *et al.*, 2017; Mizrachi *et al.*, 2017). Our study has similarly yielded novel results revealing that xylan modification processes and their associated, interdependent metabolic pathways are transcriptionally co-regulated at population level. Furthermore, the systems genetic analysis identified candidate regulatory loci harbouring drivers of this transcriptional coordination. In addition to providing a much deeper understanding of the transcriptional regulation of xylan modification and its associated metabolic precursor pathways, the systems genetics model can

be used as a tool to design and test new biotechnology approaches to alter cell wall structure and chemistry (examples provided in **Table S17**).

The early phases of SCW formation, namely the biosynthesis of cellulose, and the pathways for xylan backbone and reducing end synthesis are represented in EM5 (**Fig. 2d, Fig. S13**); (Hertzberg *et al.*, 2001). Included in this module are genes for glucuronic acid modification as well as the interconversion of nucleotide sugars (**Fig. 2d, Fig. S13**). Additionally, it was found that in this module, the expression of sugar interconversion associated genes was affected by combinations of HS_10.3 and HS_10.4, whereas HS_9.1 and HS_10.3 combinations affected mostly xylan biosynthesis genes. Nucleotide sugar interconversion pathways, including sucrose breakdown, provide precursors for cellulose and xylan biosynthesis, which are co-regulated in EM5 (**Fig. 6, Fig. S13**). Of the xylan modification genes, *EgrRWAl*, *EgrTBL29*, *EgrTBL3-B*, *EgrGUX1* and *EgrGUX2* are present in EM5 (**Fig. 4, Fig. 6**). GUX1, GUX2 and TBL29 modify the O-2 positions on the xylan backbone (Mortimer *et al.*, 2010; Yuan *et al.*, 2013; Urbanowicz *et al.*, 2014), these genes are clustered together and likely compete with each other for O-2 substrate space (Lee *et al.*, 2014); (**Fig. 4**). *EgrRWAl* is present in EM5 as acetyl-CoA needs to be transported into the Golgi to supply precursor for TBL29 to perform its function (Manabe *et al.*, 2013; Urbanowicz *et al.*, 2014); (**Fig. 4**). The transporter of UDP-GlucA into the Golgi has currently not been identified, but we can assume that it will be present in EM5 similar to the UXT1 (**Fig. S13**) homolog which transports UDP-xylose into the Golgi (Ebert *et al.*, 2015). From work in *Arabidopsis*, it is known that TBL3 adds acetyl to the O-3 position on xylose residues (Yuan *et al.*, 2016b), and these O-3 modifications have been suggested to stabilise the cellulose-xylan interaction (Pereira *et al.*, 2017), so this may be a reason for *EgrTBL3-B* to be located in EM5 (**Fig. 4**). Methionine biosynthesis from 3-PGA is also present in EM5 (Ho *et al.*, 1999; Ho & Saito, 2001), which may be an indication that the differentiating xylem requires methionine for protein synthesis, or may be accumulating for SAM biosynthesis for monolignol methylation (Do *et al.*, 2007). From the above observations, we posit that EM5 represents early SCW formation, from polysaccharide biopolymer substrate supply to the coordination of cellulose deposition with xylan deposition, and may ultrastructurally impact cellulose content and microfibril angle (Reis & Vian, 2004; Derba-Maceluch, 2015; Yang *et al.*, 2017).

The later stages of SCW formation involving monolignol biosynthesis, lignin polymerisation and programmed cell death (PCD) are present in EM1 (**Fig. 2d, Table S4**); (Hertzberg *et al.*, 2001; Doukhanina *et al.*, 2006; Escamez *et al.*, 2016). This module contains genes for

phenylalanine biosynthesis (Maeda *et al.*, 2011), and the subsequent phenylpropanoid pathway that produces monolignols required for polymeric lignin formation (Carocha *et al.*, 2015); (**Fig. 3**). The HS_9.1 and HS_10.3 combination was also prevalent in the biosynthesis of phenylalanine, SAM, phenylpropanoids and methylation of GlucA. The phenylpropanoid pathway is also involved in the methylation of H-monolignols to produce either G- or S-monolignols, a process that requires SAM as a precursor (Do *et al.*, 2007). The SAM precursor is potentially obtained from the methionine pool produced in EM5, which may be an example of a metabolite that accumulates in early cell wall and is utilised for SAM biosynthesis later (Gallardo *et al.*, 2002; Ohtani *et al.*, 2016). This methionine pool is also maintained in EM1, as the methyl donor folate pathway is present (**Fig. 3**), using glycine as a source (Douce *et al.*, 2001; Voll *et al.*, 2006; Vanholme *et al.*, 2012). The majority of SAM produced needs to be balanced between lignin and the methylation of the GlucA residues on the xylan backbone, an association that has been identified in *gx*m knockout mutants (Urbanowicz *et al.*, 2012). EgrGXM1 shares SAM with monolignol methylation which is why they are co-regulated in EM1 (**Fig. 3, Fig. 6**). Additionally, a single *GXM* gene implies that in *Eucalyptus*, expression and enzyme kinetics determine GlucA methylation rather than gene number (Yuan *et al.*, 2014a). *EgrTBL33-A* and *EgrTBL34-A* are other xylan modification genes present in EM1, but have not been fully characterised in the context of lignin biosynthesis and polymerisation (Yuan *et al.*, 2016a,c) as suggested by our data. From Yuan *et al.* 2016, it has been implied that TBL33 adds acetyl to the O-3 position, which already have a (Me)GlucA at the O-2 position (Yuan *et al.*, 2016a; **Fig. 4**). EgrTBL33-A and EgrGXM1 are both required to create the O-2 (Me)GlucA, O-3 acetyl modification, which together with EgrTBL34-A may be creating the hydrophobic pocket required for lignin polymerisation (Yuan *et al.*, 2016c; Johnson *et al.*, 2017; **Fig. 4**). An additional reason for their co-regulation may be due to the fact that acetyl-CoA needs to be balanced between both xylan acetylation and maintaining SAM levels. EM1 thus represents the late stages of SCW formation, highlighting the metabolic coordination of substrate utilisation and structural requirements for monolignol production and polymerisation into lignin as well as a potential association between xylan and lignin.

The source of Acetyl-CoA may vary between different stages of SCW deposition

The pathways providing acetyl-CoA for xylan acetylation during SCW formation are currently unclear. Using our model it was found that acetyl-CoA is potentially produced from four

different sources that are transcriptionally coordinated with xylan modification (**Fig. 5**). Unlike the sugar and SAM precursor pathways (**Fig. 3, Fig. S13**), the acetyl-CoA producing pathways were not completely co-regulated with xylan modification (**Fig. 5**). It was found that glycolysis starting from F-6P, as well as the conversion of citrate into oxaloacetate and acetyl-CoA were present in EM5, and affected by HS_10.3 (**Fig. 5**). F-6P is utilised in the oxidative pentose phosphate pathway (OPPP) and the cycling of triose phosphates (**Fig. 3, Fig. 5**), which are used for serine and phenylalanine biosynthesis (Alonso *et al.*, 2010). However, our results suggest that F-6P may also be used for acetyl-CoA production. The F-6P is probably produced by the action of FRUCTOKINASE (FRK) that is also present in EM5 (**Fig. S13, Tables S9, S14**), which may be why plants deficient in FRK activity, display decreases in TCA intermediates and SCW defects (Stein *et al.*, 2016a,b). This observation may also explain why F-6P enters glycolysis, despite glycerol being available as a carbon source (Sharples & Fry, 2007). Genes from cytosolic and plastidial glycolysis, as well as the TCA cycle were present primarily in EM2, which makes it difficult to conclusively implicate an individual precursor pathway in the production of acetyl-CoA for xylan acetylation (**Fig. 5**). One possibility is that TCA cycle needs to be balanced between energy production and citrate for acetyl-CoA production, which is perhaps why we do not see more glycolysis genes clustering in EM5. We hypothesise that cytosolic glycolysis is the likely pathway that F-6P enters to produce citrate for xylan acetylation.

It was also noted that genes involved in three other source pathways for acetyl-CoA production, namely ethanol degradation, BCAA breakdown and β -oxidation of FA, were present in EM1 (**Fig. 5**). Upregulation of BCAA- and lipid degradation during later stages of SCW has been reported (Li *et al.*, 2016), and is congruent with the presence of genes involved in these two processes in EM1. The carbon in glycolysis is required for phenylalanine and phenylpropanoid biosynthesis in late SCW formation (Ohtani *et al.*, 2016), therefore the TCA cycle may need to be supplemented by alternative pathways, such as those seen in EM1. The reduced flow of carbon from glycolysis into the TCA pathway may result in fermentation occurring to sustain energy production (Zabalza *et al.*, 2009; van Dongen *et al.*, 2011), while the ethanol produced is converted to acetyl-CoA through acetate metabolism (Turner *et al.*, 2005). Although these pathways are present in EM1, it does not conclusively mean that they contribute to acetyl-CoA production. The lipid processing genes may be associated with suberin biosynthesis (Gou *et al.*, 2009), pro-auxins require a single cycle of β -oxidation to be converted into an active form (Zolman *et al.*, 2000) and *ALCOHOL DEHYDROGENASE* may be required for acetaldehyde

detoxification (Drew, 1997; Kürsteiner *et al.*, 2003). However, the regulation of FA specific LACS6 (Fulda *et al.*, 2002) by SCW gene regulating HS_10.3 implies that FA metabolism might be involved in SCW (**Fig. 5**). These alternative pathways may be required for the TCA cycle to remain active and produce citrate for xylan modification while carbon flux is redirected to phenylalanine during late SCW formation.

Although both *ACL* genes are present in EM5, it is likely that citrate is broken down to acetyl-CoA during all of SCW formation (Li *et al.*, 2016), meaning that the citrate pool is vital for xylan acetylation (Faria-Blanc *et al.*, 2018). The presence of different acetyl-CoA producing pathways in EM5 and EM1 may indicate that the acetyl-CoA for xylan acetylation differs in origin between early and late SCW formation (**Fig. 6**). The presence of *PFK* along with acetyl-CoA producing *ACL* and the phosphorylated pathway of serine in EM5 may indicate that these pathways are, in part, coordinated together to partition acetyl-CoA between both xylan acetylation and SAM biosynthesis. This hypothesis needs to be further examined.

This study has shed light on the genetic and transcriptional regulation of xylan modification, as well as other SCW and associated metabolic processes (**Fig. 6**). Given that xylan modification patterns have an effect on SCW properties, altering the patterns may yield beneficial SCW traits for industrial applications such as biorefinery or saccharification and fermentation. We propose that altering the patterns may be accomplished by re-wiring the transcriptional circuitry of early xylan modification genes (e.g. by expressing them under the control of later stage promoters, and *vice versa*). Alternatively, we could employ metabolic engineering strategies to redirect metabolic flux of the xylan modification precursors to other biopolymers. Altering the flux of precursors has the potential to affect several other processes which rely on the same precursor pools as xylan modification. Our systems model could be used to identify such co-regulated processes and potentially predict whether any deleterious effects may arise. We posit that metabolic engineering may also be a strategy to supply sufficient precursor for the ectopic addition of monocot xylan modifications such as α 1,2 or α 1,3-linked arabinose and β 1,2-xylose to dicot xylans. The availability of a monocot systems model for xylan modification will be instrumental in identifying metabolic steps to target to achieve such ectopic modifications in a woody dicot. The increased understanding emerging from this study opens the possibility of generating a seemingly unlimited variety of xylan modification patterns. The concept of “designer biomass” was the topic of a recent review (Smith *et al.*, 2017). We propose that by understanding the specific effects each pattern has on SCW composition, patterns can essentially be “tailor-made” to produce SCW compositions or

“designer biomass” which is beneficial for each industrial processing application’s requirements.

Acknowledgements

The authors acknowledge K van der Merwe (University of Pretoria) for bioinformatics support and C Balkwill for developing the phylogenetic analysis workflow used in this study. This work was funded in part by the National Research Foundation (NRF) of South Africa - Bioinformatics and Functional Genomics Programme (BFG Grant UID 86936 and 97911), the Technology and Human Resources for Industry Programme (THRIP Grant UID 80118 and 96413), the Department of Science and Technology (DST, Strategic Grant for the *Eucalyptus* Genomics Platform) and by Sappi Forest Research through the Forest Molecular Genetics (FMG) Programme at the University of Pretoria. MPW, NC and DP acknowledge postgraduate and postdoctoral scholarship support from the NRF.

References

- Alonso AP, Piasecki RJ, Wang Y, LaClair RW, Shachar-Hill Y. 2010.** Quantifying the labeling and the levels of plant cell wall precursors using ion chromatography tandem mass spectrometry. *Plant Physiology* **153**: 915–924.
- Andriotis VME, Kruger NJ, Pike MJ, Smith AM. 2010.** Plastidial glycolysis in developing *Arabidopsis* embryos. *New Phytologist* **185**: 649–662.
- Arruda P, Kemper EL, Papes F, Leite A. 2000.** Regulation of lysine catabolism in higher plants. *Trends in Plant Science* **5**: 324–330.
- Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, *et al.* 2010.** Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**: 627.
- Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, *et al.* 2009.** Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics* **41**: 299–307.
- Bar-Peled M, O'Neill MA. 2011.** Plant nucleotide sugar formation, interconversion, and salvage by sugar recycling. *Annual Review of Plant Biology* **62**: 127–155.
- Biely P, MacKenzie CR, Puls J, Schneider H. 1986.** Cooperativity of esterases and xylanases in the enzymatic degradation of acetyl xylan. *Nature Biotechnology* **4**: 731–733.
- Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, Gerrits A, Bystrykh L V, de Haan G, Su AI, *et al.* 2008.** Genetical genomics: spotlight on QTL hotspots. *PLoS Genetics* **4**: e1000232.
- Bromley JR, Busse-Wicher M, Tryfona T, Mortimer JC, Zhang Z, Brown DM, Dupree P. 2013.** GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxytan with different substitution patterns. *The Plant Journal* **74**: 423–434.
- Brown DM, Goubet F, Wong VW, Goodacre R, Stephens E, Dupree P, Turner SR. 2007.** Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *The Plant Journal* **52**: 1154–1168.
- Brown D, Wightman R, Zhang Z, Gomez LD, Atanassov I, Bukowski J-P, Tryfona T,**

- McQueen-Mason SJ, Dupree P, Turner S. 2011.** *Arabidopsis* genes *IRREGULAR XYLEM (IRX15)* and *IRX15L* encode DUF579-containing proteins that are essential for normal xylan deposition in the secondary cell wall. *The Plant Journal* **66**: 401–413.
- Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, Bolam DN, Skaf MS, Dupree P. 2014.** The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of *Arabidopsis thaliana*. *The Plant Journal* **79**: 492–506.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009.** trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
- Carocha V, Soler M, Hefer C, Cassan-Wang H, Fevereiro P, Myburg AA, Paiva JAP, Grima-Pettenati J. 2015.** Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytologist* **206**: 1297–1313.
- Chen H-C, Song J, Wang JP, Lin Y-C, Ducoste J, Shuford CM, Liu J, Li Q, Shi R, Nepomuceno A, et al. 2014.** Systems biology of lignin biosynthesis in *Populus trichocarpa*: heteromeric 4-COUMARIC ACID:COENZYME A LIGASE protein complex formation, regulation, and numerical modeling. *The Plant Cell* **26**: 876–893.
- Christie N, Myburg AA, Joubert F, Murray SL, Carstens M, Lin Y-C, Meyer J, Crampton BG, Christensen SA, Ntuli JF, et al. 2017.** Systems genetics reveals a transcriptional network associated with susceptibility in the maize-grey leaf spot pathosystem. *The Plant Journal* **89**: 746–763.
- Civelek M, Lusi AJ. 2014.** Systems genetics approaches to understand complex traits. *Nature Review Genetics* **15**: 34–48.
- Croft D, O’Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B, et al. 2011.** Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Research* **39**: 691–697.
- Davin N, Edger PP, Hefer CA, Mizrachi E, Schuetz M, Smets E, Myburg AA, Douglas CJ, Schranz ME, Lens F. 2016.** Functional network analysis of genes differentially expressed during xylogenesis in *soc1*ful woody *Arabidopsis* plants. *The Plant Journal* **86**: 376–390.

- Derba-Maceluch M, Awano T, Takahashi J, Lucenius J, Ratke C, Kontro I, Busse-Wicher M, Kosik O, Tanaka R, Winz  l A, Kallas    et al. 2015.** Suppression of xylan endotransglycosylase *PtxtXyn10A* affects cellulose microfibril angle in secondary wall in aspen wood. *New Phytologist* **205**: 666–681.
- Djidjev HN. 2008.** A scalable multilevel algorithm for graph clustering and community structure detection - Algorithms and models for the web-graph. Aiello W, Broder A, Janssen J, Milios E, eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 117–128.
- Do C-T, Pollet B, Th  venin J, Sibout R, Denoue D, Barri  re Y, Lapierre C, Jouanin L. 2007.** Both *CAFFEOYL COENZYME A 3-O-METHYLTRANSFERASE 1* and *CAFFEIC ACID O-METHYLTRANSFERASE 1* are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta* **226**: 1117–1129.
- van Dongen JT, Gupta KJ, Ram  rez-Aguilar SJ, Ara  jo WL, Nunes-Nesi A, Fernie AR. 2011.** Regulation of respiration in plants: A role for alternative metabolic pathways. *Journal of Plant Physiology* **168**: 1434–1443.
- Douce R, Bourguignon J, Neuburger M, R  beill   F. 2001.** The *GLYCINE DECARBOXYLASE* system: a fascinating complex. *Trends in Plant Science* **6**: 167–176.
- Doukhanina E V, Chen S, van der Zalm E, Godzik A, Reed J, Dickman MB. 2006.** Identification and functional characterization of the BAG protein family in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **281**: 18793–18801.
- Dreher K. 2014.** Putting the plant metabolic network pathway databases to work: going offline to gain new capabilities. *Plant Metabolism: Methods and Protocols*. In: Sriram G. Humana Press, **1083**: 151–171.
- Drew MC. 1997.** Oxygen deficiency and root metabolism: injury and acclimation under hypoxia and anoxia. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**: 223–250.
- Eastmond PJ, Graham IA. 2001.** Re-examining the role of the glyoxylate cycle in oilseeds. *Trends in Plant Science* **6**: 72–78.
- Ebert B, Rautengarten C, Guo X, Xiong G, Stonebloom S, Smith-Moritz AM, Herter T, Chan LJG, Adams PD, Petzold CJ, et al. 2015.** Identification and characterization of a Golgi-localized *UDP-XYLOSE TRANSPORTER* family from *Arabidopsis*. *The Plant Cell* **27**: 1218–1227.

- Escamez S, André D, Zhang B, Bollhöner B, Pesquet E, Tuominen H. 2016.** METACASPASE9 modulates autophagy to confine cell death to the target cells during *Arabidopsis* vascular xylem differentiation. *Biology Open* **5**: 122–129.
- Faria-Blanc N, Mortimer J, Dupree P. 2018.** A transcriptomic analysis of xylan mutants does not support the existence of a secondary cell wall integrity system in *Arabidopsis*. *Frontiers in Plant Science* **9**: 384.
- Fatland BL, Ke J, Anderson MD, Mentzen WI, Cui LW, Allred CC, Johnston JL, Nikolau BJ, Wurtele ES. 2002.** Molecular characterization of a heteromeric ATP-CITRATE LYASE that generates cytosolic acetyl-Coenzyme A in *Arabidopsis*. *Plant Physiology* **130**: 740–756.
- Fatland BL, Nikolau BJ, Wurtele ES. 2005.** Reverse genetic characterization of cytosolic acetyl-CoA generation by ATP-CITRATE LYASE in *Arabidopsis*. *The Plant Cell* **17**: 182–203.
- Fulda M, Shockey J, Werber M, Wolter FP, Heinz E. 2002.** Two LONG-CHAIN ACYL-COA SYNTHETASES from *Arabidopsis thaliana* involved in peroxisomal fatty acid β -oxidation. *The Plant Journal* **32**: 93–103.
- Gallardo K, Job C, Groot SPC, Puype M, Demol H, Vandekerckhove J, Job D. 2002.** Importance of methionine biosynthesis for *Arabidopsis* seed germination and seedling growth. *Physiologia Plantarum* **116**: 238–247.
- Gille S, Cheng K, Skinner ME, Liepman AH, Wilkerson CG, Pauly M. 2011.** Deep sequencing of voodoo lily (*Amorphophallus konjac*): an approach to identify relevant genes involved in the synthesis of the hemicellulose glucomannan. *Planta* **234**: 515–526.
- 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**: 1178–1186.
- Gou J-Y, Yu X-H, Liu C-J. 2009.** A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 18855–18860.
- Guerriero G, Sergeant K, Hausman J-F. 2013.** Integrated -Omics: a powerful approach to understanding the heterogeneous lignification of fibre crops. *International Journal of*

Molecular Sciences **14**: 10958–10978.

Guo-Qing T, C. HS, Ralph D, C. HS. 2003. A novel C-terminal proteolytic processing of cytosolic PYRUVATE KINASE, its phosphorylation and degradation by the proteasome in developing soybean seeds. *The Plant Journal* **34**: 77–93.

Haghighat M, Teng Q, Zhong R, Ye Z-H. 2016. Evolutionary conservation of xylan biosynthetic genes in *Selaginella moellendorffii* and *Physcomitrella patens*. *Plant and Cell Physiology* **57**: 1707–1719.

Hanning I, Baumgarten K, Schott K, Heldt HW. 1999. Oxaloacetate transport into plant mitochondria. *Plant Physiology* **119**: 1025–1032.

Hao Z, Avci U, Tan L, Zhu X, Glushka J, Pattathil S, Eberhard S, Sholes T, Rothstein GE, Lukowitz W, et al. 2014. Loss of *Arabidopsis* *GAUT12/IRX8* causes anther indehiscence and leads to reduced G lignin associated with altered matrix polysaccharide deposition. *Frontiers in Plant Science* **5**: 357.

Hao Z, Mohnen D. 2014. A review of xylan and lignin biosynthesis: foundation for studying *Arabidopsis* *irregular xylem* mutants with pleiotropic phenotypes. *Critical Reviews in Biochemistry and Molecular Biology* **49**: 212–241.

Hefer CA, Mizrahi E, Myburg AA, Douglas CJ, Mansfield SD. 2015. Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: *Populus trichocarpa* and *Eucalyptus grandis*. *New Phytologist* **206**: 1391–1405.

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlén M, Teeri TT, Lundeberg J, et al. 2001. A transcriptional roadmap to wood formation. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 14732–14737.

Hildebrandt TM, Nunes Nesi A, Araújo WL, Braun H-P. 2015. Amino acid catabolism in plants. *Molecular Plant* **8**: 1563–1579.

Ho C-L, Noji M, Saito M, Saito K. 1999. Regulation of serine biosynthesis in *Arabidopsis*: crucial role of plastidic 3-phosphoglycerate dehydrogenase in non-photosynthetic tissues. *Journal of Biological Chemistry* **274**: 397–402.

Ho C-L, Saito K. 2001. Molecular biology of the plastidic phosphorylated serine biosynthetic pathway in *Arabidopsis thaliana*. *Amino acids* **20**: 243–259.

- Jansen RC, Nap JP. 2001.** Genetical genomics: the added value from segregation. *Trends in Genetics* **17**: 388–391.
- Jensen JK, Kim H, Cocuron J-C, Orlor R, Ralph J, Wilkerson CG. 2011.** The DUF579 domain containing proteins IRX15 and IRX15-L affect xylan synthesis in *Arabidopsis*. *The Plant Journal* **66**: 387–400.
- Jensen JK, Johnson NR, Wilkerson CG. 2014.** *Arabidopsis thaliana* IRX10 and two related proteins from psyllium and *Physcomitrella patens* are xylan xylosyltransferases. *The Plant Journal* **80**: 207–215.
- Jiang N, Wiemels RE, Soya A, Whitley R, Held M, Faik A. 2016.** Composition, assembly, and trafficking of a wheat xylan synthase complex. *Plant Physiology* **170**: 1999–2023.
- Johnson AM, Kim H, Ralph J, Mansfield SD. 2017.** Natural acetylation impacts carbohydrate recovery during deconstruction of *Populus trichocarpa* wood. *Biotechnology for Biofuels* **10**: 48.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014.** InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012.** KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* **40**: 109–114.
- Kullan ARK, van Dyk MM, Hefer CA, Jones N, Kanzler A, Myburg AA. 2012.** Genetic dissection of growth, wood basic density and gene expression in interspecific backcrosses of *Eucalyptus grandis* and *E. urophylla*. *BMC Genetics* **13**: 60.
- Kumar M, Thammannagowda S, Bulone V, Chiang V, Han K-H, Joshi CP, Mansfield SD, Mellerowicz E, Sundberg B, Teeri T, et al. 2009.** An update on the nomenclature for the cellulose synthase genes in *Populus*. *Trends in Plant Science* **14**: 248–254.
- Kürsteiner O, Dupuis I, Kuhlemeier C. 2003.** The *PYRUVATE DECARBOXYLASE1* gene of *Arabidopsis* is required during anoxia but not other environmental stresses. *Plant Physiology* **132**: 968–978.
- Langfelder P, Horvath S. 2008.** WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559.

- Lee C, Teng Q, Zhong R, Ye Z-H. 2011.** The four *Arabidopsis REDUCED WALL ACETYLATION* genes are expressed in secondary wall-containing cells and required for the acetylation of xylan. *Plant & Cell Physiology* **52**: 1289–1301.
- Lee C, Teng Q, Zhong R, Yuan YX, Haghghat M, Ye Z-H. 2012.** Three *Arabidopsis* DUF579 domain-containing GXM proteins are methyltransferases catalyzing 4-*O*-methylation of glucuronic acid on xylan. *Plant & Cell Physiology* **53**: 1934–1949.
- Lee C, Teng Q, Zhong R, Ye Z-H. 2014.** Alterations of the degree of xylan acetylation in *Arabidopsis* xylan mutants. *Plant Signaling & Behavior* **9**: e27797.
- Letunic I, Bork P. 2016.** Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**: 242–245.
- Li Z, Omranian N, Neumetzler L, Wang T, Herter T, Usadel B, Demura T, Giavalisco P, Nikoloski Z, Persson S. 2016.** A transcriptional and metabolic framework for secondary wall formation in *Arabidopsis*. *Plant Physiology* **172**: 1334–1351.
- Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu J, Stamatakis AP, Linder CR. 2012.** SATé-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology* **61**: 90–106.
- Maeda H, Yoo H, Dudareva N. 2011.** Prephenate aminotransferase directs plant phenylalanine biosynthesis via aroenate. *Nature Chemical Biology* **7**: 19–21.
- Manabe Y, Verherbruggen Y, Gille S, Harholt J, Chong S-L, Pawar PM, Mellerowicz EJ, Tenkanen M, Cheng K, Pauly M, *et al.* 2013.** REDUCED WALL ACETYLATION proteins play vital and distinct roles in cell wall *O*-acetylation in *Arabidopsis*. *Plant Physiology* **163**: 1107–1117.
- Marriott PE, Gómez LD, McQueen-Mason SJ. 2016.** Unlocking the potential of lignocellulosic biomass through plant science. *New Phytologist* **209**: 1366–1381.
- McDermott-Roe C, Ye J, Ahmed R, Sun X-M, Serafín A, Ware J, Bottolo L, Muckett P, Cañas X, Zhang J, *et al.* 2011.** Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature* **478**: 114.
- Mentzen WI, Peng J, Ransom N, Nikolau BJ, Wurtele ES. 2008.** Articulation of three core metabolic processes in *Arabidopsis*: Fatty acid biosynthesis, leucine catabolism and starch metabolism. *BMC Plant Biology* **8**: 76.

- Mizrachi E, Berger DK, Mansfield SD, Myburg AA. 2013.** Functional genomics and systems genetics of cellulose biosynthesis in *Eucalyptus*. PhD thesis, University of Pretoria, Pretoria, South Africa.
- Mizrachi E, Myburg AA. 2016.** Systems genetics of wood formation. *Current Opinion in Plant Biology* **30**: 94–100.
- Mizrachi E, Verbeke L, Christie N, Fierro AC, Mansfield SD, Davis MF, Gjersing E, Tuskan GA, Van Montagu M, Van de Peer Y, *et al.* 2017.** Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proceedings of the National Academy of Sciences of the United States of America* **114**: 1195–1200.
- Mortimer JC, Faria-Blanc N, Yu X, Tryfona T, Sorieul M, Ng YZ, Zhang Z, Stott K, Anders N, Dupree P. 2015.** An unusual xylan in *Arabidopsis* primary cell walls is synthesised by GUX3, IRX9L, IRX10L and IRX14. *The Plant Journal* **83**: 413–426.
- Mortimer JC, Miles GP, Brown DM, Zhang Z, Segura MP, Weimar T, Yu X, Seffen KA, Stephens E, Turner SR, *et al.* 2010.** Absence of branches from xylan in *Arabidopsis gux* mutants reveals potential for simplification of lignocellulosic biomass. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17409–17414.
- Mutwil M, Ruprecht C, Giorgi FM, Bringmann M, Usadel B, Persson S. 2009.** Transcriptional wiring of cell wall-related genes in *Arabidopsis*. *Molecular Plant* **2**: 1015–1024.
- 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.
- Ohtani M, Morisaki K, Sawada Y, Sano R, Uy ALT, Yamamoto A, Kurata T, Nakano Y, Suzuki S, Matsuda M, *et al.* 2016.** Primary metabolism during biosynthesis of secondary wall polymers of protoxylem vessel elements. *Plant Physiology* **172**: 1612–1624.
- Oikawa A, Joshi HJ, Rennie EA, Ebert B, Manisseri C, Heazlewood JL, Scheller HV, Joshua L. 2010.** An integrative approach to the identification of *Arabidopsis* and rice genes involved in xylan and secondary wall development. *PloS ONE* **5**: e15481.
- Oliver DJ, Nikolau BJ, Wurtele ES. 2009.** Acetyl-CoA—Life at the metabolic nexus. *Plant*

Science **176**: 597–601.

Ouyang B, Fei Z, Joung J-G, Kolenovsky A, Koh C, Nowak J, Caplan A, Keller WA, Cui Y, Cutler AJ, et al. 2012. Transcriptome profiling and methyl homeostasis of an *Arabidopsis* mutant deficient in S-ADENOSYLHOMOCYSTEINE HYDROLASE1 (SAHH1). *Plant Molecular Biology* **79**: 315–331.

Pauly M, Scheller H V. 2000. O-Acetylation of plant cell wall polysaccharides: identification and partial characterization of a rhamnogalacturonan O-acetyl-transferase from potato suspension-cultured cells. *Planta* **210**: 659–667.

Pawar PM-A, Ratke C, Balasubramanian VK, Chong S-L, Gandla ML, Adriasola M, Sparrman T, Hedenstrom M, Szwaj K, Derba-Maceluch M, et al. 2017. Downregulation of *RWA* genes in hybrid aspen affects xylan acetylation and wood saccharification. *New Phytologist* **214**: 1491–1505.

Pereira CS, Silveira RL, Dupree P, Skaf MS. 2017. Effects of xylan side-chain substitutions on xylan-cellulose interactions and implications for thermal pretreatment of cellulosic biomass. *Biomacromolecules* **18**: 1311–1321.

Persson S, Caffall KH, Freshour G, Hilley MT, Bauer S, Poindexter P, Hahn MG, Mohnen D, Somerville C. 2007. The *Arabidopsis irregular xylem8* mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. *The Plant Cell* **19**: 237–255.

Petersen PD, Lau J, Ebert B, Yang F, Verhertbruggen Y, Kim JS, Varanasi P, Suttangkakul A, Auer M, Loqué D, et al. 2012. Engineering of plants with improved properties as biofuels feedstocks by vessel-specific complementation of xylan biosynthesis mutants. *Biotechnology for Biofuels* **5**: 84.

Picault N, Palmieri L, Pisano I, Hodges M, Palmieri F. 2002. Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria: bacterial expression, reconstitution, functional characterization, and tissue distribution. *Journal of Biological Chemistry* **277**: 24204–24211.

Picault N, Hodges M, Palmieri L, Palmieri F. 2004. The growing family of mitochondrial carriers in *Arabidopsis*. *Trends in Plant Science* **9**: 138–146.

Pinard D, Fierro AC, Marchal K, Myburg AA, Mizrahi E. 2019. Organellar carbon

- metabolism is co-ordinated with distinct developmental phases of secondary xylem. *New Phytologist*.
- Plomion C, Leprovost G, Stokes A. 2001.** Wood formation in trees. *Plant Physiology* **127**: 1513–1523.
- Prabhakar V, Löttgert T, Geimer S, Dörmann P, Krüger S, Vijayakumar V, Schreiber L, Göbel C, Feussner K, Feussner I, et al. 2010.** Phosphoenolpyruvate provision to plastids is essential for gametophyte and sporophyte development in *Arabidopsis thaliana*. *The Plant Cell* **22**: 2594–2617.
- Price MN, Dehal PS, Arkin AP. 2010.** FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.
- Proost S, Van Bel M, Vaneechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B, Vandepoele K. 2015.** PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Research* **43**: 974–981.
- Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, et al. 2014.** Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**: 1246843.
- Ranik M, Myburg AA. 2006.** Six new *CELLULOSE SYNTHASE* genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. *Tree Physiology* **26**: 545–556.
- Ravanel S, Block MA, Rippert P, Jabrin S, Curien G, Rébeillé F, Douce R. 2004.** Methionine metabolism in plants: chloroplasts are autonomous for *de novo* methionine synthesis and can import S-adenosylmethionine from the cytosol. *The Journal of Biological Chemistry* **279**: 22548–22557.
- Reis D, Vian B. 2004.** Helicoidal pattern in secondary cell walls and possible role of xylans in their construction. *Comptes Rendus Biologies* **327**: 785–790.
- Rennie EA, Hansen SF, Baidoo EEK, Hadi MZ, Keasling JD, Scheller HV. 2012.** Three members of the *Arabidopsis* glycosyltransferase 8 family are xylan glucuronosyltransferases. *Plant Physiology* **159**: 1408–1417.

- Rennie EA, Scheller HV. 2014.** Xylan biosynthesis. *Current Opinion in Biotechnology* **26**: 100–107.
- Roka L, Koudounas K, Daras G, Zoidakis J, Vlahou A, Kalaitzis P, Hatzopoulos P. 2018.** Proteome of olive non-glandular trichomes reveals protective protein network against (a)biotic challenge. *Journal of Plant Physiology* **231**: 210–218.
- Ruprecht C, Mutwil M, Saxe F, Eder M, Nikoloski Z, Persson S. 2011.** Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Plant Physiology* **2**: 23.
- Scheller HV, Ulvskov P. 2010.** Hemicelluloses. *Annual Review of Plant Biology* **61**: 263–289.
- Schultink A, Naylor D, Dama M, Pauly M. 2015.** The role of the plant-specific ALTERED XYLOGLUCAN9 protein in *Arabidopsis* cell wall polysaccharide O-acetylation. *Plant Physiology* **167**: 1271–1283.
- Schwender J, Ohlrogge JB, Shachar-Hill Y. 2003.** A flux model of glycolysis and the oxidative pentosephosphate pathway in developing *Brassica napus* embryos. *The Journal of Biological Chemistry* **278**: 29442–29453.
- Schwender J, Shachar-Hill Y, Ohlrogge JB. 2006.** Mitochondrial metabolism in developing embryos of *Brassica napus*. *The Journal of Biological Chemistry* **281**: 34040–34047.
- Sharples SC, Fry SC. 2007.** Radioisotope ratios discriminate between competing pathways of cell wall polysaccharide and RNA biosynthesis in living plant cells. *The Plant Journal* **52**: 252–262.
- Simmons TJ, Mortimer JC, Bernardinelli OD, Pöppler A-C, Brown SP, deAzevedo ER, Dupree R, Dupree P. 2016.** Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. *Nature Communications* **7**: 13902.
- Smith PJ, Wang H-T, York WS, Peña MJ, Urbanowicz BR. 2017.** Designer biomass for next-generation biorefineries: leveraging recent insights into xylan structure and biosynthesis. *Biotechnology for Biofuels* **10**: 286.
- Song D, Sun J, Li L. 2014.** Diverse roles of PtrDUF579 proteins in *Populus* and PtrDUF579-1 function in vascular cambium proliferation during secondary growth. *Plant Molecular Biology* **85**: 601–612.

- Stančáková A, Civelek M, Saleem NK, Soininen P, Kangas AJ, Cederberg H, Paananen J, Pihlajamäki J, Bonnycastle LL, Morken MA, et al. 2012.** Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* **61**: 1895–1902.
- Stein O, Avin-Wittenberg T, Krahner I, Zemach H, Bogol V, Daron O, Aloni R, Fernie AR, Granot D. 2016a.** *Arabidopsis* FRUCTOKINASES are important for seed oil accumulation and vascular development. *Frontiers in Plant Science* **7**: 2047.
- Stein O, Damari-Weissler H, Secchi F, Rachmilevitch S, German MA, Yeselson Y, Amir R, Schaffer A, Holbrook NM, Aloni R, et al. 2016b.** The tomato plastidic fructokinase SIFRK3 plays a role in xylem development. *New Phytologist* **209**: 1484–1495.
- Suzuki S, Suzuki H. 2014.** Recent advances in forest tree biotechnology. *Plant Biotechnology* **31**: 1–9.
- Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. 2004.** MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* **37**: 914–939.
- Turner JE, Greville K, Murphy EC, Hooks MA. 2005.** Characterization of *Arabidopsis* fluoroacetate-resistant mutants reveals the principal mechanism of acetate activation for entry into the glyoxylate cycle. *The Journal of Biological Chemistry* **280**: 2780–2787.
- Urbanowicz BR, Peña MJ, Ratnaparkhe S, Avci U, Backe J, Steet HF, Foston M, Li H, O'Neill MA, Ragauskas AJ, et al. 2012.** 4-*O*-methylation of glucuronic acid in *Arabidopsis* glucuronoxylan is catalyzed by a Domain of Unknown Function family 579 protein. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 14253–14258.
- Urbanowicz BR, Peña MJ, Moniz HA, Moremen KW, York WS. 2014.** Two *Arabidopsis* proteins synthesize acetylated xylan *in vitro*. *The Plant Journal* **80**: 197–206.
- Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W. 2012.** A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *The Plant Cell* **24**: 3506–3529.
- 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.
- Voll LM, Jamai A, Renné P, Voll H, McClung CR, Weber APM. 2006.** The photorespiratory *Arabidopsis shm1*; mutant is deficient in *SHM1*. *Plant Physiology* **140**: 59–66.
- Wang S, Basten CJ, Zeng ZB. 2006.** Windows QTL Cartographer 2.5. *Department of Statistics, North Carolina State University, Raleigh, NC*.
- Wang JP, Naik PP, Chen H-C, Shi R, Lin C-Y, Liu J, Shuford CM, Li Q, Sun Y-H, Tunlaya-Anukit S, et al. 2014.** Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in *Populus trichocarpa*. *The Plant Cell* **26**: 894–914.
- Wierzbicki MP, Maloney V, Mizrahi E, Myburg AA. 2019.** Xylan in the middle: Understanding xylan biosynthesis and its metabolic dependencies toward improving wood fiber for industrial processing. *Frontiers in Plant Science* **10**: 176.
- Xiong G, Cheng K, Pauly M. 2013.** Xylan *O*-acetylation impacts xylem development and enzymatic recalcitrance as indicated by the *Arabidopsis* mutant *tbl29*. *Molecular Plant* **6**: 1373–1375.
- Yang Y, Yoo CG, Winkeler KA, Collins CM, Hinchey MAW, Jawdy SS, Gunter LE, Engle NL, Pu Y, Yang X, et al. 2017.** Overexpression of a Domain of Unknown Function 231-containing protein increases *O*-xylan acetylation and cellulose biosynthesis in *Populus*. *Biotechnology for Biofuels* **10**: 311.
- Yen JY, Senger RS, Gillaspie GE. 2013.** Systems metabolic engineering of *Arabidopsis* for increased cellulose production. MSc dissertation. Virginia Polytechnic Institute and State University, Virginia, United States of America.
- Yuan Y, Teng Q, Zhong R, Ye Z-H. 2013.** The *Arabidopsis* DUF231 domain-containing protein ESK1 mediates 2-*O*- and 3-*O*-acetylation of xylosyl residues in xylan. *Plant & Cell Physiology* **54**: 1186–1199.
- Yuan Y, Teng Q, Lee C, Zhong R, Ye Z-H. 2014a.** Modification of the degree of 4-*O*-methylation of secondary wall glucuronoxylan. *Plant Science* **219–220**: 42–50.
- Yuan Y, Teng Q, Zhong R, Ye Z-H. 2014b.** Identification and biochemical characterization

- of four wood-associated glucuronoxylan methyltransferases in *Populus*. *PLoS ONE* **9**: e87370.
- Yuan YX, Teng Q, Zhong R, Haghighat M, Richardson EA, Ye Z-H. 2016a.** Mutations of *Arabidopsis TBL32* and *TBL33* affect xylan acetylation and secondary wall deposition. *PLoS ONE* **11**: e0146460.
- Yuan Y, Teng Q, Zhong R, Ye Z-H. 2016b.** *TBL3* and *TBL31*, two *Arabidopsis* DUF231 domain proteins, are required for 3-*O*-monoacetylation of xylan. *Plant & Cell Physiology* **57**: 35–45.
- Yuan YX, Teng Q, Zhong R, Ye Z-H. 2016c.** Roles of *Arabidopsis TBL34* and *TBL35* in xylan acetylation and plant growth. *Plant Science* **243**: 120–130.
- Zabalza A, van Dongen JT, Froehlich A, Oliver SN, Faix B, Gupta KJ, Schmälzlin E, Igal M, Orcaray L, Royuela M, et al. 2009.** Regulation of respiration and fermentation to control the plant internal oxygen concentration. *Plant Physiology* **149**: 1087–1098.
- Zeng W, Jiang N, Nadella R, Killen TL, Nadella V, Faik A. 2010.** A glucurono(arabino)xylan synthase complex from wheat contains members of the GT43, GT47, and GT75 families and functions cooperatively. *Plant Physiology* **154**: 78–97.
- Zeng W, Lampugnani ER, Picard KL, Song L, Wu A-M, Farion IM, Zhao J, Ford K, Doblin MS, Bacic A. 2016.** *Asparagus* IRX9, IRX10, and IRX14A are components of an active xylan backbone synthase complex that forms in the Golgi apparatus. *Plant Physiology* **171**: 93–109.
- Zhang Y, Beard KFM, Swart C, Bergmann S, Krahnert I, Nikoloski Z, Graf A, Ratcliffe RG, Sweetlove LJ, Fernie AR, et al. 2017.** Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle. *Nature Communications* **8**: 15212.
- Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, Rhee SY. 2005.** MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiology* **138**: 27–37.
- Zhang B, Liu J-Y. 2013.** Mass spectrometric identification of *in vivo* phosphorylation sites of differentially expressed proteins in elongating cotton fiber cells. *PLoS ONE* **8**: e58758.
- Zhang J, Siika-Aho M, Tenkanen M, Viikari L. 2011.** The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed.

Biotechnology for Biofuels **4**: 60.

Zhong R, Cui D, Ye Z-H. 2017. Regiospecific acetylation of xylan is mediated by a group of DUF231-containing *O*-acetyltransferases. *Plant & Cell Physiology* **58**: 2126–2138.

Zhong R, Cui D, Ye Z-H. 2018. Members of the DUF231 family are *O*-acetyltransferases catalyzing 2-*O*- and 3-*O*-acetylation of mannan. *Plant and Cell Physiology*: <https://doi.org/10.1093/pcp/pcy159>.

Zinkgraf M, Liu L, Groover A, Filkov V. 2017. Identifying gene coexpression networks underlying the dynamic regulation of wood-forming tissues in *Populus* under diverse environmental conditions. *New Phytologist* **214**: 1464–1478.

Zolman BK, Yoder A, Bartel B. 2000. Genetic analysis of indole-3-butyric acid responses in *Arabidopsis thaliana* reveals four mutant classes. *Genetics* **156**: 1323–1337

Table 1 Mapped module eigengene QTLs (meQTLs) from co-expression analysis in xylogenic tissues of *Eucalyptus*

Module Eigengene ^a	meQTL_number ^b	Chromosome ^c	Peak_position ^d	Peak_LOD ^e	R ^{2f}	Additive ^g	LOD-1 interval ^h		LOD-2 interval ⁱ	
ME1	1	9	0.01	3.317	0.079	-0.046	0.01	5.54	0.01	12.2
ME1	2	10	77.33	2.705	0.064	0.042	68.03	84.52	66.57	87.62
ME2	1	6	22.01	3.233	0.084	-0.047	10.83	27.79	4.32	33.22
ME3	1	9	0.01	2.762	0.067	-0.042	0.01	11.94	0.01	12.58
ME4	1	3	64.58	2.545	0.066	-0.042	53.3	71.73	53.05	81.81
ME4	2	9	0.01	2.708	0.063	-0.041	0.01	7.97	0.01	12.56
ME4	3	10	77.33	3.635	0.085	0.048	69.23	84.98	66.74	87.46
ME5	1	10	77.33	3.247	0.082	0.047	70.84	86.02	66.76	87.52

^aThe names of module eigengenes (ME) that correspond to their respective expression module.

^bThe number assigned to expression quantitative trait loci (eQTLs) detected for each module eigengene.

^cChromosome on which the eQTL occurs.

^dPosition of the eQTL peak indicated in cM.

^eThe Log of odds (LOD) value for the eQTL peak.

^fPercentage of variance contributed by the eQTL to the expression variance of the module eigengene.

^gThe additive effect of the eQTL. Negative values indicated that higher expression was derived from the *Eucalyptus urophylla* allele at the locus.

^hThe range of the eQTL in the LOD-1 interval in cM.

ⁱThe range of the eQTL in the LOD-2 interval in cM.

Table 2 Functional enrichment and annotation of *Eucalyptus* xylem expressed genes present in the five expression modules and affected by one or more of the seven regulatory hotspots*

GO Biological Process					
Term ID	Description	Process involved	Genes in category	EM genes in category	Adjusted p-value
GO:0045492	xylan biosynthetic process	Xylan biosynthesis	259	47	6.10E-30
GO:0010413	glucuronoxylan metabolic process	Xylan biosynthesis	257	46	5.67E-29
GO:0010417	glucuronoxylan biosynthetic process	Xylan biosynthesis	16	9	4.26E-09
GO:0009834	secondary cell wall biogenesis	Cell wall general	60	14	5.36E-09
GO:0044036	cell wall macromolecule metabolic process	Cell wall general	97	14	4.75E-06
GO:2000652	regulation of secondary cell wall biogenesis	Transcriptional regulation	22	8	6.13E-06
GO:0009225	nucleotide-sugar metabolic process	Cellulose and xylan precursor	20	6	2.45E-03
GO:0009832	plant-type cell wall biogenesis	Cell wall general	178	14	1.00E-02
GO:0010089	xylem development	Cell wall general	89	10	1.05E-02
GO:0042546	cell wall biogenesis	Cell wall general	96	10	2.09E-02
GO Molecular Function					
GO:0009044	xylan 1,4-beta-xylosidase activity	Xylan biosynthesis	12	5	2.38E-03
GO:0015018	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase activity	Xylan biosynthesis	8	4	1.25E-02
KEGG Pathways					
ath00520	Amino sugar and nucleotide sugar metabolism	Cellulose and xylan precursor	138	15	3.35E-06
ath01230	Biosynthesis of amino acids	Xylan and lignin precursor	264	20	8.24E-06
ath01130	Biosynthesis of antibiotics	Xylan and lignin precursor	486	26	9.54E-05
ath01100	Metabolic pathways	Cell wall. general	2016	64	5.74E-04
ath00260	Glycine, serine and threonine metabolism	Xylan and lignin precursor	82	8	1.36E-02
ath00400	Phenylalanine, tyrosine and tryptophan biosynthesis	Lignin biosynthesis	47	6	2.28E-02
Mapman					
10	cell wall	Cellulose biosynthesis	340	28	7.22E-10
13	amino acid metabolism	Xylan and lignin precursor	263	17	1.85E-04

*Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Mapman were used to identify enriched

biological processes, molecular functions and metabolic pathways that are co-regulated with xylan modification, as in Pinard *et al.*, 2019.

Figure legends:

Fig. 1 Putative homologs of xylan modification genes in *Eucalyptus*.

One RWA, two GUX, three GXM and seven TBL genes were identified in *Eucalyptus* through a combinatorial approach of phylogenetics, co-expression with secondary cell wall (SCW) associated genes and network reconstruction. Black text indicates a *Eucalyptus* xylan modification gene corresponds phylogenetically to experimentally validated xylan modification genes from *Arabidopsis* and *Populus* as well as being highly co-expressed (>0.6) with other SCW associated genes. Coloured italicised text indicates partial matches, where light green text indicates phylogenetic homology only whereas orange text indicates high co-expression with SCW genes only. Stars (*) indicate potentially novel xylan modification genes.

Fig. 2 Architecture of genetic regulation and systems genetics model of *Eucalyptus* xylan modification.

(a) Seven hotspot loci identified based on expression quantitative trait loci (eQTLs) per centimorgan (cM) sliding window bins. Frequency of eQTLs present per cM sliding window bin with a cut-off set at nine eQTLs per cM sliding window bin for the four *Eucalyptus* chromosomes where the hotspots located. Hotspot detection of all 11 chromosomes is represented in **Fig. S9**. (b) Proportion of *trans*-eQTLs locating to the seven hotspot loci compared to those that do not and the number of *trans*-eQTLs which co-locate to each hotspot locus. (c) Genomic representation of all *trans*-eQTLs originating from the seven hotspots affecting expression of genes in each expression module (EM). The outer track represents density of *trans*-eQTLs per 1 cM bin for each of the 11 *Eucalyptus* chromosomes. The inner track represents the seven hotspot loci identified in this study with colouring corresponding to the colours assigned to hotspots in this study. The position of an eigen-eQTL is indicated in the innermost track with colouring corresponding to the EM. (d) Network illustration of the relationship between three layers, namely: the expression modules (EMs), the hotspot loci and the biological processes which are co-regulated with xylan modification. Five expression modules were linked to the regulatory loci, which affect the expression of genes in that module. Eight enriched processes were linked to the hotspot loci, which affect genes associated with the process. The pie chart at each process illustrates the percentage of the EM represented in the process. Nodes represent the EMs, hotspot loci and processes co-regulated with xylan modification. Node sizes of EMs and hotspot loci represent number of genes and are

transformed to facilitate clarity. Node sizes of processes do not represent number of genes. Edges connecting hotspot loci to processes represent genes within the process which are affected by the hotspot loci, whereas edges connecting EMs to hotspot loci represent genes in EMs affected by hotspot loci. Legends illustrate the amount of genes which correspond to the thickness of the edges: the edges from hotspots to processes were not transformed, whereas edges from EMs to hotspot loci were transformed for clarity. (EM1: red, EM2: yellow, EM3: green, EM4: blue, EM5: purple). (Hotspot (HS) colours: HS_3.1: orange, HS_3.3: dark green, HS_6.1: crimson, HS_9.1: turquoise, HS_10.3: sienna, HS_10.4: magenta.

Fig. 3 S-adenosylmethionine, phenylalanine and phenylpropanoid biosynthesis is highly represented in EM1 and regulated primarily by HS_9.1 and HS_10.3.

Reactions on the figure are only numbered if the genes involved are found in an expression module (EM) or affected by one or more of the regulatory hotspot loci. Elongated ovals represent metabolite, rounded rectangles represent genes (xylan modification genes are represented as membrane bound enzymes), circles represent regulatory hotspot loci and rhomboids represent pathways from where metabolites may originate. Rounded rectangles are coloured according to the expression module to which they belong; EM1: red, EM2: yellow, EM3: green, EM4: blue and EM5: purple. Circles are labelled and coloured according to the regulatory hotspot (HS) locus they represent; HS_3.1: Orange, HS_3.3: green, HS_6.1: crimson, HS_9.1: turquoise, HS_10.2: olive, HS_10.3: sienna, HS_10.4: magenta. Genes affected by one or more regulatory hotspot locus are illustrated as rounded squares with circles above/next to them; multiple stacked dots represent multiple regulatory hotspot loci affecting the gene. Circles not above/next to squares represent genes not found in any EMs but are only affected by one of the regulatory hotspot loci. Dashed lines represent transport/diffusion between two compartments. The orange oval represents the Golgi, while the remaining reactions occur within the cytosol, plastid and mitochondrion, but these compartments were not displayed for clarity. Numbers present in the figure signify reaction steps of enzymes: 1, ribose 5-phosphate isomerase; 2, transketolase; 3, transaldolase; 4, 3-deoxy-7-phosphohelptulonate synthase; 5, 3-dehydroquinate synthase; 6, dehydroquinate dehydratase; 7, shikimate kinase; 8, 3-phosphoshikimate 1-carboxyvinyltransferase; 9, chorismate synthase; 10, chorismate mutase; 11, aroenate dehydratase; 12, prephenate aminotransferase; 13, 3-phospho-D-glycerate dehydrogenase; 14, phosphoserine aminotransferase; 15, phosphoserine phosphatase; 16, serine hydroxymethyltransferase; 17, cysteine synthase/ O-acetylserine (thiol) lyase; 18, methionine synthase; 19, s-adenosylmethionine synthase; 20, caffeoyl-CoA O-

methyltransferase; 21, caffeic acid 3-O-methyltransferase. Some parts of the pathway were not shown for simplicity. Information and expression profiles on the represented pathways present in **Table S15**. **Fig. S13** illustrates the fully labelled figure including all reactions, while information on the genes and expression profiles in the (*Eucalyptus urophylla* x *E. grandis*) x *E.urophylla* F2 backcross are present in **Table S15**. A legend to the xylan molecule can be found in **Fig. 1**.

Fig. 4 Comparison of xylan modification genes present in EM5 and EM1.

Six possible xylan modification configurations are shown below the enzymes responsible for this modification and its expression module (EM) membership. (A) Xylan modification candidates present in EM5 represented by the protein families RWA, GUX and TBL. Glucuronic acid and acetyl modifications are known to compete for substrate space, which might indicate that RWA is in this expression module to supply acetyl-coenzyme A for these TBLs that compete with GUX (Lee *et al.*, 2014). (B) Xylan modification candidates present in EM1 represented by the protein families GXM and TBL. Utilisation of S-adenosylmethionine as a cofactor for methylation gives support for GXM being present in this expression module. The presence of TBL33 in this expression module may be due to its specificity towards acetylation residues with α 1,2-methylglucuronic acid modification; it may be necessary for the glucuronic acid to be methylated before TBL33 acts on the residue (Yuan *et al.*, 2016a). The * symbol nomenclature applies to TBL proteins only as they are found to display regiospecificity (Zhong *et al.*, 2017). The * symbol alongside the TBL protein indicates the preferential modification performed by that TBL protein, although alternative structural fragments can still be produced by the same TBL proteins (fragments shown without the * symbol). A legend to the xylan molecule can be found in **Fig. 1**.

Fig. 5 Four possible sources of acetyl CoA originate from glycolysis, β -oxidation of fatty acids, ethanol and branched chain amino acid degradation as genes in these pathways which are present in EM1, EM2 and EM5 and primarily regulated by HS_6.1 and HS_10.3.

Reactions on the figure are only numbered if the genes involved are found in an expression module (EM) or affected by one or more of the regulatory hotspot loci. Elongated ovals represent metabolite, rounded rectangles represent genes, circles represent regulatory hotspot loci, rhomboids represent pathways from where metabolites may originate and the purple hexagon represent xylan acetylation. Elongated ovals are coloured according to the metabolic

pathway/precursor from which acetyl-CoA is derived; (i) light blue: glycolysis and tricarboxylic acid cycle, (ii) darker green: ethanol from fermentation, (iii) greyish blue: β -oxidation of fatty acids and membrane lipids and (iv) light green: breakdown of branched chain amino acids. Rounded rectangles are coloured according to the expression module to which they belong; EM1: red, EM2: yellow, EM3: green, EM4: blue and EM5: purple. Circles are labelled and coloured according to the regulatory hotspot (HS) locus they represent; HS_3.1: Orange, HS_3.3: green, HS_6.1: crimson, HS_9.1: turquoise, HS_10.2: olive, HS_10.3: sienna, HS_10.4: magenta. Genes affected by one or more regulatory hotspot locus are illustrated as rounded squares with circles above/next to them; multiple stacked dots represent multiple regulatory hotspot loci affecting the gene. Dashed lines represent transport/diffusion between two compartments. Genes involved in plastidial glycolysis have an asterisk next to them. Numbers present in the figure signify reaction steps of enzymes: 1, phosphofructose-1-kinase; 2, fructose biphosphate aldolase; 3, triose phosphate isomerase; 4, glyceraldehyde 3-phosphate dehydrogenase; 5, phosphoglycerate kinase; 6, phosphoglyceromutase; 7, phosphoenolpyruvate carboxylase; 8, pyruvate dehydrogenase complex; 9, succinate dehydrogenase; 10, mitochondrial electron transport chain; 11, ATP-citrate lyase; 12, cytosolic isocitrate dehydrogenase; 13, glutamate synthase; 14, hydroxymethylglutaryl-CoA lyase; 15, alcohol dehydrogenase; 16, mono-/diacylglycerol lipase; 17, acyl-CoA-binding protein; 18, lipases; 19, acyl-CoA synthetase; 20, acyl-CoA oxidase; 21, catalase; 22, 3-ketoacyl-CoA thiolase; 23, isocitrate lyase. Some parts of the pathway were omitted for simplicity. The plus symbol “+” indicates that the degradation of isoleucine to 2-methyl-acetoacetyl-CoA requires several reaction steps not shown in the figure. Information and expression profiles on the represented pathways present in **Table S16**. **Fig. S15** illustrates the fully labelled figure including all reaction, while information on the genes and expression profiles in the (*Eucalyptus urophylla* x *E. grandis*) x *E. urophylla* F2 backcross are present in **Table S16**. A legend to the xylan molecule can be found in **Fig. 1**.

Fig. 6 Proposed model of metabolic coordination associated with xylan modification in *Eucalyptus* developing xylem

Arrows indicate reaction steps with black lines denoting lack of expression module (EM) membership and light blue arrows indicating enzyme activity of xylan modification genes, whereas red, greenish yellow, dark blue and purple arrows indicate EM1, EM2, EM4 and EM5

membership respectively. Red hexagons represent rhamnose and green hexagons represent galacturonic acid, whereas light blue circles on the hexagons represent the oxygen group in the sugar. A legend to the xylan molecule can be found in **Fig. 1**.

Supporting information

Fig. S1 Phylogram representing a maximum likelihood analysis of RWA proteins in *Arabidopsis*, *Eucalyptus* and *Populus* coupled with relative and absolute percentile expression in xylem and leaf, and domain structure.

Fig. S2 Phylogram representing a maximum likelihood analysis of GUX proteins in *Arabidopsis*, *Eucalyptus* and *Populus* coupled with relative and absolute percentile expression in xylem and leaf, and domain structure.

Fig. S3 Phylogram representing a maximum likelihood analysis of DUF579/GXM proteins in *Arabidopsis*, *Eucalyptus* and *Populus* coupled with relative and absolute percentile expression in xylem and leaf, and domain structure.

Fig. S4 Phylogram representing a maximum likelihood analysis of DUF231/TBL proteins from clade IV in *Arabidopsis*, *Eucalyptus* and *Populus* coupled with relative and absolute percentile expression in xylem and leaf, and domain structure.

Fig. S5 Phylogram representing a maximum likelihood analysis of the full complement of DUF231/TBL proteins in *Arabidopsis*, *Eucalyptus* and *Populus* coupled with relative and absolute percentile expression in xylem and leaf, and domain structure.

Fig. S6 F-score threshold determined from the number of genes that each query gene included in the network by using a joint likelihood curve.

Fig. S7 Co-expression and clustering of 1136 genes obtained for genes which passed set criteria divided into five expression modules.

Fig. S8 Seventy percent of genes co-expressed with xylan modification were found to have at

least one trans-eQTL affecting its expression.

Fig. S9 Identification of the seven highly enriched trans-eQTL hotspot loci.

Fig. S10 Principal component analysis of the five module eigengenes.

Fig. S11 Model of genetic regulation that shapes the expression modules.

Fig. S12 Numbers of genes affected by each of the seven hotspot loci and their expression module membership.

Fig. S13 Nucleotide sugar interconversion is highly represented in EM5 and regulated primarily by HS_10.3 and HS_10.4.

Fig. S14 Complete annotation of Fig. 3 containing genes which were not present in EMs or affected by hotspot loci.

Fig. S15 Complete annotation of Fig. 5 containing genes which were not present in EMs or affected by hotspot loci.

Table S1 *Arabidopsis*, *Eucalyptus* and *Populus* genes potentially involved in xylan modification and cumulative evidence for naming of xylan modification genes in *Populus* and *Eucalyptus*

Table S2 Confirmed *Arabidopsis* SCW xylan modification genes and probable homologs in *Eucalyptus*

Table S3 *E. grandis* query genes used for the co-expression analysis and the subset that are probable xylan modification genes.

Table S4 1112 genes co-expressed with 24 xylan modification genes

Table S5 Functional enrichment and annotation of genes present in the five expression modules produced by the co-expression algorithm*

Table S6 Expression module genes corresponding to GO, KEGG and MM terms

Table S7 *E. grandis* query genes which passed criteria set during co-expression analysis, membership to an expression module and correlation with SCW associated genes

Table S8 Identification of bins significantly enriched in *trans*-eQTLs

Table S9 Xylem expressed genes affected by seven hotspots

Table S10 Overlap of hotspot loci identified in this study with previously identified global hotspots

Table S11 Core set genes corresponding to enriched terms for GO, KEGG and MapMan

Table S12 Genes corresponding to process categories on systems model

Table S13 All genes affected by seven hotspot loci

Table S14 Genes corresponding to numbered reaction steps on Fig. S13

Table S15 Genes corresponding to numbered reaction steps on Fig. 3 as well as the genes missing from Fig. 3, indicated on Fig. S14

Table S16 Genes linked to acetyl-CoA production corresponding to the numbered reaction steps on Fig. 5 as well as the genes missing from Fig. 5, indicated on Fig. S15

Table S17 Examples of how the systems model can be used as a tool to design and test new biotechnology approaches to alter cell wall structure and chemistry

Methods S1 Network based co-expression analysis and community detection clustering.

Methods S2 Global eQTL mapping and hotspot locus detection.

Fig. 1

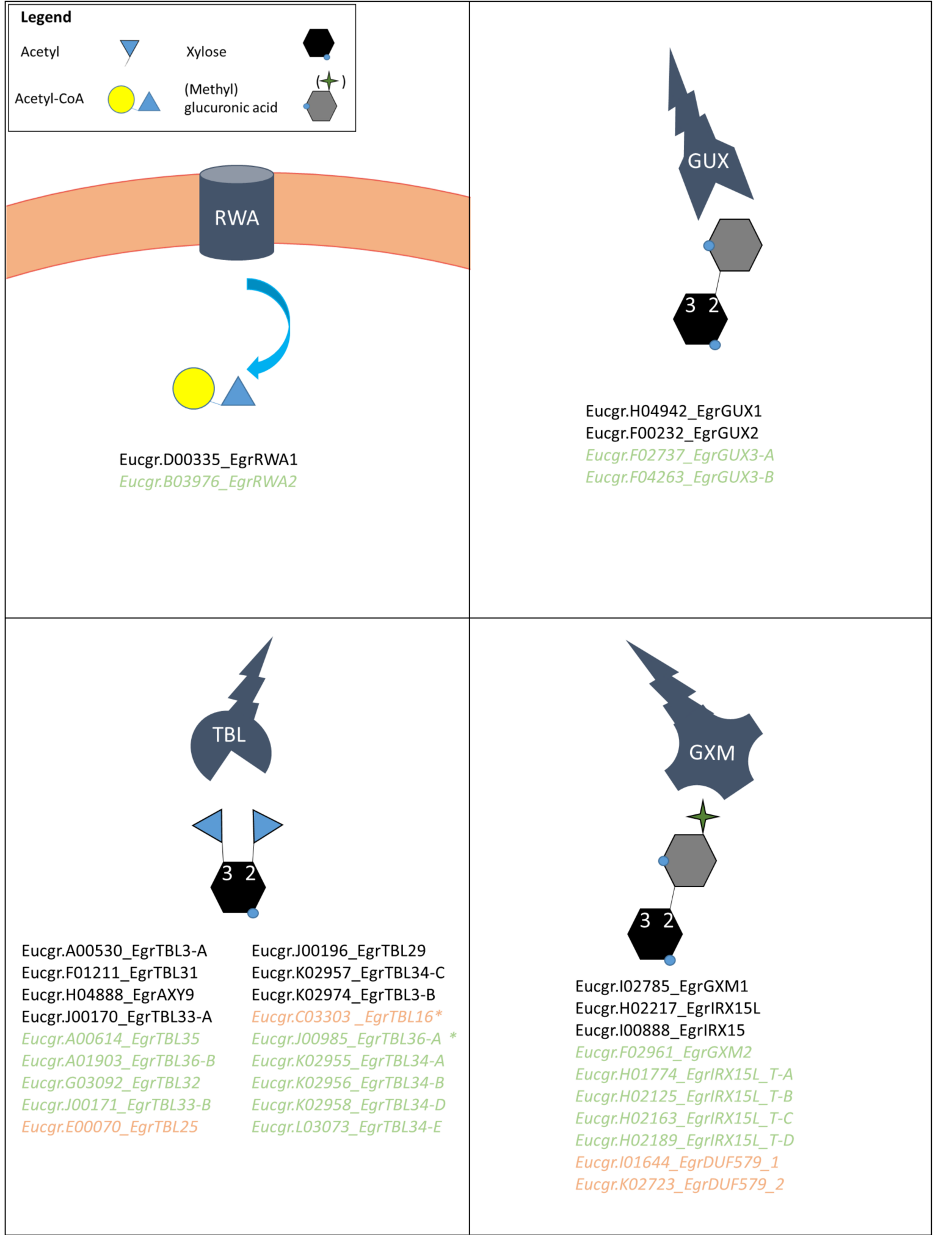


Fig. 2

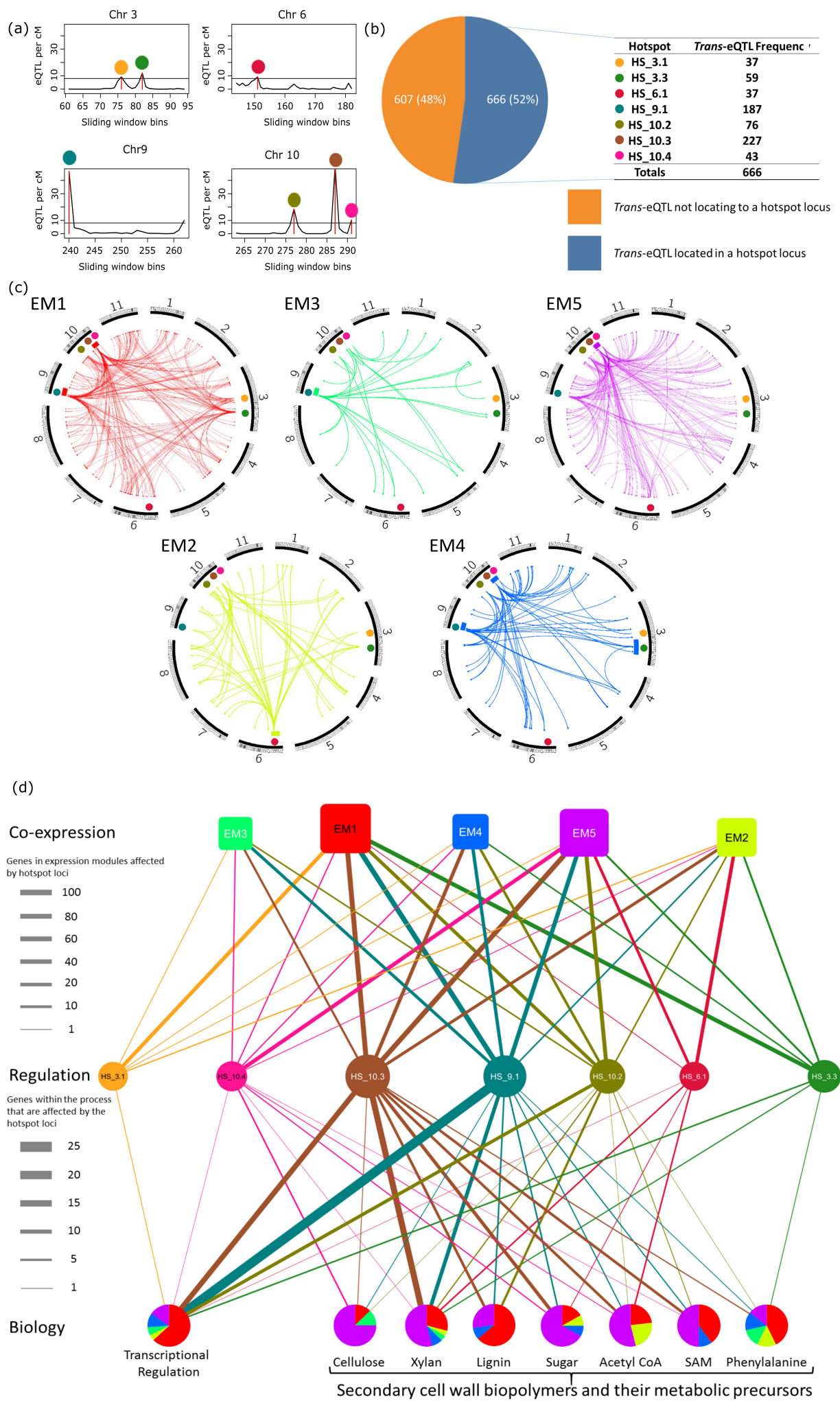


Fig. 3

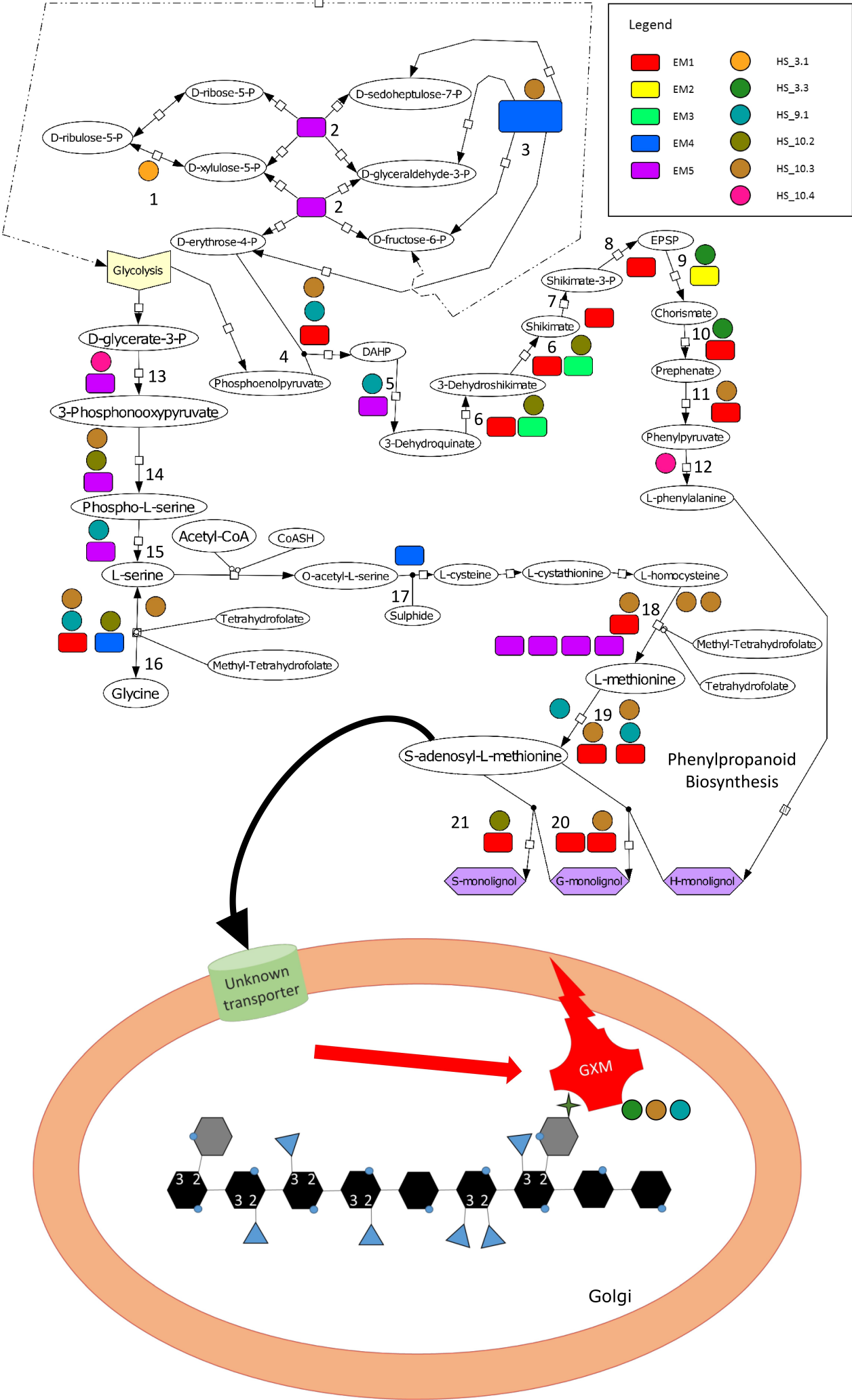
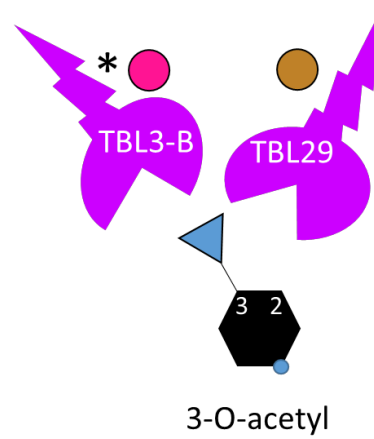
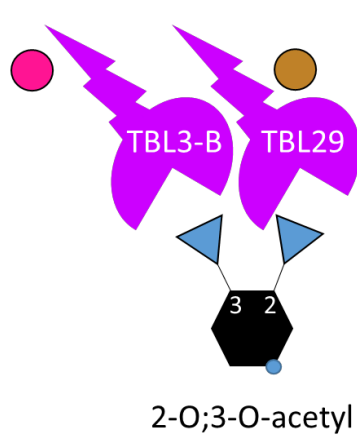
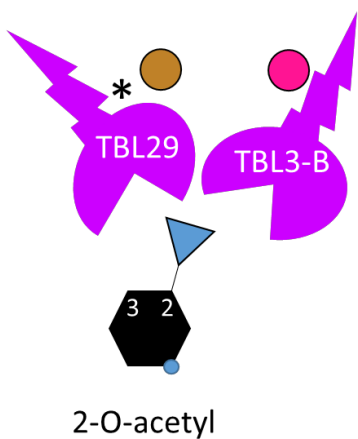
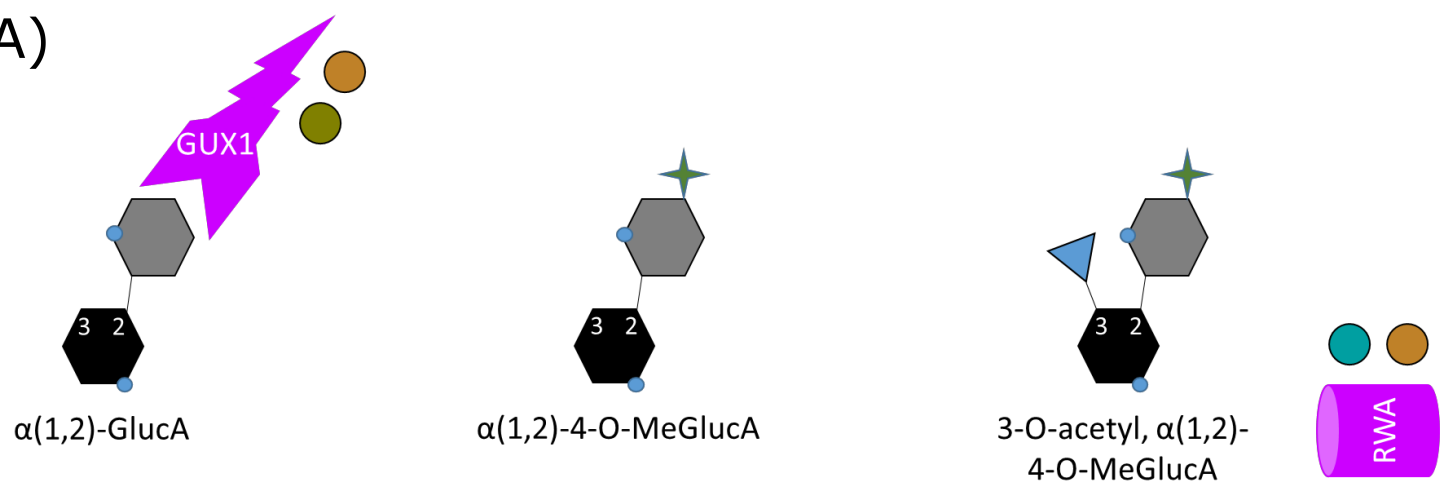


Fig. 4

(A)



Legend	
	EM1
	EM5
	HS_3.3
	HS_9.1
	HS_10.2
	HS_10.3
	HS_10.4

(B)

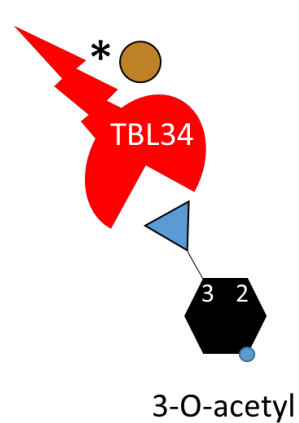
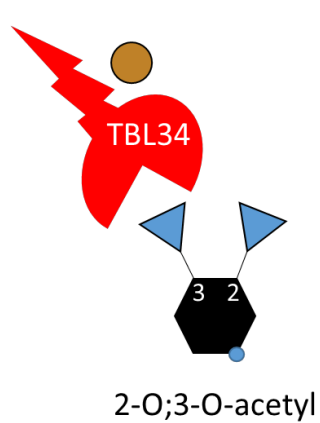
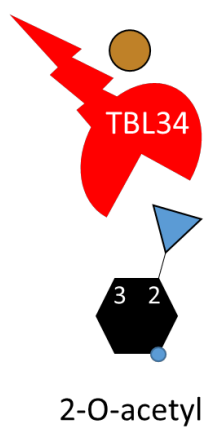
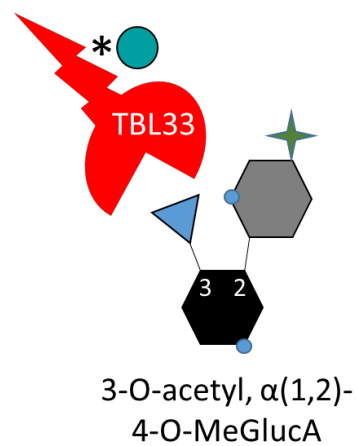
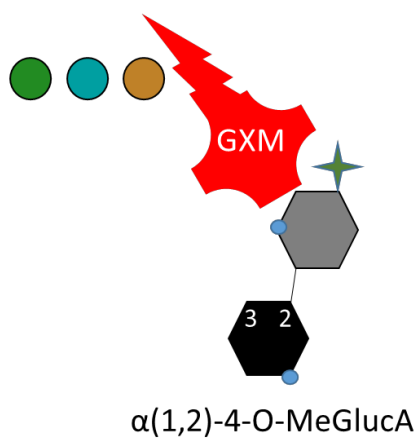
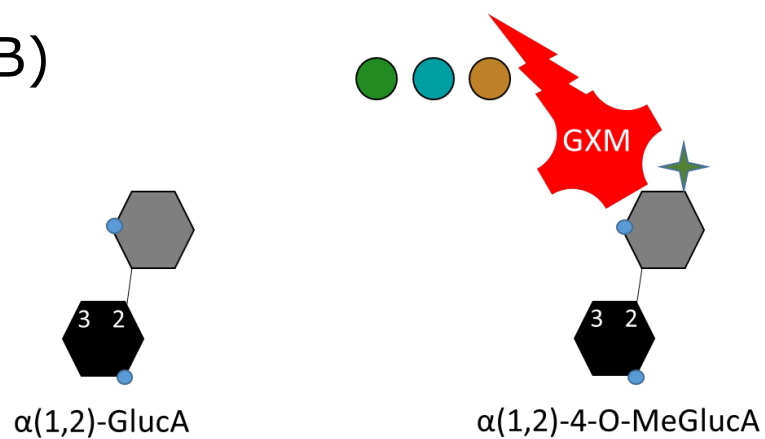


Fig. 5

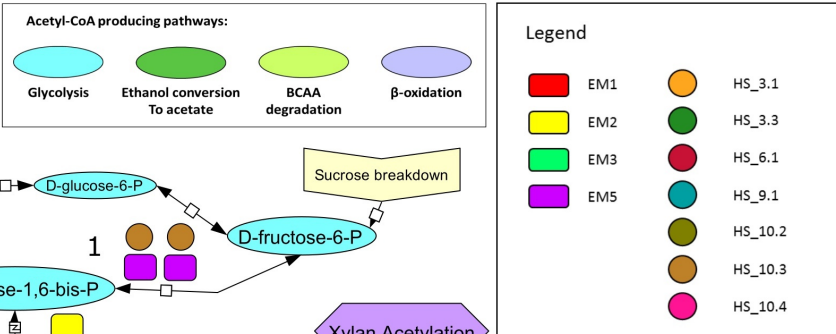


Fig. 6

