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

Jan J. Lyczakowski1,2, Li Yu2, Oliver M. Terrett2, Christina Fleischmann3, Henry Temple2, Glenn Thorlby3, Mathias Sorieul3 and Paul Dupree2

1Department of Plant Biotechnology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Gronostajowa 7, Krakow 30-387, Poland; 2Department of Biochemistry, University of Cambridge, Tennis Court Road Cambridge, CB2 1QW, UK; 3Scion, 49 Sala Street, Rotorua 3020, New Zealand

Abstract

Wood of coniferous trees (softwood), is a globally significant carbon sink and an important source of biomass. Despite that, little is known about the genetic basis of softwood cell wall biosynthesis. Branching of xylan, one of the main hemicelluloses in softwood secondary cell walls, with glucuronic acid (GlcA) is critical for biomass recalcitrance. Here, we investigate the decoration patterns of xylan by conifer GluC Uronic acid substitution of Xylan (GUX) enzymes.

- Through molecular phylogenetics we identify two distinct conifer GUX clades. Using transcriptional profiling we show that the genes are preferentially expressed in secondary cell wall forming tissues. With in vitro and in planta assays we demonstrate that conifer GUX enzymes from both clades are active glucouronyltransferases.
- Conifer GUX enzymes from each clade have different specific activities. While members of clade one add evenly spaced GlcA branches, the members of clade two are also capable of glucoroniating two consecutive xyloses. Importantly, these types of xylan patterning are present in softwood.
- As xylan patterning might modulate xylan-cellulose and xylan-lignin interactions, our results further the understanding of softwood cell wall biosynthesis and provide breeding or genetic engineering targets that can be used to modify softwood properties.

Introduction

The majority of terrestrial carbon is stored in forests as wood (Pan et al., 2011; Ramage et al., 2017; Bar-On et al., 2018). Plantation forests for commercial timber production are mainly growing coniferous species. For example, conifers account for as much as 98% of trees planted in the US managed forests (Sedjo, 2001). The abundance and sustainability of softwood position it as excellent feedstocks for use in alternative industrial processes, tackling the climate emergency. Unfortunately, progress in conifer breeding and genetic engineering aiming to improve softwood suitability for industrial application is hindered by the poor understanding of the genetic basis for its biosynthesis.

Softwood consists mainly 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. Cellulose is believed to be the main load-bearing structure of softwood cell walls. The microfibrils are a crystalline assembly of 18 or 24 glucan chains (Fernandes et al., 2011; Jarvis, 2018). 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 arabinogluconoroxylan (AGX), account for 10–30% and 5–15% of softwood secondary cell walls, respectively (Scheller & 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 (Willför 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 & 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 (Terrett & Dupree, 2019; Berglund et al., 2020a,b; Martínez-Abad et al., 2020). Both in vitro binding assays (Whitney et al., 1998; Berglund et al., 2020a,b) and in planta analysis (Terashima et al., 2009; Yu et al., 2018; Terrett et al., 2019) suggest that GGM can interact with cellulose microfibrils. 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 &
Both AGX and GGM are located <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., 2018). In softwood, the glucuronic acid sidechains of AGX were shown, using solid-state NMR, to locate within bonding distance to cell wall polysaccharides.

To understand how xylan contributes to the softwood cell wall properties it is necessary to fully characterise its structure. 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 xylan branches allows the hemicellulose 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). In Arabidopsis 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 microfibrils. This incompatibility patterning, in Arabidopsis, is laid out by GUX2 enzymes (Bromley et al., 2013).

In conifers (pine and Douglas fir), the GlcA layout on xylan was reported to be simpler than in Arabidopsis (Busse-Wicher et al., 2016b). It consists primarily of an even spacing of MeGlcA branches every six xylose units. In addition to the MeGlcA, conifer xylan also has Ara decorations. These Ara branches are placed predominantly with an even spacing of two xylose units from the MeGlcA. A spacing equal to six xylose units between MeGlcA branches was confirmed also in spruce (Martinez-Abad et al., 2017). Interestingly, the presence of a second pattern consisting in MeGlcA branches on two consecutive xyloses along the AGX backbone was also reported in larch, two sycamore species and in spruce (Shimizu et al., 1978; Yamasaki et al., 2011; Martinez-Abad et al., 2017). The biological functions of these two types of xylan patterning in conifers and the enzymes responsible for their biosynthesis are unknown. It is however possible, that similarly to angiosperms, the patterning of gymnosperm xylan with GlcA may be important for the control of its interaction with the cellulose microfibril.

Here, we report the discovery of a second conifer GUX clade, first clade was described by Lyczakowski et al., 2017, 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 cell wall properties, our results should influence future genetic engineering or breeding approaches aimed at improving softwood suitability for applications in the bioeconomy.

### Materials and Methods

**Sampling, 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 : 8 h, light : 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 *Pinus radiata* (*Pr*) GUX1 and PUGUX2 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) from three individual biological replicates of plants with each biological replicate consisting of a pooled sample of 36 plants. Constructs used for *Pinus taeda* (*Pr*) GUX2 overexpression in gux1/2/3 *A. thaliana* and for *Picea glauca* (*Pg*) GUX1 and PUGUX2 overexpression in esk-kak were prepared using OpenPlant common syntax GoldenGate assembly (Patron et al., 2015). Arabidopsis transformation was performed using the floral dip method as previously described (Clough & Bent, 1998).

**Phylogenetic and bioinformatic analysis**

Coding sequences of *Arabidopsis thaliana* (*At*) GUX and *Populus trichocarpa* (*Pt*) GUX (Kumar et al., 2019) were used in phylogenetic analysis. AtGUX1, AtGUX2 and AtGUX3 were used to identify putative GUX encoding transcripts from Coniferophyta and Gnetophyta transcriptomic data available from 1000 Plant Transcriptomes, Congenie and Gymno Plaza BLAST services (Nystedt et al., 2013; Matasci et al., 2014; Proost et al., 2015). In-house *Pinus radiata* 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 (Supporting Information Table S1). All amino acid sequences were reconstructed from transcripts with ExPaSy translate tool. A maximum likelihood phylogenetic tree was constructed using MEGA6 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. The putative cellulose, mannan, xylan and lignin biosynthesis genes were identified using Congenie BLAST service using *A. thaliana* protein sequences as a query.
**Pinus radiata** RNA isolation and cDNA synthesis

Total RNA was extracted from six different tissue types (root, cambium/xylem, phloem, vegetative bud, needles and green stem). Samples were collected from four individual 5-yr-old clonal *Pinus radiata* trees growing in Rotorua, New Zealand. For the collection of developing xylem and phloem, a section of bark was scored and peeled away from the stem. The exposed cambium and youngest layers of xylem were collected by scraping and the phloem peeled from the inside of the bark. All samples were snap frozen and stored at −80°C. The RNA extractions were performed on 50 mg of plant material from each tissue type using the Plant RNA Isolation Mini Kit (Agilent, Santa Clara, CA, USA). 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 (QuantiBio, Beverly, MA, USA) with 500 ng of DNasel-treated total RNA as a template.

**ddPCR analysis of PrGUX1 and PrGUX2 expression**

The ddPCRs were performed using the Bio-Rad QX200 system (Bio-Rad, Hercules, CA, USA). The 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. Then, 20 μl of the ddPCR reactions along with 70 μl of droplet generation oil for EvaGreen (Bio-Rad) were loaded onto an 8-channel DG8TM 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 s, annealing/extension at 60°C × 60 s); signal stabilization at 4°C × 5 min and 90°C × 5 min, ramp rate set at 2°C s⁻¹) 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. No amplification was observed from RNA-only controls (no RT).

**Nicotiana benthamiana** expression and Western blot analysis

Blunt ended *myc*-tagged GUX and *A*MAGT1 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 *N*rd (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 Lyczkowski et al. (2017). Protein concentration in the membranes was determined fluorometrically using modified Bradford reagent (Expe- deon). Western blot analysis, was performed as described in Lyczkowski et al. (2017) with anti-*myc* antibody (rabbit; Santa-Cruz, A14) and anti-rabbit HRP linked antibody (mouse; Bio-Rad, 170-6515).

**In vitro activity assay for GUX**

For each reaction 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 X-100, 10 mM UDP-GlcA). UDP-GlcA was replaced with water in certain reactions to control for nonspecific 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-min-long 100°C heat treatment. For reactions including acetylxylan esterase (Acetylxylan esterase CE4 from *Clostridium thermocellum*, NzyTech CZ00341) 5 μl (0.2 U) of enzyme suspension was added to the reaction following 3 h from its start. Additional UDP-GlcA (U6751; Sigma-Aldrich) was introduced into the reaction upon addition of the esterase enzyme, leading to doubling of the initial concentration in the reaction tube from 5 mM to 10 mM. 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 from Bligh & Dyer (1959).

**Carbohydrate digestions, polysaccharide analysis by gel electrophoresis (PACE) and saccharification**

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 (20 U) 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⁻¹ purified *Eucarya chrysantha* (Ec) enzyme was applied onto in vitro reaction products for 1 h 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⁻¹ fungal GH115 was added to oligosaccharides resuspended in 0.1 M ammonium acetate pH 5.5 buffer and incubated overnight at 55°C. Before GH115 treatment, the GH11 enzyme was heat inactivated by incubating the oligosaccharide sample at 120°C for 10 min. Following digestion the oligosaccharides were analysed with PAGE as previously described (Mortimer et al., 2010). PAGE band intensity was performed using ImageJ software and the degree of glucuronidation was quantified as described by Mortimer (2017). All saccharification experiments were...
performed on AIR as previously described (Lyczakowski et al., 2017) without any pretreatment of the feedstock.

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, two clades of GUX enzymes produce specific GlcA distribution patterns on xylan backbone (Bromley et al., 2013). Therefore, the presence of two distinct GlcA patterns on gymnosperm xylan (Shimizu et al., 1978; Yamasaki et al., 2011; Martínez-Abad et al., 2017) prompted us to investigate the presence of additional clades of GUX enzymes within this taxonomic group. We used Arabidopsis GUX1, GUX2 and GUX3 sequences to search for putative gymnosperm xylan glucuronyltransferases. A highly conserved region, containing the GT8 domain, was observed in all gymnosperm protein sequences that had high similarity with Arabidopsis GUX1, GUX2 and GUX3 and a maximum likelihood phylogenetic tree was constructed using these amino acid sequences (Fig. 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 GUX2/4/5 clade than to the other conifer clade. 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 PaGUX2 mRNAs are enriched in phloem and xylem (Figs S1, 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 tracheids. Therefore, to further evaluate the likely biological role of putative conifer GUX enzymes, expression of PaGUX1 and PaGUX2, together with the expression of other putative cell wall biosynthesis genes, such as putative cellulose synthase complex members (CesAs) or cellulose synthase like A (CslA) family members, was studied using the Norwood tool (Fig. 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 function in secondary cell wall biosynthesis, was also observed for both *PaGUX1* and *PaGUX2*.

To validate our *in silico* results and to extend expression profiling to another conifer species, we performed an absolute quantification of conifer *GUX1* and *GUX2* expression in four individual clonal *Pinus radiata* (Pr) trees using droplet digital (dd) PCR (Fig. 2). We studied the expression of the genes in xylem/cambium, phloem, root, vegetative bud, green stem and needles. Expression of both *PrGUX1* and *PrGUX2* was detected in all tissues tested, with the exception of phloem, which had only trace amounts of *PrGUX2* mRNA. In line with the *in silico* analysis for *Picea abies*, we have detected strong *PrGUX1* and *PrGUX2* expression in green stems and needles. The strongest expression detected for both *PrGUX1* and *PrGUX2* was in xylem/cambium. In xylem, as in all tissues tested, the expression of *PrGUX1* was six to eleven times higher than that of *PrGUX2*. This suggests that *PrGUX1* 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 conifer taxa. This hypothesis is further supported by a recent report indicating that a *Pinus densiflora* gene homologous to *AtGUX1* is expressed more strongly in xylem 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. benthamiana* for an in vitro activity assay (Lyczakowski et al., 2017). To extend this analysis and study conifer GUX2 clade representatives, we expressed myc-tagged *PgGUX1* and *PgGUX2* enzymes and *Picea sitchensis* (Ps) GUX2 enzyme (Fig. S4). We selected *P. taeda* GUX1 and GUX2 as their protein sequences are identical to those in *P. radiata* (Fig. 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 *ArMAGT1* (Yu et al., 2018) was used as a negative control to monitor for tobacco’s potential endogenous glucuronyltransferase activity. In contrast with 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 (Fig. 3a). In the tobacco system, no xylan glucuronidation background activity was detected as the reaction products coming from the incubation in the presence of...
At - Arabidopsis thaliana
Ptr - Populus trichocarpa
Pe - Picea engelmannii
AI - Abies lasiocarpa
Pp - Pinus ponderosa
Wm - Weleitchia mirabilis
Gm - Gnetum montanum
Js - Juniperus scopulorum
Ls - Larix speciosa
Tb - Taxus baccata
Pg - Taxus glaucra
Pa - Picea abies
Pt - Pinus taeda
Ps - Picea sitchensis
Pr - Pinus radiata
Selmo - Selaginella moellendorffii

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 PgGUX1 and PgGUX2 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 analyised by PACE (Fig. 3b). As expected, the removal of GlcA from the XUXX structure resulted in its co-migration with the xylotetraose (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 at 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 nonreducing 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 cleavage site (Fig. 3c). Therefore, the structure of the unknown oligosaccharide must correspond to XUXX. The identity of the oligosaccharide was 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 XUXX backbone. This supports our conclusion that conifer GUX2 enzymes are likely to generate an oligosaccharide with GlcA present on consecutive xyloses. This structure was already reported in softwood (Shimizu et al., 1978; Yamasaki et al., 2011; Martínez-Abad et al., 2017).

Xylan acetylation modifies and reduces conifer GUX activity

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 performed an in vitro GlcAT reaction to which we added CE4, a xylan acetylesterase enzyme. The CE4 enzyme 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 h after the start of the in vitro glucuronidation reaction to enable some GlcA to be transferred to avoid acceptor precipitation. Following an overnight incubation of the gux1/2 xylan acceptor with PgGUX1 and PgGUX2 enzymes in the presence and absence of CE4 the reaction products were extracted, digested with GH11 and analysed by PACE (Fig. 4a).

GH11 cleavage of all reactions catalysed by PgGUX1 and PgGUX2 in the presence of UDP-GlcA released xylose, xylobiose and XUXX. Independently of the esterase presence, reactions catalysed by PgGUX1 produced an additional band migrating between xylopentaose and xylotetraose standards. However, the intensity of this novel band was increased in the reactions including the CE4 enzyme. Partial xylan deacetylation increased the amount of this oligosaccharide facilitating analysis of its structure. To investigate the identity of this band, oligosaccharides released by the xylanase GH11 were further digested with alpha-glucuronidase GH115 and analysed by PACE (Fig. 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 (Fig. S6b). The presence of this structure is consistent with GH11 specificity, as the enzyme cannot 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 X₅ and X₄ standards, as singly glucuronidated xylohexaose would migrate between X₆ and X₅ oligosaccharides (Bromley et al., 2013). Taking all evidence into consideration, the oligosaccharide is highly likely to be XUXUXX. The reaction catalysed by PgGUX2 produced XUXX and XUXUXX in similar frequency in the absence of acetylxylan esterase, while in the presence of the esterase XUXUXX became the dominant glucuronidated product (Fig. 4b). Having established the likely identity of all glucuronidated structures observed in the PACE experiment (Fig. 4a) we were able to compare the efficiency of in vitro glucuronidation in the presence and absence of xylan.
acetylesterase (Fig. 4b). Our analysis indicates that the addition of the esterase increases the efficiency and alters the specific glucuronidation of \( \text{PgGUX1} \) and \( \text{PtGUX2} \). It is possible that using an acceptor with reduced acetylation allows the conifer GUX enzymes to exert \textit{in vitro} glucuronyltransferase activities more similar to their \textit{in planta} function.

Conifer 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 \textit{in vitro} glucuronyltransferase activity of \( \text{PgGUX1} \) and \( \text{PtGUX2} \), we decided to evaluate its impact on the longer distance GlcA patterning. For this purpose, we analysed the GlcA decoration pattern of the xylan products generated during \textit{in vitro} reactions by \( \text{PgGUX1} \) and \( \text{PtGUX2} \) in the presence or absence of the xylan acetylesterase. The products were deacetylated and digested with xylanase GH30, which enables analysis of GlcA patterning on xylan (Fig. 5a). The released oligosaccharides were derivatised with the 8-aminonaphthalene-1,3,6-trisulfonic acid, disodium salt (ANTS) fluorophore and analysed by PACE (Fig. 5b). In addition to the xylo-oligosaccharide standard, wild-type (WT) \textit{A. thaliana} AIR was digested with GH30 to produce known glucuronidated xylo-oligosaccharide markers.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3}
\caption{\textit{In vitro} activity of two conifer GlucUronic acid substitution of Xylan (GUX) enzymes. (a) Results of GH11 digestion of \textit{in vitro} activity products of glucuronidation reaction catalysed by \textit{Pinus taeda} (Pt) GUX1 and GUX2, \textit{Picea glauca} (Pg) GUX1 and \textit{Picea sitchensis} (Ps) GUX2. Glucuronidated oligosaccharides are marked. Control reaction, performed using membranes enriched for \textit{Arabidopsis} MAGT1 did not release any glucuronidated oligosaccharides. (b) GH115 treatment of PGUX2 \textit{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), red arrow shows GH11 digestion point and blue arrows show GH115 de-glucuronidation. GlcA is not visible in PACE (b) since oligosaccharides are labelled with the ANTS fluorophore only once, after GH11 digestion. GH11 requires three contiguous unsubstituted xyloses on positions \(-2, -1\) and \(+1\) from the digestion point to cleave the xylan backbone. Therefore, the presence of two consecutive decorated xyloses on positions \(+2\) and \(+3\) from one digestion event will lead to a next hydrolysis event happening only between xyloses \(+5\) and \(+6\) from the original hydrolysis point, if these positions are unbranched. This results in formation of xylopentaose oligosaccharide with two glucuronidated xyloses: XUUXX.
\end{figure}
For *PgGUX1*, in the absence of xylan acetylationerase, 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 acetylerase to a glucuronidation catalysed by *PgGUX1*. The 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 (Fig. 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. Upon the addition of xylan acetylase to a glucuronidation catalysed by *PgGUX2*, the digestion pattern produced by GH30 was modified. 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 (Fig. 4a).

Together with the GH11 data (Fig. 4a), the results of the GH30 analysis provide further evidence that the pattern of GlcA branches added by both *PgGUX1* and *PgGUX2* 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 acetylase enabled *PgGUX2* 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 conifer GUX2 enzymes. For this purpose, we expressed *PtGUX2* under the control of *IRX3*, 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 (*PtGUX1* and *PtGUX2*) and their xylan structure was analysed by PACE after xylanase GH11 digestion (Fig. 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 (Fig. 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. To this end, we generated *A. thaliana* *esk-kak* mutant plants expressing *PgGUX1* or *PtGUX2* (lines *Pg1 esk-kak* and *Pt2 esk-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 *Pg1 esk-kak* and *Pt2 esk-kak* plants was digested with xylanase GH11 and the resulting oligosaccharides were separated by PACE (Fig. 6c). The detected glucuronidated structures were primarily XUXUXX and XUUXX in *esk-kak* plants overexpressing *PgGUX1* and *PtGUX2*, respectively. Interestingly, when
overexpressed in the *esk-kak* background, the *Pg*GUX2 enzyme also generated structures resulting in XUXUXX formation by GH11. In *esk-kak* plants overexpressing the *Pg*GUX1 and *Pt*GUX2 enzymes the total degree of glucuronidation exceeded 20% of the backbone (Fig. 6d). This is more than observed previously for *Pg*GUX1 (Lyczakowski et al., 2017) or *Pt*GUX2 expressed in *gux1/2/3* plants. This observation strengthens the hypothesis that a low degree of xylan acetylation is favourable to conifer GUX activity.

Impact of *Pt*GUX2 expression on saccharification of *gux1/2/3* *A. thaliana* biomass

The absence of [Me]GlcA branches on xylan of *gux1/2/3* Arabidopsis is associated with a significant increase in glucose and xylose release from biomass during enzymatic saccharification (Lyczakowski et al., 2017). Therefore, to investigate if the GlcA branches introduced by the *Pt*GUX2 enzyme are capable of rescuing the mutant phenotype we performed a saccharification assay on cell wall material isolated from WT, *gux1/2/3* and *gux1/2/3* plants expressing *Pt*GUX2. In the *gux1/2/3* plants expressing *Pt*GUX2 the release of both glucose (Fig. 7a) and xylose (Fig. 7b) was reduced to that measured for WT Arabidopsis. This suggested that a softwood-like pattern of GlcA branches, introduced by the *Pt*GUX2 enzyme, can participate in the maintenance of biomass recalcitrance.

**Discussion**

The progression towards a more sustainable economy requires an optimised use of softwood, which represent a significant proportion of the lignocellulosic biomass processed in biorefineries.
Despite its importance very little information is known about the molecular basis of softwood cell wall biosynthesis. In this work we investigated the gene expression profiles and biochemical activities of conifer xylan glucuronyltransferases. We found two distinct conifer clades of GUX enzymes expressed during secondary cell wall synthesis. Each clade produced a specific glucuronic acid substitution pattern. The conifer GUX enzymes expressed during secondary cell wall synthesis. Each clade produced a specific glucuronic acid substitution pattern. The conifer GUX enzymes (Lyczakowski et al., 2017). Here, after re-analysing transcriptomic and genomic data of conifer species, we discovered a second clade of conifer GUX enzymes. Our phylogenetic analysis (Fig. 1) indicated that while the previously characterised conifer GUX clade shares a high sequence similarity with hardwood GUX1/3 sequences, the newly identified sequences were more comparable to the hardwood GUX2/4/5 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 (Fig. S5). This is consistent with the theory that most of gymnosperm evolution happened before 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 (Buschiazzo et al., 2012; Pavy 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.

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 & 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, were mainly interspaced by 4, 5 or 6 unsubstituted xylosyl units. Softwood xylan is also decorated with arabinose residues. However, these decorations were 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 suggested 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 predominant 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 Martínez-Abad et al. (2017). 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., 2013). Therefore, with xylan in a two-fold screw conformation, AaGUX1 enzymes produce a pattern with all the GlcA projecting from 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). AaGUX2 places GlcA decorations projecting on both sides of a potential 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 a 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 (Martínez-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 indicate that xylan with GlcA on consecutive xylosyl monomers may also interact with certain parts of the hydrophilic face of the cellulose microfibril (Martínez-Abad et al., 2017). Convergently, recently published ssNMR experiments on spruce indicated 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 that 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 intrahemicellulose 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 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 with 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 (Watansbe & Koshijima, 1988). In this model, a small number of linkages might maintain full xylan–lignin cross-linking (Giummarella & 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 that can be more easily converted into pulp or biofuels. Moreover, as xylan–cellulose 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.

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 that was used in this work. Enzymes used in this study (GH115 and EcGH30) were a kind gift from Kristian Krogh (Novozymes A/S, Denmark). Dr Krogh also acted as a supporting supervisor during OMT doctoral training at the University of Cambridge. 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 JKL as part of the SONATINA 3 programme (project number 2019/32/C/ZN3/00392). 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. The authors declare no conflict of interest.

Author contributions
J JL 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.

ORCID
Paul Dupree https://orcid.org/0000-0001-9270-6286
Jan J. Lyczakowski https://orcid.org/0000-0002-7694-8629
Mathias Soriuel https://orcid.org/0000-0001-7326-3707
Henry Temple https://orcid.org/0000-0002-0415-6403
Oliver M. Terrett https://orcid.org/0000-0002-3796-2858
Glenn Thorlby https://orcid.org/0000-0002-6742-900X
Li Yu https://orcid.org/0000-0001-8820-6705

Data availability
The data that support the findings of this study are available in the main text and in the Supporting Information for this article. Sequences for newly identified P. radiata GUX were deposited in GenBank and access codes are provided in the Supporting
Information. All plant material described in this manuscript is available on a reasonable request from the corresponding authors.

References


**Supporting Information**

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

**Fig. S1** Expression profiling for *Picea abies* (Pa) GUX1.

**Fig. S2** Expression profiling for *Picea abies* (Pa) GUX2.

**Fig. S3** Expression profiles of conifer genes likely to be involved in softwood biosynthesis.

**Fig. S4** Western blot analysis.

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

**Fig. S6** Annotation of the XUXUXX structure.

**Table S1** Primers used in this study.

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

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