Two softwood GUX clades are responsible for distinct glucuronic acid patterns on xylan.

Jan J Lyczakowski^{1,2,*}, Li Yu², Oliver M Terrett^{2,#}, Christina Fleischmann³, Henry Temple², Glenn Thorlby³, Mathias Sorieul³ and Paul Dupree^{2,*}

¹ Department of Plant Biotechnology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Gronostajowa 7, 30-387 Krakow, Poland

² Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, United Kingdom

³ Scion, 49 Sala Street, Rotorua 3020, New Zealand

current address: Department of Biology, ETH Zürich, Universitätstrasse 2, 8092 Zürich, Switzerland

* address correspondence to Jan J. Lyczakowski (jan.lyczakowski@uj.edu.pl) and Paul Dupree (pd101@cam.ac.uk).

Keywords: softwood, plant cell walls, conifers, xylan, glucuronic acid, biomass recalcitrance

Abstract:

Wood of coniferous (softwood) trees, is a globally significant carbon sink and an important source of biomass for industrial applications. Despite its importance, very little is known about the genetic basis of softwood biosynthesis. Glucomannan and xylan are the main hemicelluloses in softwood secondary cell walls. Xylan interacts with the cellulose fibrils in a two-fold screw configuration. Moreover, we have shown that xylan GUX (GlucUronic acid substitution of Xylan)-dependent branching with glucuronic acid is critical for biomass recalcitrance. Here, we investigated the decoration patterns of xylan by softwood GUX enzymes. Using *in vitro* and *in planta* assays we demonstrate that two distinct clades of conifer GUX enzymes are active glucuronyltransferases. Interestingly, these enzymes have different specific activities, with one adding evenly spaced GlcA branches and the other one being also capable of glucuronidating two consecutive xyloses. Since xylan patterning might modulate xylan-cellulose and xylan-lignin interactions, our result further the understanding of softwood biosynthesis and can contribute to strategies aimed at modifying softwood cell wall properties.

Introduction:

The majority of terrestrial carbon is stored in forests as wood (Bar-On et al., 2018, Ramage et al., 2017, Pan et al., 2011). The main commercially used type of timber comes from softwoods. For example, softwoods account for as much as 98% of trees planted in the US managed forests (Sedjo, 2001). The abundance and sustainability of softwoods position them as perfect feedstocks for use in alternative industrial processes, tackling the climate emergency. Unfortunately, progress in conifer breeding or genetic engineering aiming to improve softwood suitability for industrial application is hindered by the poor understanding of the genetic basis for its biosynthesis.

Softwood is mainly made of tracheid cells (Ramage et al., 2017). Their secondary cell walls are composed of cellulose microfibrils embedded in a matrix of hemicellulose polysaccharides and hydrophobic lignin polymers. Cellulose is believed to be the main load-bearing structure of softwood cell walls. It constitutes a crystalline assembly of 18 or 24 glucan chains (Jarvis, 2018, Fernandes et al., 2011).

In conifers, lignin is mainly made out of guaiacyl monomers. It impregnates softwood secondary cell walls and enables water transport in trees (Vanholme et al., 2010). The two main types of hemicelluloses; Galactoglucomannan (GGM) and arabinoglucuronoxylan (AGX), account for 10-30% and 5-15% of softwood secondary cell walls, respectively (Scheller and Ulvskov, 2010). The GGM backbone consists of β -1,4 linked mannose and glucose monomers. Some mannose subunits are branched with α -1,6 linked galactose or acetyl substituents linked to carbon 2 or 3 (Willfor et al., 2008). The AGX backbone is composed of β -1,4 linked xylose monomers with α -1,2-linked 4-O-methyl glucuronic acid (MeGlcA) and α -1,3-linked arabinofuranose (Ara) branches (Timell and Syracuse, 1967, Willför et al., 2005).

The exact biological functions of conifer hemicelluloses are still debated. However, they are likely to mediate cross-linking of different cell wall components and modulate the mechanical properties of the cell wall (Martínez-Abad et al., 2020, Terrett and Dupree, 2019, Berglund et al., 2020a, Berglund et al., 2020b). Both *in vitro* binding assays (Whitney et al., 1998, Berglund et al., 2020a, Berglund et al., 2020b) and *in planta* analysis (Terashima et al., 2009, Yu et al., 2018, Terrett et al., 2019) suggest that GGM can interact with the cellulose microfibril. Recent two-dimensional solid state nuclear magnetic resonance (ssNMR) analysis of spruce wood demonstrated that GGM may directly interact with cellulose *in planta* and extend its crystalline structure (Terrett et al., 2019). Interestingly, the same study indicated that a majority of AGX can also directly bind to cellulose microfibrils. In addition to their interaction with cellulose, hemicelluloses can also bind lignin (Terrett and Dupree, 2019). Both AGX and GGM are located less than 1 nm away from lignin in spruce wood (Terrett et al., 2019). GGM is likely to form direct covalent linkages with lignin (Nishimura et al., 2018). In softwood, the glucuronic acid sidechains of AGX were proposed to form ester linkages with lignin (Oinonen et al., 2015). However, this xylan-lignin complex was never clearly identified. In Arabidopsis, the presence of GlcA on xylan has been proven critical for the maintenance of plant biomass resistance to enzymatic degradation (Lyczakowski et al., 2017). Therefore, this plausible xylan-lignin complex may be important for industrially relevant properties of softwood.

To understand how xylan contributes to the softwood cell wall properties it is necessary to fully characterise its structure. In angiosperms, xylan is decorated with

GlcA and acetyl side branches. The addition of the GlcA onto the xylan backbone is carried by GlucUronic acid substitution of Xylan (GUX) enzymes (Mortimer et al., 2010, Rennie et al., 2012). In Arabidopsis, an even pattern of branches allows the xylan to bind onto the hydrophilic faces of cellulose microfibrils via hydrogen bonds (Busse-Wicher et al., 2014, Busse-Wicher et al., 2016a, Grantham et al., 2017). The addition of evenly spaced GlcA side branches is catalysed by GUX1 enzymes (Bromley et al., 2013). In addition to this even patterning, some of the GlcA decorations are distributed with uneven spacing. This generates a xylan domain unable to bind in the same way to the hydrophilic surface of the cellulose microfibril. This incompatible patterning, in Arabidopsis, is laid out by GUX2 enzymes.

In softwoods (pine and Douglas fir), the GlcA layout on xylan was reported by Busse-Wicher to be simpler than in Arabidopsis (Busse-Wicher et al., 2016b). It consisted primarily of an even spacing of MeGlcA branches every six xylose units. In addition to the MeGlcA, softwood xylan also has Ara decorations. These Ara branches are placed predominantly with an even spacing of two xylose units from the MeGlcA. Interestingly, some reports indicated the presence of more complex AGX decoration patterns in softwoods (Shimizu et al., 1978, Yamasaki et al., 2011, Martinez-Abad et al., 2017). Martinez-Abad et al., confirmed that the main spacing between MeGlcA branches is equal to six xylose units in spruce. However, similarly to Shimizu et al., and Yamasaki et al., their work also reported the presence of MeGlcA branches on two consecutive xyloses along the AGX backbone. The biological functions of these two types of xylan patterning in softwoods and the enzymes responsible for their biosynthesis are unknown. It is however possible, that similarly to angiosperms, the patterning of softwood xylan with GlcA may be important for the control of its interaction with the cellulose microfibril.

Here, we report the discovery of a second softwood GUX clade and the characterisation of the biosynthetic activities of the two clades identified. The two GUX enzymes have distinct specific activities enabling them to lay out different patterns of GlcA branches. Our results demonstrate the biosynthetic role of GUX enzymes for the patterning of xylan in softwoods. Since xylan patterning is likely to be important for the cell wall properties of softwood, our results should influence future genetic engineering or breeding approaches aimed at improving softwood suitability for applications in the bioeconomy.

Materials and Methods:

Plant material, growth and Arabidopsis transformation

Arabidopsis thaliana (*At*) plants of the Columbia-0 ecotype were grown in a cabinet maintained at 21 °C, with a 16 h light, 8 h dark photoperiod. Two previously described *A. thaliana* insertion lines *gux1/2/3* (Mortimer et al., 2015) and *esk-kak* (Grantham et al., 2017) were used for transformation experiments. For *PrGUX1* and *PrGUX2* expression profiling, *Pinus radiata* was grown by Scion, New Zealand. For biomass analysis, alcohol insoluble residues (AIR) were prepared as previously described (Mortimer et al., 2010). Constructs used for *PtGUX2* over-expression in *gux1/2/3 A. thaliana* and for *PgGUX1* and *PtGUX2* overexpression in *esk-kak* were prepared using OpenPlant common syntax GoldenGate assembly (Patron et al., 2015). Arabidopsis transformation was performed using floral dip as previously described (Clough and Bent, 1998).

Phylogenetic and bioinformatic analysis

The coding sequences of *A. thaliana* GUX1, GUX2 and GUX3 were used to identify putative GUX encoding transcripts from *Populus trichocarpa* using data available via the NCBI BLAST service. The same Arabidopsis sequences were used as a query to identify transcripts encoding putative GUX enzymes from Coniferophyta and Gnetophyta transcriptomic data available from 1000 Plant Transcriptomes, Congenie and Gymno Plaza BLAST services (Matasci et al., 2014, Nystedt et al., 2013, Proost et al., 2015). In house *Pinus radiata* (*Pr*) genome and transcriptome resources (Scion, New Zealand) were used to recover full length DNA sequences for *PrGUX1* and *PrGUX2* genes. The sequences were confirmed by amplification and sequencing from genomic DNA and primers for droplet digital (dd) PCR analysis were designed (Table S1). All amino acid sequences were reconstructed from transcripts with ExPASy translate tool and aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm. A maximum likelihood phylogenetic tree was constructed using MEGA 6 software (Tamura et al., 2013) including the gamma distribution algorithm in the reconstruction. Table S2 summarises codes required to access gymnosperm GUX sequences studied.

For the computational expression profiling of *Picea abies* (*Pa*) GUX enzymes Congenie (Nystedt et al., 2013) and Norwood (Jokipii-Lukkari et al., 2017)

bioinformatics tools were used. For the Congenie analysis, the exlImage tool was used to visualise the expression across the entire organism. For NorWood analysis, the expression profiles of putative *PaGUX* were compared to these responsible for the biosynthesis of other cell wall polysaccharides and lignin. The putative cellulose, mannan, xylan and lignin biosynthesis genes were identified using Congenie BLAST service using *A. thaliana* protein sequences as a query. Data visualisation was performed by an imaging script built into the NorWood tool.

***Pinus radiata* RNA isolation and cDNA synthesis**

Total RNA was extracted from 6 different tissue types (root, cambium/xylem, phloem, vegetative bud, needles and green stem) of 4 individual clonal *Pinus radiata* trees using the Plant RNA Isolation Mini Kit (Agilent, Santa Clara, CA). 50 mg of plant material from each tissue type were used for the RNA extractions. Concentration and purity of total RNA was determined fluorometrically (Qubit, Invitrogen) and the integrity of the RNA samples was verified using 1.2% agarose formaldehyde gels. Reverse transcription (RT) was carried out using the qScript XLT cDNA Super Mix Kit (Quantabio, Beverly, MA) with 500 ng of DNaseI-treated total RNA as a template.

ddPCR analysis of *PrGUX1* and *PrGUX2* expression

ddPCR was performed using the Bio-Rad QX200 system (Bio-Rad, Hercules, CA) according to the manufacturer's instructions. ddPCR reactions contained 11 µL of ddPCR EvaGreen Supermix (Bio-Rad), primers at a final concentration of 150 nM each, 2 µL of a 1:30 cDNA template dilution and ultrapure water to a final volume of 22 µL. Twenty µL of the ddPCR reactions along with 70 µL of droplet generation oil for EvaGreen (Bio-Rad) were loaded onto an 8-channel DG8™ cartridge for droplet generation in the QX200 Droplet Generator (Bio-Rad). Droplets were transferred into a 96-well plate (Bio-Rad) and a two-step thermocycling protocol [(enzyme activation at 95 °C × 5 min); 40 cycles × (denaturation at 95 °C × 30 sec, annealing/extension at 60 °C × 60 sec); signal stabilization at 4 °C × 5 min and 90 °C × 5 min, ramp rate set at 2 °C/sec] was performed in a thermal cycler (Bio-Rad C1000 Touch). Droplets were quantified in a QX200 Droplet Reader (Bio-Rad) and results were analysed with QuantaSoft™ Analysis Pro Software v.1.0.596. Statistical analysis was performed using R. Estimated marginal means were shown with bars representing the 95% confidence intervals. Four biological replicates were used for all tissue types

except for vegetative bud where only 3 were available. RNA-only controls (no RT) were performed to check for the presence of contaminating genomic DNA. No amplification was observed from these controls.

***N. benthamiana* expression and western blot analysis**

Blunt ended Myc tagged GUX and AtMAGT1 CDSs were amplified from a synthetic construct (synthesised by Genewiz) with Q5 DNA Polymerase (NEB) using primers detailed in Table S1. The PCR product was ligated into NruI (NEB) digested pEAQ-HT *Nicotiana benthamiana* overexpression vector (Sainsbury et al., 2009) using T4 DNA ligase (Thermo-Fisher Scientific). For overexpression in tobacco AGL-1 *Agrobacterium* was used and a previously described infiltration protocol was followed (Sparkes et al., 2006). The membrane collection protocol was adapted after Rennie et al. 2012 and performed as described in Lyczakowski et al., 2017. Protein concentration in the membranes fraction was quantified using modified Bradford reagent (Expedeon). For western blot analysis, each well of SDS-PAGE (10–15% gradient, BioRad) was loaded with 2.5 or 5 µg of GUX enriched *N. benthamiana* leaf membrane protein. Following the run, the gel was transferred onto nitrocellulose membrane using iBlot system (Life Technologies). The membrane was blocked overnight in 5% milk powder solution in TBS solution. The following day it was probed with 1:2000 anti-Myc primary antibody (rabbit polyclonal, Santa-Cruz, A14) and with 1:10000 mouse anti-rabbit HRP linked secondary antibody (Bio-Rad, 170-6515). Amersham ECL prime HRP substrate (GE-Lifesciences) was used to obtain signal from membrane bound antibodies.

***In vitro* activity assay for GUX**

Each reaction mix was prepared as described by Rennie et al., 2012 with the omission of UDP-Xylose. In brief, dried aliquots of acetylated heteroxylan (extracted following a protocol described by Busse-Wicher et al., 2014) were resuspended in 30 µL of a reaction master-mix (0.5 mM DTT, 10 mM MnCl₂, 10 mM MgCl₂, 2% Triton-X100, 10 mM UDP-GlcA). UDP-GlcA concentration was changed for some reactions or the sugar nucleotide was replaced with water in certain reactions to control for non-specific glucuronidation. The reaction mix was amended with 30 µL of undiluted microsomal proteins extracted from *N. benthamiana* leaves. For most experiments the reaction was performed for 5 h. at room temperature and terminated with a 10-

minute-long 100 °C heat treatment. In the case of reactions including acetylxyylan esterase (Acetylxyylan esterase CE4 from *Clostridium thermocellum*, NzyTech CZ00341) 5 µL of enzyme suspension was added to the reaction following 3 h from its start. Additional UDP-GlcA was introduced into the reaction upon addition of the esterase enzyme, leading to doubling of the initial concentration in the reaction tube. Reactions performed in the presence of the esterase enzyme were carried out overnight. Following completion of the reaction, the polysaccharides products were extracted using a Methanol: Chloroform lipid removal protocol adapted after (Bligh and Dyer, 1959) and described fully in Lyczakowski et al., 2017.

Carbohydrate digestions and PACE analysis

De-acetylated AIR (0.25 mg) or *in vitro* reaction products were digested with *Neocallimastix patriciarum* GH11 enzyme overnight at 30 °C by amending the suspension with 2 µL of enzyme stock (Megazyme). These conditions achieved complete digestion of accessible xylan to xylose, xylobiose and glucuronidated oligosaccharides. For GH30 digestion 1 µL of 1 mg/mL purified *Erwinia chrysanthemi* enzyme was applied onto *in vitro* reaction products for 1 hour at room temperature. EcGH30 used in this work was a kind gift from Dr Kristian Krogh (Novozymes A/S, Denmark). Fungal GH115 glucuronidase used in this work was also received from Dr Kristian Krogh. For de-glucuronidation, 1 µL of 1 mg/mL fungal GH115 was added to oligosaccharides re-suspended in 0.1 M ammonium acetate pH = 5.5 buffer and incubated overnight at 55 °C. Prior to GH115 treatment, the GH11 enzyme was heat inactivated by incubating the oligosaccharide sample at 120 °C for 10 minutes. Following digestion the oligosaccharides were derivatised with 8-aminonaphthalene-1,3,6-trisulphonic acid (ANTS; Invitrogen) and analysed with PACE as previously described (Mortimer et al., 2010). PACE band intensity analysis was performed with ImageJ.

Saccharification experiments

All saccharification experiments were performed on AIR using Cellic Ctec2® (Novozymes A/S, Denmark). For saccharification 0.5 mg of AIR was used and no pre-treatment was applied. The saccharification and sugar yield analysis were performed as previously described (Lyczakowski et al., 2017).

Statistical analysis and sampling

For all quantitative experiments, unless otherwise stated, three biological replicates of plant material were grown and analysed. Each biological replicate of AIR consisted of a pooled sample of 36 plants. For each biological replicate, 3 technical replicates were analysed for the quantitation of sugar release in saccharification. Statistical analysis was performed using packages available in the R software. Statistical tests used to compare average measurements for samples are mentioned in Figure legends. The variance between each pair compared was estimated to be similar with Levene's test.

Results:

Conifer genomes encode two distinct GUX homologues

In our previous work we identified a single conifer GUX enzyme clade and characterised the *in vitro* activity of its representative from *Picea glauca* (Lyczakowski et al., 2017). In dicots, GUX enzymes produce specific GlcA distribution patterns on xylan backbones (Bromley et al., 2013). Therefore, the recent detection of two distinct GlcA patterns on spruce xylan (Martinez-Abad et al., 2017) prompted us to reanalyse transcriptomic and genomic data available for coniferous plants. We investigated the presence of additional clades of GUX enzymes within this taxonomic group. We used Arabidopsis GUX1 and GUX2 sequences to search for putative gymnosperm xylan glucuronyltransferases. A highly conserved region, containing the GT8 domain, was selected in all gymnosperm protein sequences with high homology to Arabidopsis GUX1 or GUX2 and a maximum likelihood phylogenetic tree was constructed using these amino acid sequences (Figure 1). The phylogenetic analysis indicated that, similarly to eudicots, conifers have two distinct clades of putative GUX enzymes. Sequences from the previously characterised conifer clade (Lyczakowski et al., 2017) were observed to be more similar to eudicot GUX1 and GUX3 sequences. The newly identified clade of putative conifer GUX enzymes was more similar to eudicot GUX2 than to the other conifer clade. Therefore, in line with previous nomenclature, we labelled the conifer sequences as members of GUX1 and GUX2 clades.

Conifer GUX1 and GUX2 enzymes are likely to be involved in the wood formation process

To determine the possible biological function of GUX1 and GUX2 in conifers, we first analysed their expression profiles. We started with information available in public databases such as Congenie (Nystedt et al., 2013) and Norwood (Jokipii-Lukkari et al., 2017). The Congenie exHeatmap tool indicates that both *Picea abies* (*Pa*) GUX1 and *Pa*GUX2 mRNAs are enriched in phloem, cambium and xylem (Figure S1 and S2). Developing vegetative shoots also had high expression levels. The Norwood database provides a more specific analysis by allowing the mapping of conifer gene expression across different elements of the stem, from cambium to mature xylem vessels. Therefore, to further evaluate the likely biological role of putative conifer GUX enzymes, expression of *Pa*GUX1 and *Pa*GUX2, together with the expression of other putative cell wall biosynthesis genes, was studied using the Norwood tool (Figure S3). Most of the studied cell wall biosynthesis genes shared a similar high expression profile predominantly in the developing tracheids. This expression profile, which can be associated with gene's function in secondary cell wall biosynthesis, was also observed for both *Pa*GUX1 and *Pa*GUX2.

In order to validate our *in silico* results and to extend expression profiling to another conifer species, we have performed an absolute quantification of conifer GUX1 and GUX2 expression in four individual clonal *Pinus radiata* (*Pr*) trees using droplet digital (dd) PCR (Figure 2). We studied the expression of the genes in xylem/cambium, phloem, root, vegetative bud, green stem and needles. Expression of both *Pr*GUX1 and *Pr*GUX2 was detected in all tissues tested, with the exception of phloem which had only trace amounts of *Pr*GUX2 mRNA. In line with the *in silico* analysis for *Picea abies*, we have detected strong *Pr*GUX1 and *Pr*GUX2 expression in green stems and needles. The strongest expression detected for both *Pr*GUX1 and *Pr*GUX2 was in xylem/cambium. In xylem, as in all tissues tested, the expression of *Pr*GUX1 was about six to eleven times higher than that of *Pr*GUX2. This suggests that *Pr*GUX1 may be responsible for the glucuronidation of the majority of the xylan in this species. In summary, together with the Congenie and Norwood based analyses, our absolute quantification results support the hypothesis

that putative conifer GUX enzymes are strongly expressed in developing xylem and, therefore, are likely to be involved in the conifer secondary cell wall biosynthesis process. Importantly, similar *GUX* expression profiles were obtained for both *Picea abies* and *Pinus radiata*, suggesting that the two GUX enzyme clades may contribute to xylan formation across many softwood taxa. This hypothesis is further supported by a recent report indicating that a GUX homologue in *Pinus densiflora* is expressed more strongly in xylem and in cambium than in other conifer tissues (Kim et al., 2021).

Conifer GUX1 and GUX2 are xylan glucuronyltransferases with distinct specificities

We previously demonstrated that *PgGUX1* is an active xylan glucuronyltransferase by using transient expression in *N. bethamiana* for an *in vitro* activity assay (Lyczakowski et al., 2017). To extend this analysis and study conifer GUX2 clade representatives, we expressed myc-tagged *Pinus taeda* (*Pt*) GUX1 and *PtGUX2* enzymes and *Picea sitchensis* (*Ps*) GUX2 enzyme (Figure S4). We selected *P. taeda* GUX1 and GUX2 as their protein sequences are identical to those in *P. radiata* (Figure S5). Therefore, the GUXs enzymatic activity is likely to be identical in these two species. *PgGUX1* was used as a positive control. An Arabidopsis glucomannan galactosyltransferase *AtMAGT1* was used as a negative control to monitor for tobacco's potential endogenous glucuronyltransferase activity. Contrary to hardwood, softwood's xylan is not acetylated. However, xylan lacking acetylation and GlcA side branches is poorly soluble. Therefore, to maintain the acceptor solubility during the *in vitro* reaction, we have used acetylated xylan extracted from the Arabidopsis *gux1/2* mutant. Membrane preparations enriched with the myc-tagged proteins of interest were incubated with acetylated polymeric xylan lacking [Me]GlcA branches in the presence or absence of UDP-GlcA. Following incubation, the reaction products were extracted, deacetylated, digested with a GH11 xylanase and analysed with PACE (Figure 3a). In the tobacco system, no xylan glucuronidation background activity was detected as the reaction products coming from the incubation in the presence of *AtMAGT1* were fully converted to xylose and xylobiose. A GlcA-xylotetraose (XUXX) product was observed in all reactions in which putative

conifer GUX enzymes were incubated in the presence of the acceptor and UDP-GlcA. This demonstrates that conifer GUX1 and GUX2 enzymes were able to add glucuronic acid onto an acetylated xylan backbone *in vitro*.

We noticed that the reaction catalysed by both *PtGUX2* and *PsGUX2* enzymes produced an additional product migrating above XUXX. This product was not detected in GUX1 catalysed reactions. Therefore, the activity enabling its formation was specific to representatives of the conifer GUX2 clade. To characterise this novel structure, GH11 products were further digested with α -glucuronidase GH115 and the resulting oligosaccharides analysed by PACE (Figure 3b). As expected, the removal of GlcA from the XUXX structure resulted in its co-migration with the xyloetraose (X₄) standard. The GH115 treatment of the uncharacterised oligosaccharide altered its migration profile to match the xylopentaose (X₅) standard, indicating that the structure has five xyloses on its backbone. GH11 needs two unsubstituted xyloses on the -1 and -2 sites and one unsubstituted xylose on the +1 site to digest xylan (Vardakou et al., 2008, Paes et al., 2012). Therefore, oligosaccharides produced by GH11 xylanases have one unsubstituted xylosyl residue at the non-reducing end, and two unsubstituted xylosyl residues at the reducing end. Consequently, a GH11 released oligosaccharide with five xylosyl monomers in the backbone must contain two branches on positions -3 and -4 from the GH11 restriction site (Figure 3c). Therefore, the structure of the unknown oligosaccharide must correspond to XUUXX. The identity of the oligosaccharide is supported by its PACE migration profile. The oligosaccharide migrates between X₄ and X₃ oligosaccharide standards, whereas a singly glucuronidated xylopentaose migrates between the X₅ and X₄ standards (Bromley et al., 2013). The increase in the extent of PACE migration originates from the presence of two negatively charged GlcA molecules on the XUUXX backbone. This supports our conclusion that conifer GUX2 enzymes are likely to generate an oligosaccharide with GlcA present on consecutive xyloses. The same structure was already reported in softwood (Shimizu et al., 1978, Yamasaki et al., 2011, Martinez-Abad et al., 2017).

Conifer GUX activity is stimulated and modified by xylan deacetylation

As softwood xylan is not acetylated, we wanted to investigate the effect of removing xylan acetyl groups on the activity of conifer GUX1 and GUX2 enzymes. To investigate the impact of xylan acetylation we have performed an *in vitro* GlcAT reaction to which we have added CE4, a xylan acetyl esterase enzyme. The CE4 esterase removes single 2 or 3 linked acetyl groups from xylose monomers which do not carry GlcA decorations (Taylor et al., 2006). The enzyme was added 3 hours after the start of the *in vitro* glucuronidation reaction to enable some GlcA to be transferred to avoid the acceptor precipitation. Following an overnight incubation of the *gux1/2* xylan acceptor with *Pg*GUX1 and *Pt*GUX2 enzymes in the presence and absence of CE4 esterase the reaction products were extracted, digested with GH11 and analysed by PACE (Figure 4a).

GH11 cleavage of all reactions catalysed by *Pg*GUX1 and *Pt*GUX2 in the presence of UDP-GlcA released xylose, xylobiose and XUXX. Independantly of the esterase presence, reactions catalysed by *Pg*GUX1 produced an additional band migrating between xylopentaose and xylotetraose standards. However, the intensity of this novel band was increased in the reaction that included the esterase. To investigate the identity of this band, oligosaccharides released by the xylanase GH11 were further digested with alpha-glucuronidase GH115 and analysed by PACE (Figure S6a). Following the GH115 treatment xylotetraose and xylohexaose were detected. The xylotetraose must have originated from XUXX de-glucuronidation. Therefore, the novel oligosaccharide had six xyloses in the backbone. Since the GH11 enzyme requires two unsubstituted xyloses at the -1 and -2 sites and one unsubstituted xylose at the +1 site (Vardakou et al., 2008, Paes et al., 2012), an oligosaccharide with six xylosyl monomers in the backbone is likely to contain two GlcA branches on positions -3 and -5 from the GH11 cleavage sites (Figure S6b). The presence of this structure is consistent with GH11 specificity, as the enzyme can not degrade oligosaccharides in which only a single unbranched xylose separates two branched backbone subunits. The identity of the oligosaccharide is also supported by its PACE migration profile between X_5 and X_4 standards, as singly glucuronidated xylohexaose would migrate between X_6 and X_5 oligosaccharides (Bromley et al., 2013). Taking all evidence into consideration, the oligosaccharide is highly likely to be XUXUXX. The reaction catalysed by *Pt*GUX2 produced XUXX and XUUXX in similar proportion in the absence of acetylxylan esterase, while in the presence of the esterase XUUXX

became the dominant glucuronidated product (Figure 4b). Having established the likely identity of all glucuronidated structures observed in the PACE experiment (Figure 4a) we were able to compare the efficiency of *in vitro* glucuronidation in the presence and absence of xylan acetyltransferase (Figure 4b). Our analysis indicates that the addition of the esterase increases the efficiency and alters the specific glucuronidation of *PgGUX1* and *PtGUX2*. It is possible that using an acceptor with reduced acetylation allows the conifer GUX enzymes to exert *in vitro* glucuronyltransferase activities more similar to their *in planta* function.

Softwood GUX1 and GUX2 branching patterns are different and dependant on the degree of acetylation of the xylan acceptor

As the acetylation level of the acceptor influences the specific *in vitro* glucuronyltransferase activity of *PgGUX1* and *PtGUX2*, we decided to evaluate its impact on the GlcA patterning. For this purpose we analysed the GlcA decoration pattern of the xylan products generated during *in vitro* reactions by *PgGUX1* and *PtGUX2* in the presence or absence of the xylan acetyltransferase. The products were deacetylated and digested with xylanase GH30, which enables analysis of GlcA patterning on xylan (Figure 5a). The released oligosaccharides were derivatised with the ANTS fluorophore and analysed by PACE (Figure 5b). In addition to the xylo-oligosaccharide standard, wild type (WT) *A. thaliana* AIR was digested with GH30 to produce known glucuronidated xylooligosaccharide markers.

For *PgGUX1*, in the absence of xylan acetyltransferase, the most frequent spacing of GlcA was equal to 6 xylose units. Oligosaccharides with a longer degree of polymerisation (DP) were also released. The pattern changed substantially upon the addition of xylan acetyltransferase to the reaction. Partial deacetylation of the acceptor resulted in *PgGUX1* generating a denser GlcA patterning *in vitro* with spacings of 6, 4 and 2 xylosyl units preferred. For *PtGUX2*, in the absence of the xylan acetyltransferase, GH30 digestion of the glucuronoxylan released a mixture of single GlcA branched oligosaccharides with a second band migrating close to it. The GH11 digestion of the products of the same reaction resulted in production of singly and doubly glucuronidated oligosaccharides (Figure 4a). Therefore, the GH30-produced oligosaccharides migrating close to known digestion products might be carrying two

GlcA on consecutive xyloses. This structure is unlikely to be digested further by GH30 due to the steric hindrance of the two consecutive decorations. The addition of xylan acetyltransferase to *PtGUX2* catalysed glucuronidation changed the GH30 digestion pattern. The released oligosaccharides were shorter and almost all of them migrated like putative oligosaccharides bearing two GlcA branches. This alteration in the ratio of doubly and singly glucuronidated oligosaccharides, is in line with the GH11 digestion results (Figure 4a).

Together with the GH11 data (Figure 4a), the results of the GH30 analysis provide further evidence that the pattern of GlcA branches added by both *PgGUX1* and *PtGUX2* is sensitive to the degree of acetylation of the xylan acceptor. For both enzymes, partial removal of acetyl branches resulted in generation of more densely glucuronidated xylan molecules. This is indicated by both a significant increase in the degree of glucuronidation in the GH11 analysis, and the release of shorter oligosaccharides after the GH30 digestion. Moreover, the addition of the acetyltransferase enabled *PtGUX2* to add GlcA to neighbouring xylose units with higher efficiency, suggesting this might be a preferred mode of glucuronidation for this enzyme.

Conifer GUX enzymes are able to glucuronidate xylan *in planta*

We wanted to observe the *in planta* glucuronyltransferase activity of enzymes within the softwood putative GUX2 clade. For this purpose, we expressed *PtGUX2* under the control of *pIRX3*, a strong secondary cell wall-specific Arabidopsis promoter, in a *gux1/2/3* Arabidopsis background. The *gux1/2/3* plants lack any detectable xylan glucuronidation (Mortimer et al., 2015). Therefore, any GlcA detected on the xylan of *gux1/2/3* plants would be introduced by *PtGUX2*. Two individual lines were obtained (*PtGUX2.1* and *PtGUX2.2*) and their xylan structure was analysed by PACE after xylanase GH11 digestion (Figure 6a). The *PtGUX2* expression led to restoration of GlcA side chains on *gux1/2/3* xylan. The total degree of xylan glucuronidation in the transgenic plants exceeded the one observed in WT Arabidopsis (Figure 6b). Moreover, for both transgenic lines the XUXX and the XUUXX oligosaccharides were detected after GH11 digestion of xylan. This further confirmed that the clade 2 conifer GUX enzymes are capable of glucuronidating consecutive xylosyl units.

We then investigated the effect of xylan acetylation on the *in planta* activity of conifer GUX enzymes. Thus, we have generated *A. thaliana esk-kak* mutant plants expressing *PgGUX1* or *PtGUX2* (lines *Pg1esk-kak* and *Pt2esk-kak*, respectively) under the control of *IRX3* promoter. The *esk-kak* plants have an approximate 50% reduction in xylan acetylation levels (Xiong et al., 2013, Grantham et al., 2017). Therefore, in this background, the activity of conifer GUX enzymes may be higher than within the *gux1/2/3* plants where the xylan is fully acetylated. However, in the *Arabidopsis esk-kak* mutant background the native GUX enzymes contribute to glucuronidation. AIR isolated from basal stem of the *Pg1esk-kak* and *Pt2esk-kak* plants was digested with xylanase GH11 and the resulting oligosaccharides were separated by PACE (Figure 6c). The detected glucuronidated structures were primarily XUXUXX and XUUXX in *esk-kak* plants overexpressing *PgGUX1* and *PtGUX2*, respectively. Interestingly, when over-expressed in the *esk-kak* background, the *PtGUX2* enzyme also generated structures resulting in XUXUXX formation by GH11. In *esk-kak* plants overexpressing the *PgGUX1* and *PtGUX2* enzymes the total degree of glucuronidation was 20% higher than the WT (Figure 6d). This is more than what we observed for *PgGUX1* (Lyczakowski et al., 2017) or *PtGUX2* expressed in *gux1/2/3* plants. This observation strengthens the hypothesis that a low degree of xylan acetylation is favourable to conifer GUX activity.

GlcA pattern introduced by *PtGUX2* is able to confer biomass recalcitrance in the *Arabidopsis* model

The absence of [Me]GlcA branches on xylan of *gux1/2/3* *Arabidopsis* is associated with a significant reduction in biomass recalcitrance to enzymatic digestion (Lyczakowski et al., 2017). Therefore, to investigate if the GlcA branches introduced by the *PtGUX2* enzyme are capable of rescuing the mutant phenotype we have performed a saccharification assay on cell wall material isolated from WT, *gux1/2/3* and *gux1/2/3* plants expressing *PtGUX2*. In the *gux1/2/3* plants expressing *PtGUX2* the release of both glucose (Figure 7a) and xylose (Figure 7b) was reduced to that measured for WT *Arabidopsis*. This suggests that a softwood-like pattern of GlcA branches, introduced by the *PtGUX2* enzyme, can participate in the maintenance of biomass recalcitrance.

Discussion:

To progress towards a more sustainable economy, it is important to optimise the use of lignocellulosic biomass for various purposes. Conifer wood, accounts for a significant proportion of lignocellulosic biomass, especially in the Northern boreal forests. Despite the importance of softwood, very little is known about the molecular basis of its biosynthesis. In this work we investigated the gene expression profiles and biochemical activities of softwood xylan glucuronyltransferases. We found two distinct softwood clades of GUX enzymes expressed during secondary cell wall synthesis. Each clade produces a specific glucuronic acid substitution pattern. The conifer *GUX* genes could be important breeding or engineering targets for modification of softwood biomass, particularly to modify the digestibility of softwood.

In our previous work, we discovered a single clade of conifer GUX enzymes (Lyczakowski et al., 2017). Here, we re-analysed transcriptomic and genomic data of conifer species, this led to the discovery of a second clade of conifer GUX enzymes. Our phylogenetic analysis (Figure 1) indicates, that, while the previously characterised conifer GUX clade is more homologous to hardwood GUX1/3 sequences, the newly identified sequences are more homologous to the hardwood GUX2 clade. This suggests a potential functional divergence of conifer GUX genes, similar to that seen in dicot species (Bromley et al., 2013). Interestingly, conifer GUX sequences share an exceptionally high degree of sequence similarity within each clade (Figure S5). This is in line with previously published data indicating that the majority of gymnosperm evolution happened prior to the lineage split with angiosperms and that after this event the pace of genome evolution has been slower in the gymnosperm lineage than in the angiosperm one (Pavy et al., 2012, Buschiazzi et al., 2012). Therefore, it is likely that the functional characterisation performed in this work may be an accurate representation of GUX activities for the entire taxon. With that in mind, we propose a nomenclature for conifer GUX clades which is in line with their similarity to eudicot GUX1/3 or GUX2 sequences.

In our work, we were able to demonstrate, using both *in vitro* assays and *in vivo* complementation, that some conifer genomes encode a minimum of two distinct active xylan glucuronyltransferases. Importantly, our results indicate that the two

conifer GUX enzymes may be responsible for the synthesis of different patterns of GlcA branches. This conclusion is supported by the results of our *in vitro* GUX activity experiments. In particular, by decreasing the acetylation level of the *in vitro* xylan acceptor it was possible to shift conifer GUX activities towards generating softwood specific types of patterning. As the conifer AGX is not acetylated (Scheller and Ulvskov, 2010) it is possible that these low acetylation conditions are more similar to the *in planta* environment. When acting on the xylan acceptor with enzyme-reduced acetylation, clade 1 enzymes were observed to add GlcA predominantly onto every 6th, 4th and 2nd xylosyl unit. In the same conditions, clade 2 enzymes added GlcA mainly onto consecutive xylosyl units. The two adjacent xylose backbone units, each bearing a GlcA branch, are mainly interspaced by 4, 5 or 6 unsubstituted xylosyl units. Softwood xylan is decorated with arabinose residues. However, these decorations are absent in our *in vitro* and *in planta* acceptors. This difference is probably influencing the specific activity of conifer GUX. Therefore, the relationship between the patterning of glucuronic acid and arabinose substitutions in conifers remains to be investigated.

Our results suggest that the two distinct clades of GUX enzymes in conifers are responsible for the two patterns of MeGlcA branches observed in softwood. The conifer clade 1 enzymes could be responsible for the synthesis of the main xylan pattern in softwood (Busse-Wicher et al., 2016b), with a spacing of six xylose units between the GlcA branches. Clade 2 of conifer GUX enzymes could be responsible for addition of GlcA branches onto two consecutive xylose residues, forming the structure described by Martinez-Abad et al.,. To confirm this hypothesis, we are currently analysing transgenic conifer plants lacking GUX1 or GUX2 activity.

Patterning of GlcA branches is thought to be a determinant of the compatibility of xylan molecules for their interaction with the hydrophilic face of the cellulose fibril (Grantham et al., 2017). In *A. thaliana*, the AtGUX1 enzyme is responsible for the addition of evenly spaced GlcA branches whereas the AtGUX2 enzyme adds GlcA branches without this preference (Bromley et al., 2013). Therefore, with xylan in a two-fold screw conformation, AtGUX1s produce a pattern with all the GlcA facing the same side of the backbone. The other side of xylan, free of decorations, is therefore compatible with the cellulose fibril and can adsorb onto it (Busse-Wicher et al., 2016a). AtGUX2 places GlcA decorations on both side of the two-fold screw xylan.

This pattern, known as incompatible, is likely to alter the interaction between xylan and the hydrophilic surface of the cellulose microfibril.

Similarly, to the *A. thaliana* model, both compatible and incompatible GlcA patterns may be present in softwood. The majority of softwood AGX has the MeGlcA branch on every 6th monomer (Busse-Wicher et al., 2016b) and as such can be considered compatible with adsorption onto the hydrophilic surface of the cellulose microfibril. A minor proportion of softwood AGX has GlcA branches on two consecutive monomers (Martinez-Abad et al., 2017). According to the current xylan-cellulose interaction model this structure could have a reduced capacity to adsorb onto the hydrophilic face of the cellulose fibril (Busse-Wicher et al., 2016a, Grantham et al., 2017). However, molecular dynamics simulations performed by Martinez-Abad et al., indicate that xylan with GlcA on consecutive xylosyl monomers may also interact with certain parts of the hydrophilic face of the cellulose microfibril. Convergently, recently published ssNMR experiments on spruce indicate that both two-fold, cellulose bound, and three-fold, soluble, xylan structures are present in softwood (Terrett et al., 2019). The cellulose face(s) upon which the xylan binds remains to be discovered but it is likely that at least some of the softwood AGX has a structure which is incompatible with binding the hydrophilic surface of the cellulose microfibril. Noticeably, the extractability of spruce xylan with consecutive MeGlcA branches is lower than that of xylan with evenly spaced MeGlcA branches (Martínez-Abad et al., 2020). This extractability difference might originate from xylan's interaction capacity with lignin, which can be regulated partially by its glucuronidation (Martínez-Abad et al., 2020). When investigating softwood properties it is also important to account for the role of the predominant hemicellulose - the GGM. Similar to AGX, GGM can bind to cellulose (Terrett et al., 2019) and lignin (Nishimura et al., 2018). Like AGX, GGM also shows diverse extractability patterns (Martínez-Abad et al., 2020). It is therefore likely that the potential interactions between AGX and GGM may also contribute to conifer cell wall properties. This intra-hemicellulose interaction might also be influenced by AGX and GGM decoration patterns.

Xylan glucuronidation, in addition to determining the capacity of the hemicellulose to interact with the hydrophilic surface of the cellulose microfibril, is critical for the maintenance of plant biomass recalcitrance to enzymatic digestion (Lyczakowski et al., 2017). Importantly, our previous work suggested that reducing the amount of

[Me]GlcA branching by 30% had no effect on the cell wall recalcitrance (Lyczakowski et al., 2017). In *gux1/2/3* plants overexpressing the PtGUX2 enzyme, we observed a 20% increase in [Me]GlcA levels compared to WT. However, this did not result in any further increase in biomass recalcitrance. These observations are consistent with the proposed role of GlcA in the formation of ester linkages with lignin (Watanabe and Koshijima, 1988). In this model, a small number of linkages might maintain full xylan-lignin cross-linking (Giummarella and Lawoko, 2016). Therefore, a partial change in the content of [Me]GlcA branches does not result in recalcitrance reduction or further resistance gain to enzymatic digestion. This may be important for conifer GUX mutagenesis experiments aiming to generate softwood material with reduced xylan glucuronidation and possibly with reduced wood recalcitrance. Since the absence of xylan glucuronidation is required to decrease biomass recalcitrance, homologues of both conifer GUX may need to be simultaneously targeted for mutagenesis to achieve a significant reduction in softwood recalcitrance. However, even small incremental improvements in saccharification or pulping efficiency provided by a single GUX mutant could be relevant in the context of industrial processes. For example, the reduction of GlcA during the kraft pulping process would reduce the formation of undesired chemical by-products such as hexenuronic acid which leads to the creation of hazardous organic halogens (Danielsson et al., 2006).

In summary, our work identifies and describes the function of two clades of conifer GUX enzymes. We were able to demonstrate that clade 1 enzymes are likely to generate compatible pattern of GlcA branches and that clade 2 enzymes have the unique capacity of adding GlcA branches onto two consecutive xylose monomers. Since xylan glucuronidation is important for the maintenance of biomass recalcitrance, our work may guide mutagenesis or breeding approaches aiming to generate softwood biomass which can be more easily converted into pulp or biofuels. Moreover, as xylan-celulose interaction may influence the mechanical properties of secondary cell walls (Simmons et al., 2016, Grantham et al., 2017) modulating the relative contribution clade 1 and clade 2 GUX activity to the process of xylan glucuronidation in softwood may prove a valid strategy to modify wood strength. As such, our work may be applied for the modification of softwood material

to make it more suitable for applications in sustainable industries such as pulping, biofuels or wood-based construction.

Main text figure legends:

Figure 1: Phylogenetic analysis of dicot and gymnosperm GUX enzymes. Analysis based on amino acid sequences of GT8 domains. Hardwood and softwood sequences are shown in green and blue, respectively. Specific clades are annotated on the phylogenetic tree. Putative GUX of the moss *Physcomitrella patens*, shown in orange, was used as an outgroup.

Figure 2: Absolute quantification of *P. radiata* GUX1 and GUX2 expression using ddPCR. Concentrations are presented as copies per μL when an equal amount of cDNA was used for each of 6 tissue types. Bars represent estimated marginal means $\pm$ 95% confidence intervals.

Figure 3: *In vitro* activity of two conifer GUX enzymes. Results of GH11 digestion of *in vitro* activity products of glucuronidation reaction catalysed by *Pinus taeda* (Pt) GUX1 and GUX2, *Picea glauca* (Pg) GUX1 and *Picea sitchensis* (Ps) GUX2. Glucuronidated oligosaccharides are marked. Control reaction, performed using membranes enriched for Arabidopsis MAGT1 did not release any glucuronidated oligosaccharides. B) GH115 treatment of PtGUX2 *in vitro* reaction products digested with GH11. Removal of glucuronic acid from the structure unique to clade 2 conifer GUX shifts it to the X5 position. C) Diagram presenting a GH11 digestion profile expected from a polysaccharide containing some GlcA on consecutive xylosyl units. In C) and red arrow shows GH11 digestion points.

Figure 4: Reduction in xylan acetylation increases the activity of conifer GUX enzymes *in vitro*. A) PACE of a GH11 digest of PgGUX1 and PtGUX2 *in vitro* glucuronidation products obtained with and without xylan acetyl esterase (E) in the reaction mix. Arabidopsis MAGT1 is used as a negative control. B) Quantification of the degree of glucuronidation obtained for the reaction catalysed by PgGUX1 and PtGUX2 in the absence and presence of the xylan acetyl esterase in the reaction mix. Three replicates of the reaction were performed and analysed.

Figure 5: Supplementation with a xylan acetyl-esterase changes the pattern of GlcA branches introduced by conifer GUX enzymes *in vitro*. A) Diagram demonstrating GH30 activity on xylan with individual or consecutive GlcA branches. Red arrows indicate digestion points by GH30. DP is degree of polymerisation. B) PACE of a GH30 digest of PgGUX1 and PtGUX2 *in vitro* activity products obtained in reactions performed with and without the xylan acetyl esterase. WT *A. thaliana* biomass was digested to obtain a standard of singly glucuronidated oligosaccharides (marked on the left edge of the gel, UX on the reducing end). Predicted annotation of putative doubly glucuronidated oligosaccharides (marked on the right edge of the gel, UUX on the reducing end) is proposed based on the identity of the most proximal singly glucuronidated oligosaccharide. Change in band migration is likely to originate from increase in charge (positive effect on band migration) and mass (negative effect on band migration) of the oligosaccharide due to presence of second GlcA branch. PtGUX2 reaction performed in the absence of the esterase enzyme releases a mix of singly and doubly glucuronidated structures.

Figure 6: Analysis of transgenic Arabidopsis plants over-expressing the PtGUX2 enzyme under the control of a secondary cell wall specific *IRX3* promoter. A) PACE of a GH11 digest of Arabidopsis WT, *gux1/2/3* and two independent *gux1/2/3* lines overexpressing the PtGUX2 enzyme. X1 to X6 marker (M) is used B) Quantification of the degree of glucuronidation in WT and transgenic lines. In line with previous reports (Mortimer et al., 2015 and Lyczakowski et al., 2017) *gux1/2/3* plants have no detectable (nd) GlcA on xylan. C) PACE analysis of xylan structure in WT, *esk-kak* and *PgGUX1* (lines *Pg1esk-kak*) and *PtGUX2* (lines *Pt2esk-kak*) expressing *esk-kak* using xylanase GH11. D) Quantification of the degree of glucuronidation obtained for WT, *esk-kak* and *PgGUX1* and *PtGUX2* expressing *esk-kak*. Three biological replicates of biomass were analysed. Grey coloration of the bar indicates the GlcA contributed by the XUXX structure. Black bar denotes GlcA contributed by the XUUXX and XUXUXX structures. Error bars represent standard deviation. Student's t-test was used to evaluate significance of the difference between the degree of glucuronidation, **p value ≤ 0.01 ; ***p value ≤ 0.001 .

Figure 7: Impact of *PtGUX2* expression on recalcitrance of *gux1/2/3 A. thaliana* biomass. Saccharification of AIR from Arabidopsis WT, *gux1/2/3* and *PtGUX2* lines.

Glucose (A) and xylose (B) release from AIR. Light grey bars denote datasets which are significantly different from the WT. For all graphs error bars represent standard deviation. Student's t-test was used to evaluate significance of the difference between WT and transgenic lines, *p value ≤ 0.05 ; **p value ≤ 0.01 ; ***p value ≤ 0.001 .

Supporting information figure legends:

Table S1: Primers used in this study.

Table S2: Database codes for gymnosperm GUX sequences used in this study.

Figure S1: Expression profiling for *Picea abies* (Pa) *GUX1* using exHeatmap tool. Legend explaining the colour-coding is provided.

Figure S2: Expression profiling for *Picea abies* (Pa) *GUX2* using exHeatmap tool. Legend explaining the colour-coding is provided.

Figure S3: Expression profiles of conifer genes likely to be involved in softwood biosynthesis using the Norwood tool. Expression profiling of *Picea abies* genes likely to be involved in the biosynthesis of cellulose, galactoglucomannan (GGM), xylan and lignin across cambium (C), secondary cell wall formation zone (SCW) and mature xylem/latewood (MX). Read enrichment is marked with red and read depletion is marked with blue. Most *P. abies* candidates (column 1 of the table) were identified based on having the highest homology score to known *A. thaliana* enzymes (column 2). Due to their potential role in xylan branching two representatives were selected for putative xylan glucuronyltransferases (GUX, marked in red) and putative arabinosyltransferases (GT61). Analysis reports data for three trees.

Figure S4: Western blot analysis confirming the successful expression of all four conifer GUXs analysed in the *in vitro* assay.

Figure S5: Sequence similarity matrix generated based on whole length sequences of conifer GUX1 and GUX2 enzymes analysed in this study.

Figure S6: Annotation of the XUXUX structure. A) GH115 treatment of PgGUX1 *in vitro* reaction products, obtained in the presence of xylan acetyl esterase, digested

with GH11. Removal of glucuronic acid from the structure enriched in conifer GUX1 glucuronidation events performed in the presence of xylan acetyl esterase shifts it to X₆ position. B) Diagram presenting a GH11 digestion profile expected from a polysaccharide containing some GlcA interspaced by just one xylosyl unit. In B) and red arrow shows GH11 digestion points, and red cross indicates that GH11 is unable to digest xylan in a specific position leading to XUXUXX formation.

Author contributions:

JJL designed the study, performed most polysaccharide analyses and genetic engineering, and co-wrote the paper; LY contributed to biochemical analyses, assignment of new polysaccharide structures and manuscript writing; OMT supported the biochemical analysis, isolation of stable transgenic Arabidopsis lines and co-wrote the paper; CF isolated RNA from radiata pine and performed RT-PCR experiments profiling GUX expression; HT contributed to assignment of new polysaccharide structures and isolation of stable transgenic Arabidopsis lines; GT was involved in the design of RT-PCR experiments and manuscript preparation; MS designed parts of the study and co-wrote the manuscript; PD designed the study, interpreted data and co-wrote the manuscript.

Acknowledgements:

We would like to acknowledge Prof. George Lomonossoff (John Innes Centre, UK), who developed the pEAQ-HyperTrans expression system used in this study. Plant Bioscience Limited supplied the pEAQ-HT vector which was used in this work. Enzymes used in this study (GH115 and EcGH30) were a kind gift from Dr Kristian Krogh (Novozymes A/S, Denmark). Dr Krogh also acted as a supporting supervisor during OMT doctoral training at The University of Cambridge.

Funding:

This work was supported by the Leverhulme Trust Centre for Natural Material Innovation and by Biotechnology and Biological Sciences Research Council

(BBSRC) of the UK as part of the OpenPlant Synthetic Biology Research Centre (Reference BB/L014130/1), Cambridge BBSRC-DTP Programme (Reference BB/J014540/1) and iCASE studentship (Reference BB/M015432/1). Work was also supported by a grant from the National Science Centre Poland awarded to JJL as part of the SONATINA 3 programme (project number 2019/32/C/NZ3/00392). We acknowledge the funding received from the New Zealand Ministry of Business, Innovation and Employment (MBIE) Endeavour Fund (Contract No. C04X1707, Fibre Grand Design) for supporting this work.

Conflict of interest: The authors declare no conflict of interest.

Material availability statement: All plant material described in this manuscript is available on a reasonable request from the corresponding authors.

References:

- BAR-ON, Y. M., PHILLIPS, R. & MILO, R. 2018. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 6506-6511.
- BERGLUND, J., KISHANI, S., DE CARVALHO, D. M., LAWOKO, M., WOHLERT, J., HENRIKSSON, G., LINDSTROM, M. E., WAGBERG, L. & VILAPLANA, F. 2020a. Acetylation and Sugar Composition Influence the (In)Solubility of Plant beta-Mannans and Their Interaction with Cellulose Surfaces. *Acs Sustainable Chemistry & Engineering*, 8, 10027-10040.
- BERGLUND, J., MIKKELSEN, D., FLANAGAN, B. M., DHITAL, S., GAUNITZ, S., HENRIKSSON, G., LINDSTRÖM, M. E., YAKUBOV, G. E., GIDLEY, M. J. & VILAPLANA, F. 2020b. Wood hemicelluloses exert distinct biomechanical contributions to cellulose fibrillar networks. *Nature Communications*, 11.
- BLIGH, E. G. & DYER, W. J. 1959. A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION. *Canadian Journal of Biochemistry and Physiology*, 37, 911-917.
- BROMLEY, J. R., BUSSE-WICHER, M., TRYFONA, T., MORTIMER, J. C., ZHANG, Z., BROWN, D. M. & DUPREE, P. 2013. GUX1 and GUX2 glucuronyltransferases decorate distinct domains of glucuronoxylan with different substitution patterns. *Plant Journal*, 74, 423-434.
- BUSCHIAZZO, E., RITLAND, C., BOHLMANN, J. & RITLAND, K. 2012. Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *Bmc Evolutionary Biology*, 12.
- BUSSE-WICHER, M., GOMES, T. C. F., TRYFONA, T., NIKOLOVSKI, N., STOTT, K., GRANTHAM, N. J., BOLAM, D. N., SKAF, M. S. & 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*. *Plant Journal*, 79, 492-506.
- BUSSE-WICHER, M., GRANTHAM, N. J., LYCZAKOWSKI, J. J., NIKOLOVSKI, N. & DUPREE, P. 2016a. Xylan decoration patterns and the plant secondary cell wall molecular architecture. *Biochemical Society Transactions*, 44, 74-78.

- BUSSE-WICHER, M., LI, A., SILVEIRA, R. L., PEREIRA, C. S., TRYFONA, T., GOMES, T. C. F., SKAF, M. S. & DUPREE, P. 2016b. Evolution of Xylan Substitution Patterns in Gymnosperms and Angiosperms: Implications for Xylan Interaction with Cellulose. *Plant Physiology*, 171, 2418-2431.
- CLOUGH, S. J. & BENT, A. F. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, 16, 735-743.
- DANIELSSON, S., KISARA, K. & LINDSTROM, M. E. 2006. Kinetic study of hexenuronic and methylglucuronic acid reactions in pulp and in dissolved xylan during kraft pulping of hardwood. *Industrial & Engineering Chemistry Research*, 45, 2174-2178.
- FERNANDES, A. N., THOMAS, L. H., ALTANER, C. M., CALLOW, P., FORSYTH, V. T., APPERLEY, D. C., KENNEDY, C. J. & JARVIS, M. C. 2011. Nanostructure of cellulose microfibrils in spruce wood. *Proceedings of the National Academy of Sciences of the United States of America*, 108, E1195-E1203.
- GIUMMARELLA, N. & LAWOKO, M. 2016. Structural Basis for the Formation and Regulation of Lignin-Xylan Bonds in Birch. *Acs Sustainable Chemistry & Engineering*, 4, 5319-5326.
- GRANTHAM, N. J., WURMAN-RODRICH, J., TERRETT, O. M., LYCZAKOWSKI, J. J., STOTT, K., IUGA, D., SIMMONS, T. J., DURAND-TARDIF, M., BROWN, S. P., DUPREE, R., BUSSE-WICHER, M. & DUPREE, P. 2017. An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. *Nature Plants*, 3, 859-865.
- JARVIS, M. C. 2018. Structure of native cellulose microfibrils, the starting point for nanocellulose manufacture. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 376.
- JOKIPII-LUKKARI, S., SUNDELL, D., NILSSON, O., HVIDSTEN, T. R., STREET, N. R. & TUOMINEN, H. 2017. NorWood: a gene expression resource for evo-devo studies of conifer wood development. *New Phytologist*, 216, 482-494.
- KIM, M.-H., TRAN, T. N. A., CHO, J.-S., PARK, E.-J., LEE, H., KIM, D.-G., HWANG, S. & KO, J.-H. 2021. Wood transcriptome analysis of *Pinus densiflora* identifies genes critical for secondary cell wall formation and NAC transcription factors involved in tracheid formation. *Tree Physiology*, tpab001.
- LYCZAKOWSKI, J. J., WICHER, K. B., TERRETT, O. M., FARIA-BLANC, N., YU, X. L., BROWN, D., KROGH, K., DUPREE, P. & BUSSE-WICHER, M. 2017. Removal of glucuronic acid from xylan is a strategy to improve the conversion of plant biomass to sugars for bioenergy. *Biotechnology for Biofuels*, 10.
- MARTINEZ-ABAD, A., BERGLUND, J., TORIZ, G., GATENHOLM, P., HENRIKSSON, G., LINDSTROM, M., WOHLERT, J. & VILAPLANA, F. 2017. Regular Motifs in Xylan Modulate Molecular Flexibility and Interactions with Cellulose Surfaces. *Plant Physiology*, 175, 1579-1592.
- MARTÍNEZ-ABAD, A., JIMÉNEZ-QUERO, A., WOHLERT, J. & VILAPLANA, F. 2020. Influence of the Molecular Motifs of Mannan and Xylan Populations on their Recalcitrance and Organization in Spruce Softwoods. *Green Chemistry*.
- MATASCI, N., HUNG, L. H., YAN, Z. X., CARPENTER, E. J., WICKETT, N. J., MIRARAB, S., NGUYEN, N., WARNOW, T., AYYAMPALAYAM, S., BARKER, M., BURLEIGH, J. G., GITZENDANNER, M. A., WAFULA, E., DER, J. P., DEPAMPHILIS, C. W., ROURE, B., PHILIPPE, H., RUHFEL, B. R., MILES, N. W., GRAHAM, S. W., MATHEWS, S., SUREK, B., MELKONIAN, M., SOLTIS, D. E., SOLTIS, P. S., ROTHFELS, C., POKORNY, L., SHAW, J. A., DEGIRONIMO, L., STEVENSON, D. W., VILLARREAL, J. C., CHEN, T., KUTCHAN, T. M., ROLF, M., BAUCOM, R. S., DEYHOLOS, M. K., SAMUDRALA, R., TIAN, Z. J., WU, X. L., SUN, X., ZHANG, Y., WANG, J., LEEBENS-MACK, J. & WONG, G. K. S. 2014. Data access for the 1,000 Plants (1KP) project. *Gigascience*, 3.
- MORTIMER, J. C., FARIA-BLANC, N., YU, X., TRYFONA, T., SORIEUL, M., NG, Y. Z., 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. *Plant Journal*, 83, 413-527.

- MORTIMER, J. C., MILES, G. P., BROWN, D. M., ZHANG, Z., SEGURA, M. P., WEIMAR, T., YU, X., SEFFEN, K. A., STEPHENS, E., TURNER, S. R. & DUPREE, P. 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.
- NISHIMURA, H., KAMIYA, A., NAGATA, T., KATAHIRA, M. & WATANABE, T. 2018. Direct evidence for alpha ether linkage between lignin and carbohydrates in wood cell walls. *Scientific Reports*, 8.
- NYSTEDT, B., STREET, N. R., WETTERBOM, A., ZUCCOLO, A., LIN, Y. C., SCOFIELD, D. G., VEZZI, F., DELHOMME, N., GIACOMELLO, S., ALEXEYENKO, A., VICEDOMINI, R., SAHLIN, K., SHERWOOD, E., ELFSTRAND, M., GRAMZOW, L., HOLMBERG, K., HALLMAN, J., KEECH, O., KLASSON, L., KORIABINE, M., KUCUKOGLU, M., KALLER, M., LUTHMAN, J., LYSHOLM, F., NIITTYLA, T., OLSON, A., RILAKOVIC, N., RITLAND, C., ROSSELLO, J. A., SENA, J., SVENSSON, T., TALAVERA-LOPEZ, C., THEISSEN, G., TUOMINEN, H., VANNESTE, K., WU, Z. Q., ZHANG, B., ZERBE, P., ARVESTAD, L., BHALERO, R., BOHLMANN, J., BOUSQUET, J., GIL, R. G., HVIDSTEN, T. R., DE JONG, P., MACKAY, J., MORGANTE, M., RITLAND, K., SUNDBERG, B., THOMPSON, S. L., VAN DE PEER, Y., ANDERSSON, B., NILSSON, O., INGVARSSON, P. K., LUNDEBERG, J. & JANSSEN, S. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497, 579-584.
- OINONEN, P., ZHANG, L. M., LAWOKO, M. & HENRIKSSON, G. 2015. On the formation of lignin polysaccharide networks in Norway spruce. *Phytochemistry*, 111, 177-184.
- PAES, G., BERRIN, J.-G. & BEAUGRAND, J. 2012. GH11 xylanases: Structure/function/properties relationships and applications. *Biotechnology Advances*, 30, 564-592.
- PAN, Y. D., BIRDSEY, R. A., FANG, J. Y., HOUGHTON, R., KAUPPI, P. E., KURZ, W. A., PHILLIPS, O. L., SHVIDENKO, A., LEWIS, S. L., CANADELL, J. G., CIAIS, P., JACKSON, R. B., PACALA, S. W., MCGUIRE, A. D., PIAO, S. L., RAUTIAINEN, A., SITCH, S. & HAYES, D. 2011. A Large and Persistent Carbon Sink in the World's Forests. *Science*, 333, 988-993.
- PATRON, N., ORZAEZ, D., MARILLONNET, S., WARZECHA, H., MATTHEWMAN, C., YOULES, M., RAITSKIN, O., LEVEAU, A., FARRE, G., ROGERS, C., SMITH, A., HIBBERD, J., WEBB, A. A. R., LOCKE, J., SCHORNACK, S., AJIOKA, J., BAULCOMBE, D. C., ZIPFEL, C., KAMOUN, S., JONES, J. D. G., KUHN, H., ROBATZEK, S., VAN ESSE, H. P., SANDERS, D., OLDROYD, G., MARTIN, C., FIELD, R., O'CONNOR, S., FOX, S., WULFF, B., MILLER, B., BREAKSPEAR, A., RADHAKRISHNAN, G., DELAUX, P. M., LOQUE, D., GRANELL, A., TISSIER, A., SHIH, P., BRUTNELL, T. P., QUICK, W. P., RISCHER, H., FRASER, P. D., AHARONI, A., RAINES, C., SOUTH, P. F., ANE, J. M., HAMBERGER, B. R., LANGDALE, J., STOUGAARD, J., BOUWMEESTER, H., UDVARDI, M., MURRAY, J. A. H., NTOUKAKIS, V., SCHAFER, P., DENBY, K., EDWARDS, K. J., OSBOURN, A. & HASELOFF, J. 2015. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist*, 208, 13-19.
- PAVY, N., PELGAS, B., LAROCHE, J., RIGAULT, P., ISABEL, N. & BOUSQUET, J. 2012. A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *Bmc Biology*, 10.
- PROOST, S., VAN BEL, M., VANEECHOUTTE, D., VAN DE PEER, Y., INZE, D., MUELLER-ROEBER, B. & VANDEPOELE, K. 2015. PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Research*, 43, D974-D981.
- RAMAGE, M. H., BURRIDGE, H., BUSSE-WICHER, M., FEREDAY, G., REYNOLDS, T., SHAH, D. U., WU, G. L., YU, L., FLEMING, P., DENSLEY-TINGLEY, D., ALLWOOD, J., DUPREE, P., LINDEN, P. F. & SCHERMAN, O. 2017. The wood from the trees: The use of timber in construction. *Renewable & Sustainable Energy Reviews*, 68, 333-359.
- RENNIE, E. A., HANSEN, S. F., BAIDOO, E. E. K., HADI, M. Z., KEASLING, J. D. & SCHELLER, H. V. 2012. Three Members of the Arabidopsis Glycosyltransferase Family 8 Are Xylan Glucuronosyltransferases. *Plant Physiology*, 159, 1408-1417.

- SAINSBURY, F., THUENEMANN, E. C. & LOMONOSSOFF, G. P. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal*, 7, 682-693.
- SCHELLER, H. V. & ULVSKOV, P. 2010. Hemicelluloses. *Annual Review of Plant Biology*, Vol 61, 61, 263-289.
- SEDJO, R. 2001. Biotechnology's Potential Contribution to Global Wood Supply and Forest Conservation
- SHIMIZU, K., HASHI, M. & SAKURAI, K. 1978. ISOLATION FROM A SOFTWOOD XYLAN OF OLIGOSACCHARIDES CONTAINING 2 4-O-METHYL-D-GLUCURONIC ACID RESIDUES. *Carbohydrate Research*, 62, 117-126.
- SIMMONS, T. J., MORTIMER, J. C., BERNARDINELLI, O. D., POPPLER, A. C., BROWN, S. P., DEAZEVEDO, E. R., DUPREE, R. & DUPREE, P. 2016. Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. *Nature Communications*, 7.
- SPARKES, I. A., RUNIONS, J., KEARNS, A. & HAWES, C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocols*, 1, 2019-2025.
- TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729.
- TAYLOR, E. J., GLOSTER, T. M., TURKENBURG, J. P., VINCENT, F., BRZOWSKI, A. M., DUPONT, C., SHARECK, F., CENTENO, M. S. J., PRATES, J. A. M., PUCHART, V., FERREIRA, L. M. A., FONTES, C., BIELY, P. & DAVIES, G. J. 2006. Structure and activity of two metal ion-dependent acetylxylan esterases involved in plant cell wall degradation reveals a close similarity to peptidoglycan deacetylases. *Journal of Biological Chemistry*, 281, 10968-10975.
- TERASHIMA, N., KITANO, K., KOJIMA, M., YOSHIDA, M., YAMAMOTO, H. & WESTERMARK, U. 2009. Nanostructural assembly of cellulose, hemicellulose, and lignin in the middle layer of secondary wall of ginkgo tracheid. *Journal of Wood Science*, 55, 409-416.
- TERRETT, O. & DUPREE, P. 2019. Covalent interactions between lignin and hemicelluloses in plant secondary cell walls. *Current Opinion in Biotechnology*, 56, 97-104.
- TERRETT, O. M., LYCZAKOWSKI, J. J., YU, L., IUGA, D., FRANKS, W. T., BROWN, S. P., DUPREE, R. & DUPREE, P. 2019. Molecular architecture of softwood revealed by solid-state NMR. *Nature Communications*, 10.
- TIMELL, T. E. & SYRACUSE, N. Y. 1967. Recent progress in the chemistry of wood hemicelluloses. *Wood Science and Technology*, 1.
- VANHOLME, R., DEMEDTS, B., MORREEL, K., RALPH, J. & BOERJAN, W. 2010. Lignin Biosynthesis and Structure. *Plant Physiology*, 153, 895-905.
- VARDAKOU, M., DUMON, C., MURRAY, J. W., CHRISTAKOPOULOS, P., WEINER, D. P., JUGE, N., LEWIS, R. J., GILBERT, H. J. & FLINT, J. E. 2008. Understanding the structural basis for substrate and inhibitor recognition in eukaryotic GH11 xylanases. *Journal of Molecular Biology*, 375, 1293-1305.
- WATANABE, T. & KOSHIIJIMA, T. 1988. EVIDENCE FOR AN ESTER LINKAGE BETWEEN LIGNIN AND GLUCURONIC-ACID IN LIGNIN CARBOHYDRATE COMPLEXES BY DDQ-OXIDATION. *Agricultural and Biological Chemistry*, 52, 2953-2955.
- WHITNEY, S. E. C., BRIGHAM, J. E., DARKE, A. H., REID, J. S. G. & GIDLEY, M. J. 1998. Structural aspects of the interaction of mannan-based polysaccharides with bacterial cellulose. *Carbohydrate Research*, 307, 299-309.
- WILLFOR, S., SUNDBERG, K., TENKANEN, M. & HOLMBOM, B. 2008. Spruce-derived mannans - A potential raw material for hydrocolloids and novel advanced natural materials. *Carbohydrate Polymers*, 72, 197-210.
- WILLFÖR, S., SUNDBERG, A., HEMMING, J. & HOLMBOM, B. 2005. Polysaccharides in some industrially important softwood species. *Wood Science and Technology*, 39, 245-257.

- XIONG, G. Y., 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.
- YAMASAKI, T., ENOMOTO, A., KATO, A., ISHII, T. & SHIMIZU, K. 2011. Structural unit of xylans from sugi (*Cryptomeria japonica*) and hinoki (*Chamaecyparis obtusa*). *Journal of Wood Science*, 57, 76-84.
- YU, L., LYCZAKOWSKI, J. J., PEREIRA, C. S., KOTAKE, T., YU, X. L., LI, A., MOGELSVANG, S., SKAF, M. S. & DUPREE, P. 2018. The Patterned Structure of Galactoglucomannan Suggests It May Bind to Cellulose in Seed Mucilage. *Plant Physiology*, 178, 1011-1026.

A)

bioRxiv preprint doi: <https://doi.org/10.1101/2021.01.28.426387>; this version posted January 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B)

C)

A)

B)

A)

B)