Identification of a putative rhamnogalacturonan-II CMP-β-Kdo transferase through a callus-based gene editing method which overcomes embryo lethality.

Yuan Zhang[1,2], Deepak Sharma[3], Yan Liang[1,2], Nick Downs[1,2], Fleur Dolman[4], Kristen Thorne[3], Jose Henrique Pereira[1,2], Paul Adams[1,2], Henrik V. Scheller[1,2,5], Malcolm O'Neill[3], Breeanna Urbanowicz[3], Jenny C. Mortimer[1,2,4]

[1]Joint BioEnergy Institute California, USA,
[2]Lawrence Berkeley National Laboratory, USA,
[3]Complex Carbohydrate Research Center, The University of Georgia, USA,
[4]University of Adelaide, Australia,
[5]University of California, Berkeley, USA

Corresponding author:
J.C. Mortimer, School of Agriculture, Food and Wine, Waite Research Institute, University of Adelaide, Glen Osmond, 5064, Australia. (T) +61-88-313-3615 (E) jenny.mortimer@adelaide.edu.au

Abstract

The pectin rhamnogalacturonan II (RG-II) is an extraordinarily complex plant carbohydrate, crucial for developmental processes. RG-II contains over 20 distinct glycosidic linkages involving 12 distinct sugars; its biosynthesis is predicted to require numerous glycosyltransferases (GTs). RG-II's low abundance in the plant cell wall belies its vital role in plant development. Minor structural modifications lead to lethality or severe growth impairment, posing significant challenges for GT identification via reverse genetics. Here we developed a novel method to generate viable loss-of-function Arabidopsis mutants in callus tissue via gene editing. We combined this with a candidate gene approach to characterize the GT29 RCKT1/MGP2. Analysis of *rckt1* callus revealed a loss of 3-deoxy-D-manno-octulosonic acid (Kdo) from RG-II sidechain C, suggesting RCKT1's role as the RG-II CMP-β-Kdo transferase1. RCKT1 becomes only the second confirmed GT implicated in RG-II biosynthesis. This discovery provides insight into RG-II's structural impact on plant cell walls, as well as a method to further uncover the machinery required for the synthesis of this enigmatic polymer.

The plant cell wall, a dynamic and complex polysaccharide-rich structure consists primarily of the polysaccharides cellulose, hemicellulose and pectins [1]. The specific composition and abundance of these components vary among species and cell types. The significance and diversity of these structurally complex cell wall polysaccharides are evident from the numerous genes involved in their biosynthesis, accounting for over 10% of the plant genome. Glycosyltransferases (GTs) are crucial enzymes in cell wall biosynthesis, facilitating glycosidic bond formation in polysaccharides. Arabidopsis, a model plant, contains 567 predicted GTs across 44 CAZy families, with few characterized functions [2].

Rhamnogalacturonan II (RG-II), a pectin, stands out among cell wall components for its complexity [3–6]. RG-II consists of a short stretch of homogalacturonan (HG) backbone decorated with four distinct side chains (A-D) and two monosaccharide substituents (**Supplementary Fig. 1**). These side chains are composed of 12 different sugars in over 20 different glycosidic linkages, making RG-II arguably the most structurally complex polysaccharide on earth, while being the most structurally conserved in land plants [7,8] (Scheller, 2007 https://doi.org/10.1111/j.1399-3054.2006.00834.x). RG-II predominantly exists as a dimer, formed by a borate diester crosslinking the apiosyl residue (Api) from side chain A in each monomer [8–10], a process crucial for proper cell wall structure and function, accounting for the absolute requirement for boron in normal plant growth and development [11,12].

Mutations affecting RG-II biosynthesis genes can result in severe growth defects or plant lethality, complicating RG-II studies. For example, the Arabidopsis mutant *mur1*, which has impaired GDP-L-fucose (Fuc) biosynthesis, demonstrated that disrupted RG-II side-chain A biosynthesis leads to severe dwarfism [13,14]. Similar phenotypes were observed in the GFP1-silenced mutant, which is defective in the major Golgi-localized GDP-L-Fuc transporter [15], as well as in mutants affecting the synthesis and transport of other sugars required for RG-II biosynthesis, including GDP-L-galactose (Gal) [1617], UDP-D-glucuronic acid (GlcA), a precursor for UDP-D-apiose (Api) [18], and CMP-β-3-deoxy-D-manno-octulosonic acid (Kdo) [19]. Beyond its established role in plant development and reproduction, RG-II and its borate crosslinking have been suggested to play a critical role in cell-cell adhesion, although this aspect has received comparatively less attention. Notably, another study on the Arabidopsis *mur1* mutant revealed a deficiency in cell adhesion, specifically at the interface between stele and cortex cells, which is likely a consequence of the reduced crosslinking of RG-II [20]. The maize mutant known as "*rotten ear*" has reduced boron transport due to a mutation in the boron efflux transporter BOR1, and not only exhibits defective growth and reproduction, but also displays disrupted cell wall organization and loosened cell adhesion in the corn ear [21]. This phenotype is likely due to the insufficient borate

cross-linked RG-II, especially considering that the sole experimentally demonstrated molecular function of boron in plants is the dimerization of RG-II.

Identifying the GTs responsible for forming RG-II glycosyl linkages is essential for understanding its pivotal role in maintaining cell wall integrity and supporting plant growth. Reverse genetics is a commonly used strategy to investigate the function of putative GTs and has provided important insights into the biosynthesis of other cell wall polysaccharides. However, its success has been extremely limited in characterizing RG-II specific GTs, primarily due to the lethal effects of even minor modifications to the RG-II structure. The only RG-II GTs characterized to date are GAUT4, which is responsible for polymerizing the homogalacturonan backbone of RG-II [22], and RGXT1-4, a family of GTs that add xylose to side chain A of RG-II [23–26]. While the loss-of-function mutation in both GAUT4 and RGXT4 are lethal, two distinct strategies were employed to elucidate their functions: RNA interference (RNAi) gene silencing technology and pollen-rescue. Using RNAi, *GAUT4*-knockdown switchgrass and poplar were generated which exhibited substantially reduced content and crosslinking of RG-II [22]. The pollen-rescue strategy enabled the creation of a viable homozygous RGXT4 (also known as MGP4) loss-of-function mutant, which showed decreased xylose content and RG-II dimerization [26]. Despite these successes, both strategies are constrained by inherent limitations: insufficient knockdown of the target gene expression by RNAi can obscure the consequences of gene loss-of-function and result in invisible phenotypic change, while the pollen rescue strategy is only effective if the lethality is associated with the defective male gametophyte. Both methods are time consuming and laborious, require multiple generations of crossbreeding and screening, and it can be challenging to obtain stable lines. Hence, there is a need to establish a system that enables the generation of viable null GT mutants that allows direct investigation of their function.

Previously, when investigating Golgi-localized GTs and nucleotide sugar transporters involved in the glycosylation of glycosylinositolphosphorylceramides (GIPCs), we observed that while the mutant plants exhibited severe growth phenotypes, null mutant callus material grew well in tissue culture [27–29]. GIPCs, like RG-II, have been proposed to have a role in cell-cell adhesion [3029]. This observation led us to hypothesize that it may be possible to cultivate plants disrupted in these classes of genes that are otherwise lethal. Accordingly, we developed a CRISPR-mediated tissue culture transformation approach generating viable loss-of-function mutants as callus. This technique was validated by targeting the well characterized AtGMT1 [28] and later applied to the characterization of a GT known as MGP2, now referred to as RCKT1 (RG-II CMP-β-Kdo transferase1). Our approach obtained viable RCKT1 loss-of-function callus, allowing us to identify

it as the likely CMP-β-Kdo transferase required for RG-II side chain C biosynthesis. This strategy offers a valuable platform for investigating the function of GTs critical for plant development.

Results

Bioinformatic analysis identified Arabidopsis GT29 family members as highly co-expressed with primary cell wall- and GIPC-related genes

Cell-to-cell adhesion plays a fundamental role in plants [31]. The control and regulation of cell adhesion are critical for various cellular processes that are vital for maintaining plant integrity and tissue structure [32]. To identify GTs implicated in this essential process, we conducted a comprehensive bioinformatic screening strategy based on co-expression analysis. We focused on *AtGMT1*, a gene encoding a mannosyltransferase that is specifically involved in the biosynthesis of GIPCs in Arabidopsis. Notably, disruption of *AtGMT1* not only caused a severe plant growth but also led to reduced cell-cell adhesion [28,30].

Our analysis included a total of 441 *Arabidopsis* genes, spanning all CAZy GT families except for GT1, since GT1s are usually responsible for the glycosylation of metabolites. Co-expression analysis was conducted across an extensive publicly available array of 113 developmental RNA seq datasets, representing 11 distinct tissue and organ types. From the clade of genes exhibiting expression patterns highly similar to *AtGMT1*, we identified 75 candidates with a likely Golgi subcellular localization based on SUBA (Subcellular Location Database for Arabidopsis Proteins) (**Supplementary Data 1**). These candidates were regarded as putative GTs relevant to cell-cell adhesion.

Remarkably, our bioinformatic screening identified all three GT29 family members, namely At1g08280 (GALT29A), At1g08660 (RCKT1, also known as MGP2), and At3g48820 (SIA2). They are all annotated as sialyltransferase-like (ST-like) enzymes due to their possession of sialyl motifs, characteristic features of mammalian STs in family GT29. In particular, GALT29A exhibited all four conserved sialyl motifs (L, S, III, VS), while RCKT1 and SIA2 exhibit three well-defined motifs (L, S, III). However, none of these enzymes exhibited detectable ST activity in enzymatic assays [33], which is not unexpected considering the apparent absence of sialic acid in plants. Among these enzymes, GALT29A has been proposed as a potential galactosyltransferase, implicated in the glycosylation of the arabinogalactan proteins (AGPs). However, the functions of RCKT1 and SIA2 remain unknown.

While the presence of sialyl motifs does not confer ST activity to the plant GT29 enzymes, it raises the possibility that they can use donor substrates structurally similar to CMP-Sialic acid,

the sugar donor for STs. *RCKT1* and *SIA2* have been proposed as candidate enzymes for the transfer of 2-keto-3-deoxy-D-manno-octulosonic acid (Kdo) or 2-keto-3-deoxy-D-lyxo-heptulosaric acid (Dha), both rare sugars found in plants with structural similarities to sialic acid [34,35]. CMP-β-Kdo acts as the activated form of Kdo, exclusively found in RG-II. Dha is also unique to RG-II, although the activated form of Dha remains to be identified.

Despite extensive efforts, homozygous mutants of these three genes have never been isolated, suggesting a lethal or severely deleterious phenotype of the mutant plants, and the essential role of these GTs in plant development. Motivated by this challenge, we embarked on the development of a CRISPR-mediated tissue culture transformation strategy to generate viable loss-of-function mutants in the form of callus for studying the function of GTs.

Developing an efficient system for creating biallelic mutant calli with GT knockout

CRISPR/Cas9 technology is a powerful tool for targeted genome editing in plants, enabling efficient generation of the knockout mutants for functional genetics studies. Here, we optimized a previously reported tissue culture-based transformation method [36], and used it to introduce the CRISPR/Cas9 system, with the aim of circumventing the need to produce whole plants and viable seeds (**Fig. 1**). Subsequently, the isolated clonal knockout line can be indefinitely propagated as callus tissue, which is ideal for downstream biochemical and functional characterization.

We made use of a modular cloning system based on [37], which includes three modules: a T-DNA destination vector, a Cas9 entry vector, and a guide RNA (gRNA) entry vector (**Supplementary Fig. 2**). To boost the rate of editing, each construct contained three sgRNAs for each gene target. Furthermore, since the most common outcome following CRISPR/Cas9 editing entails the generation of small insertions and deletions (indels) through non-homologous end joining (NHEJ), the disruption of gene function primarily stems from frameshift mutations. To maximize the impact of such disruption, the sgRNAs were designed to target the protein-coding region proximal to the 5'-untranslated region (5'-UTR).

Assessing the system with a known GT as the target for knockout

To test the effectiveness of the tailored CRISPR/Cas9 toolbox and tissue culture transformation method in creating GT knockout mutant calli, we targeted *AtGMT1* as a proof of concept. *AtGMT1* encodes a mannosyltransferase that glycosylates GIPCs in Arabidopsis. Loss of *AtGMT1* function led to severely stunted growth, significantly reduced abundance of mannose-carrying GIPCs, and defective cell-cell adhesion [28].

The CRISPR/Cas9 strategy targeted three distinct sites in *AtGMT1*: two on Exon1 and one on Exon2 (**Fig. 2 a,b**). The primary roots of ten-day-old Arabidopsis seedlings were used for *Agrobacterium*-mediated transformation, yielding 16 viable calli out of 300 infected root explants (5.3%). The stable integration of the CRISPR/Cas9 cassette in these transformants was indicated by their resistance to kanamycin selection and the detection of DsRed fluorescence, further validated via PCR.

To characterize the editing events in each transformant, the genomic region spanning the three target sites underwent PCR amplification, Sanger sequencing, and subsequent analysis using the DECODR v3.0 (https://decodr.org/) tool [38]. Among the 16 transformants, Cas9-induced mutations were introduced into 13 lines (81%) at the sgRNA3 target site in Exon2. This resulted in 7 homozygous mutants (with two identical mutant alleles), 1 biallelic mutant (with two distinct mutant alleles) and 5 chimeric mutants (carrying two or more types of alleles, with or without mutation) (**Fig. 2c**). The 7 editing events detected in the homozygous lines consisted of 5 frameshifting indel mutations, expected to result in truncated proteins, and 2 in-frame deletion mutations that would lead to the loss of only one amino acid (**Fig. 2c**). The single biallelic mutant had one allele with a frameshift insertion, and another with a frame-preserving deletion (**Fig. 2c**). All the transgenic lines remained unedited at the sgRNA1 and sgRNA2 target sites located on Exon1, while 3 lines completely escaped targeted editing of *AtGMT1* by the CRISPR/Cas9 complex. In summary, this strategy enables the efficient creation of clonal knockout lines in the form of callus, where every individual cell carries the same frameshift mutation within the target gene.

Validating the function of *gmt1* by GIPC analysis

To assess the suitability of homozygous frameshift mutants generated as callus for the functional characterization of gene activity, we conducted a GIPC analysis of the callus lines. The aim was to determine whether the *Atgmt1* CRISPR mutant (*gmt1-4*) created in this study exhibited a reduction in GIPC mannosylation, mirroring the characteristics observed in the previously characterized *Atgmt1* T-DNA insertional mutant (*gmt1-3*) [28]. A transgenic callus line without CRISPR editing was included as a non-edited (NE) control. Additionally, we also examined callus samples derived from seedling roots of both the Col-0 wild-type and *gmt1-3* T-DNA insertional mutant, serving as negative and positive controls, respectively.

GIPCs were extracted from callus tissue cultivated in liquid media, and the degree of glycosylation determined using a previously established thin layer chromatography (TLC) method [39]. The TLC assay revealed an identical GIPC separation pattern in both *gmt1-4* and *gmt1-3*

mutants, which distinctly differed from the pattern observed in both the Col-0 wild-type and NE control lines (**Supplementary Fig. 5).** Notably, the GIPC bands exhibited a larger shift in both *gmt1* mutants, indicating a reduction in GIPC molecular weight due to the absence of mannosylation, consistent with previous findings [28]. Furthermore, these bands displayed a compact and tightly clustered appearance in contrast to the dispersed bands observed in the negative control lines. This distinctive feature marked a significant decrease in the diversity of GIPC species when proper mannosylation of the sugar headgroup is absent.

To further determine the mannose content present in the extracted GIPCs, the monosaccharides were liberated from the GIPC sugar headgroup by trifluoroacetic acid (TFA) hydrolysis and subsequently quantified by high-performance anion exchange chromatography coupled with pulsed amperometric detection (HPAEC-PAD) as previously described [28]. It was found that the mannose content was significantly reduced (3-7 fold) in both *gmt1* mutants compared to that in the wild-type and NE control (**Table 1**). Both the qualitative TLC assay and the quantitative glycan profiling demonstrate that the *gmt1* CRISPR callus mutant reconstituted the phenotype observed in the *gmt1-3* T-DNA insertional mutant.

Creation of the homozygous *RCKT1* knockout mutant callus

Utilizing the efficient and robust GT gene knockout system established in this study, we next targeted the GT29 candidate identified in the bioinformatic screen. The tri-gRNA module in the transformation vector was designed to target 3 distinct sites within the *RCKT1* coding region (Exon 2, 3, and 5) (**Fig. 3a**). We obtained 16 viable calli out of 410 infected root explants (3.9% efficiency). Cas9-mediated editing was restricted to the sgRNA2 target site of *RCKT1*. Genotyping of the transgenic lines identified the presence of 1 homozygous mutant, 2 heterozygous/biallelic mutants, 11 chimeric mutants, and 1 non-edited control line (**Fig. 3b**). The homozygous mutant *rckt1-21* carried a "-49bp/-49bp" biallelic mutation, disrupting the open reading frame and introducing an early stop codon (**Fig. 3b**). Consequently, the predicted product of the *RCKT1* mutant allele in *rckt1-21* is a truncated protein consisting of 104 amino acid residues at the N-terminus, with 73 residues preserved and 31 altered. This mutant RCKT1 variant is expected to lack the entire catalytic domain, rendering *rckt1-21* an ideal knockout line for cell wall analysis. In contrast, the two heterozygous mutants, *rckt1-1* and *rckt1-9*, harbored a "-1/-3" and "+1/-6" biallelic mutation, respectively (**Fig. 3b**). These mutations resulted in one allele carrying a frameshift indel mutation (-1 or +1), while the other allele containing a frame-preserving deletion (-3 or -6), which led to the loss of one or two amino acid(s) (**Fig. 3b**). It's unknown whether such biallelic mutations would cause the complete loss of *RCKT1* function in these heterozygous

mutants. Lastly, all the chimeric mutants were found to possess wild-type alleles at varying percentage (6-52%), which remained recalcitrant to CIRSPR/Cas9 editing.

In summary, the homozygous *rckt1-21*, the heterozygous/biallelic *rckt1-1* and *rckt1-9* represent promising gene edited lines that can be utilized for functional studies of *AtRCKT1*. Additionally, the non-edited control *rckt1-6*, featuring the T-DNA insertion but lacking any *AtRCKT1* editing, serves as an ideal control for comparison.

The absence of RG-II Kdo from the *rckt1* mutant

To explore the effect of *RCKT1* mutation on cell wall biosynthesis, we grew the *rckt1* callus lines, along with relevant controls, in liquid culture, and isolated a cell wall rich fraction (alcohol-insoluble residue, AIR). No significant changes were found in the overall monosaccharide composition of the cell wall in the *rckt1* mutants when compared to that of the wild-type and non-edited control line (**Fig. 4, Supplementary Data 2**). To prepare an RG-II rich fraction, the AIR was hydrolyzed with an endopolygalacturonase (EPG) followed by size-exclusion chromatography (SEC) as previously described [40]. This procedure not only allowed for the separation of RG-II from the other pectic domains, but also enabled the ratio of RG-II monomer to dimer to be determined. Approximately ~80% of the total RG-II in the Col-0 wild-type and non-edited control was found as a dimer (**Supplementary Fig. 6**, **Supplementary Table 4),** a result consistent with previous reports such as [417,41]. However, in both the homozygous *rckt1-21* mutant and the heterozygous/biallelic *rckt1-9* mutant, a drastic reduction in RG-II dimerization was observed, with the monomeric form becoming the predominant configuration of RG-II, accounting for 78% in *rckt1-21* and 74% in *rckt1-9* (**Supplementary Fig. 6 and Supplementary Table 4**), suggesting a disruption to normal RG-II biosynthesis.

Due to the chemical diversity of the RG-II sugar components, there is no single method capable of measuring all RG-II constituents. This is in part due to the challenge in obtaining standards, but also because sugars such as Kdo and Dha are particularly susceptible to cross-ring cleavage by TFA hydrolysis, the most common method used for determining monosaccharide composition. Therefore, we prepared trimethylsilyl methyl-ester methyl glycoside (TMS) derivatives of the samples for analysis by Gas-liquid chromatography with electron impact mass spectrometry (GLC-EI-MS) [40], and showed that while Dha is present in both controls and *rckt1* mutants, Kdo is detected in the controls, but absent in both the *rckt1-21* homozygous mutant and the *rckt1-9* heterozygous/biallelic mutant (**Fig. 5**).

Moreover, the absence of Kdo from the two *rckt1* mutants was corroborated through solution state nuclear magnetic resonance spectroscopy (NMR) analysis, which enabled the non-

destructive identification of glycosyl constituents within intact RG-II. To facilitate the detection of Kdo and Dha, the isolated RG-II was first treated with HCl, converting all dimers into monomers, and subsequently de-esterified to remove acetyl- and methyl- esters for NMR structural analysis. In the ^{1}H-NMR spectrum, the diagnostic resonance of Kdo (H3', δ 1.95 ppm) and Dha (H3', δ 1.88 ppm) were clearly discerned in the wild-type RG-II, while the Kdo signal was absent in the two *rckt1* mutants (**Fig. 6a**). Similarly, in the 2D TOCSY NMR spectrum, the characteristic H3 and H3' resonance (δ 1.88 and 2.13 ppm?) of Kdo were exclusively observed in the wild-type RG-II, whereas the Dha H3 and H3' resonance (δ 1.78? and 2.05 ppm?) were present in both the control and the *rckt1* mutants (**Fig. 6b**).

In summary, both the MS-based and the NMR-based techniques conclusively demonstrate the absence of Kdo from the disaccharide side-chain C of RG-II in the homozygous mutant and the *rckt1-9* heterozygous/biallelic mutant. This compelling evidence strongly supports the hypothesis that RCKT1 functions as a putative Kdo transferase essential to RG-II side-chain C biosynthesis.

Predicted structure of RCKT1 in the presence of CMP-β-Kdo

Heterologous expression of plant GTs has been largely unsuccessful [42,43]. There are additional challenges in characterizing RCKT1 since the acceptor is unknown, and the CMP-β-Kdo substrate is extremely unstable, with an estimated half-life of ~34 minutes [44,45]. Therefore, in order to gain some initial structural information about RCKT1 enzyme, an AlphaFold prediction [46] was used to create RCKT1 structural model. The RCKT1 model shows a predicted N-terminal transmembrane domain from residue 1 to 36, where the transmembrane α-helix started at residue 12 until residue 36. The general architecture of RCKT1 model can be described as a mix of αβ folds, composed of 1 central β-sheet consisting of 7 parallel β-strands surrounded by 14 α-helices and 2 small β-sheets (**Fig. 7a,b**). The closest structural homolog from Protein Data Bank is the sialyltransferases from mammalian *Sus Scrofa* (wild pig) (PDB ID 2WNB) [47], which showed 32% identity with RCKT1. In addition, the sialyltransferases (CstII) from *Campylobacter jejuni* (PDB ID 1RO7) [48], which also showed similar folding compared to RCKT1 model, was solved in complex with the donor nucleotide sugar cytidine 5'-monophosphate (CMP)-3FNeuAc. Superposition of CStII structure and RCKT1 model, allowed us to extract the position of the CMP binding site and consequently model the CMP-β-Kdo into the RCKT1 substrate donor active site. The CMP-β-Kdo binding site of RCKT1 showed the residues Phe306, Thr307 and Thr316 directly contact the cytidine nucleotide by the main and side chain atoms of these residues. Moreover, the Trp315 makes aromatic stacking interactions. The ribose portion of the substrate showed interaction with

the main chain N atoms of Asn186 and Gly287, and the Kdo sugar region showed 2 hydrogen bonds with the single residue Thr236 (**Fig. 7c**). It will be important to test the importance of these predicted residues in future experiments.

Discussion

Advantages of, and potential extensions to, a callus-based gene editing method

In this study, we demonstrated the effectiveness of our callus-based gene editing method as a reverse genetics approach for functional characterization of putative GTs critical for growth and development. Nevertheless, we believe that its potential extends beyond its current application and that it could be further optimized to serve as a powerful tool for both reverse and forward genetic investigations.

One valuable application of our strategy is the creation of higher-order mutants of GTs with potentially redundant functions. Gene redundancy, in addition to lethality, significantly hampers the characterization of plant GTs [42]. For instance, in the case of the GAUT family, which includes 15 Arabidopsis members, only a few have yielded mutants with observable phenotypes that assist in revealing gene function [49] [50] [22]. The multiplexed CRISPR/Cas9 toolbox employed in our study allows the assembly of up to eight sgRNAs into a single T-DNA vector, a crucial feature that can enable targeted editing at multiple genes simultaneously [37].

Furthermore, our method can be adapted for use in forward screens to identify candidate GTs involved in specific biological processes, such as cell-cell adhesion. Known cell adhesion mutants exhibit a decreased regeneration phenotype [51]. An intriguing possibility is to create a CRISPR callus mutant library targeting all 75 candidates in our target clade, and assess organ regeneration capability in Arabidopsis. A similar approach has been reported for the generation of a large-scale mutant rice collection using the CRISPR/Cas9 system [52].

To make our method applicable for the studies proposed above, it's crucial to further enhance the efficiency of CRISPR/Cas9 editing. As exemplified in the cases of GMT1 and RCKT1 in our study, out of three bioinformatically predicted top-performing sgRNAs, only one successfully created a CRISPR mutation at the target site. Drawing from the work of Liang and colleagues, who developed a prescreening method to assess sgRNA efficiency via the tobacco (*Nicotiana benthamiana*) transient expression system [53], this approach could be applied to select highly efficient sgRNAs for the targeted genes of interest. Moreover, improvements in the CRISPR/Cas9 binary vector design, including the use of stronger promoters to drive higher expression of sgRNA

and Cas9, as well as the utilization of intronized Cas9 [54], can further optimize its editing efficiency in plant species.

RG-II structure integrity is important for its function

Prior to our investigation into RCKT1, it has been widely recognized that maintaining the structural integrity of RG-II is essential for its dimerization capacity, a key factor in its functions related to plant growth, development, and viability. This consensus has primarily been supported by research on plant mutants deficient in RG-II specific GTs, as well as biosynthetic enzymes and transporters responsible for RG-II specific nucleotide sugars. As previously described, knockdown of the RG-II side-chain A GT RGXTs [26], and disruption of the synthesis and transport of RG-II nucleotide sugars, such as GDP-L-Fuc [13,1415], GDP-L-Gal [1617], UDP-D-Api [18], all lead to severe growth and developmental defects in the mutant plants. The compositional and structural analysis of the mutant RG-II strongly suggested that the truncation of side-chain A, due to the absence of its D-Xyl, L-Fuc, L-Gal or D-Api was responsible for the defective RG-II dimerization. This finding aligns with the established understanding that borate cross-linking between two RG-II monomers is mediated by D-Api residues in side-chain A [8,9].

In contrast to the findings mentioned above, our study on RCKT1 introduces a novel perspective by revealing that even a structural modification in side-chain C, a region seemingly distant from the crucial boron-binding D-Api residue, can disrupt the borate cross-linking of RG-II. This unexpected revelation broadens the prevailing hypothesis that any alteration to the RG-II structure may exert a substantial impact on its dimerization, consequently compromising the formation of the pectic network *in muro*.

The CRISPR-mediated reverse genetics strategy developed in this study empowers us to further test this hypothesis. By generating loss-of-function mutants of putative RG-II GTs, we could potentially create a diverse array of RG-II glycoforms *in situ*, allowing for an in-depth exploration of the relationship between RG-II structure and functionality. Beyond RG-II dimerization, we can also delve into how structural modifications of RG-II influence its interaction with other pectic domains and various cell wall components. Techniques such as X-ray scattering, or NMR may facilitate this investigation. Consequently, this would open new avenues for us to probe how the pectic network contributes to the mechanical strength and physical properties of the plant primary wall, which play a substantial role in shaping the growth and development of plant cells, tissues, and organs [565758]. The formation of pectic network involves multiple levels of crosslinking, including but not limited to the backbone glycosidic linkages of the three pectic domains, the calcium crosslinking of HG, and the borate ester crosslinking of RG-II [59]. Although

its specific role in maintaining the physical property of plant cell wall is far from clear, RGII-boron cross-linking has been suggested to regulate the wall pore size [60], mechanical strength [61] and extensibility [1].

Challenges to confirm the enzymatic activity of RCKT1 *in vitro*

The findings presented in this study have shed light on the crucial role of RCKT1 in the incorporation of Kdo into RG-II, as demonstrated through the analysis of RG-II composition in the *AtRCKT1* null mutants. To bolster our understanding of RCKT1's function as an RGII-specific Kdo transferase, it is essential to conduct *in vitro* enzymatic assay. However, it is a particularly challenging proposition for reasons which include the difficulties in the heterologous expression of a functional RCKT1 enzyme, the biosynthesis of the donor substrate CMP-β-Kdo, and acquisition of the appropriate RG-II glycoform to serve as the acceptor substrate.

It's worth noting that the plant cell wall GTs have been notoriously challenging to express as a stable and active recombinant protein in orthogonal hosts such as bacteria and yeast [42,43]. Our own attempt to express RCKT1 using the *Escherichia coli* Origami B strain was also unsuccessful. Nonetheless, the recent use of the Human Embryonic Kidney (HEK293) cell system [62], has shown some promise for the production of soluble plant GTs [63–69]. However, a greater challenge arises from the inherent instability of CMP-β-Kdo, the essential donor substrate for the RCKT1 enzyme. CMP-β-Kdo is highly prone to rapid hydrolysis, with a half-life of ~34 mins, resulting in the formation of CMP and free Kdo [44,45]. To address this obstacle, a potential approach involves the *in situ* generation of CMP-β-Kdo using CTP, Kdo and purified CMP-β-Kdo synthetase (KdsB) from *E. coli*. This strategy has proven successful in studies focused on β-Kdo-Transferases involved in bacterial glycolipid biosynthesis [70–72]. Finally, an appropriate RG-II glycoform, capable of accepting Kdo from the nucleotide sugar donor CMP-β-Kdo, is critical for the success of the glycosylation reaction. As of now, the overall process of the RG-II biosynthesis remains mysterious, especially with regard to the sequence of polysaccharide synthesis and the arrangement of the four side-chains. It is plausible that an RG-II lacking side-chain C could serve as an ideal acceptor for Kdo. In such a scenario, the *RCKT1* callus mutant generated in this study could potentially offer the required glycoform present in its cell wall. However, it is also possible that other glycoforms, lacking other side-chains in addition to side-chain C, may be necessary, presenting a more intricate challenge in terms of isolation.

In addition to confirming its enzymatic activity via *in vitro* biochemical assay, it is also important to obtain structural information that enable a comprehensive understanding of its functional mechanism at the molecular level. While it is beyond the scope of this paper to

experimentally determine the crystal structure of heterologously expressed RCKT1, as an initial effort, AlphaFold was used to predict the structural model of RCKT1. It provided insights into the plausible substrate binding site and putative critical residues involved in substrate interaction. To investigate the specific role of the hypothesized binding residues, future work could include generating a series of RCKT1 variants for kinetic analysis.

Materials and Methods

Plant material and incubation conditions

Arabidopsis thaliana ecotype Col-0 or *gmt1-3* seeds were surface sterilized using a 10% (v/v) commercial bleach (Clorox, 6.5%, v/v) in ethanol solution for 10 minutes, and then washed twice with 100% ethanol, and left to air dry overnight. Sterilized seeds were inoculated onto B5 agar media (**Supplementary Table 5**) and stratified in the dark for three days at 4°C. Afterward, the plates were transferred to the growth chamber (8,000 lux, 16:8 L:D, 22°C, 75% humidity). After 10 days, the primary roots were isolated from the seedlings for root transformation as described in the section below. *gmt1-3* allele is as described in [28]. The callus culture of *gmt1-3* was derived from Arabidopsis roots as described in [73].

Calli were maintained by subculturing onto freshly prepared callus induction media (CIM) (**Supplementary Table 5**) in 25mm petri dishes every 14 days. The desirable callus tissue used for each subculture should be creamy and white, indicating healthy status. Calli plates were incubated in the dark of a growth chamber (25°C, 50% humidity). For growing liquid cultures, callus tissue (approx. 100 mg) was inoculated into 1 L flasks containing 250 mL liquid CIM media and shaken at 100 rpm in the dark at 23°C, a method modified from [74]. Callus from liquid cultures was harvested after 14 days of growth. The tissue was collected into falcon tubes and lyophilized at -50°C for 5 days. Fully lyophilized calli were stored at -80°C for downstream biochemical analysis.

Screening novel GTs important for plant cell-cell adhesion by a bioinformatic pipeline

To identify putative GTs involved in the plant cell-cell adhesion process, we set up a bioinformatics pipeline based on the following filtering process.

Filter I: All the Arabidopsis GT genes with the GT1 family removed. Arabidopsis GT genes were obtained from the CAZy database (http://www.cazy.org/). GT genes from GT1 family were removed since they only glycosylated small molecules and are unlikely to be involved in the cell wall biosynthesis. This step resulted in 441 genes from 49 GT families.

Filter II: Genes co-expressed with known Arabidopsis GT gene involved in cell-cell adhesion. The Arabidopsis *GMT1* gene encoding the GIPC Mannosyltransferase was selected as the query gene for co-expression analysis. To perform the co-expression analysis of the Arabidopsis GT genes, 113 developmental RNA seq datasets across 11 tissue and organ types were retrieved from ThaleMine (https://bar.utoronto.ca/thalemine/begin.do) [75]. Genes with similar expression patterns across the datasets were grouped based on hierarchical clustering using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/). The gene clusters were visualized using TreeView (https://jtreeview.sourceforge.net/) and the large clade containing the query gene *GMT1* was extracted.

Filter III: Genes encoding GT localized in Golgi. The subcellular localization of the extracted GTs is obtained by SUBA (https://suba.live/). This step resulted in 75 Golgi-localized candidates as putative GTs most likely to be relevant to plant cell-cell adhesion.

Guide RNA design and transformation vector construction

For the three sgRNA designed to target each gene, DNA oligonucleotides were synthesized by IDT (**Supplementary table 1**) and subsequently annealed and phosphorylated. The

phosphorylated sgDNA were individually cloned into BsmBI site of the Golden Gate entry vectors pYPQ13XA or pYPQ13XB to create pYPQ131A-AtU6-gRNA1, pYPQ132B-AtU3-gRNA2, and pYPQ133A-AtU6-gRNA3. Then the three gRNA expression cassettes were cloned into the recipient vector pYPQ143Amp by Golden Gate assembly to create the vector that contains the Tri-sgRNA module. The Tri-sgRNA module and Cas9 were assembled into the pTkan destination vector by Gateway cloning.

This modular cloning toolkit for CRISPR/Cas9 vector construction was originally developed by [37], and modified in this study. The antibiotic selection makers of the Golden Gate recipient vector pYPQ143 and the Cas9 entry vector pYPQ150 from the original toolkit were replaced by the ampicillin resistance marker (AmpR) to be compatible with the pTKan destination vector we had in house.

Whole plasmid sequencing was performed to validate the sequences of the different components in the transformation vectors targeting AtGM1 and AtRCKT1. The final constructs were transformed into *Agrobacterium fabrum* (previously known as *Agrobacterium tumefaciens*) GV3101 for Arabidopsis root transformation.

Arabidopsis root transformation

Arabidopsis thaliana ecotype Col-0 was transformed with *Agrobacterium fabrum* GV3101 carrying the CRISPR/Cas9 vectors. The method was modified from Vergunst et al. [36]. Sterilized Col-0 seeds were sowed on B5 agar media (**Supplementary Table 5**) and cultured in the growth chamber (8,000 lux, 16:8 L:D, 22C°, 75% humidity). Primary roots were isolated from 10d old seedlings using a sterile razor blade and transferred onto CIM media (**Supplementary Table 5**) to be precultured for three days. *A. tumefacient* GV3101 carrying the transformation vector was grown on LB solid media with antibiotics (rifampin, gentamicin, spectinomycin) and incubated at 30°C for two days. A single *A. tumefacient* colony was inoculated into liquid LB media with antibiotics (rifampin, gentamicin, spectinomycin) and cultured overnight at 30°C by shaking at 200 rpm. *A. tumefacient* cells were pelleted and resuspended in liquid B5 media (final OD600 = 0.1). The isolated primary roots were soaked in the *A. tumefacient* suspension for five minutes, then sliced into 0.5-1 cm explants using a sterile razor blade. The infected root explants were blot dried on sterile filter paper and plated onto CIM media with a lower concentration of D-Glucose (1.8 g/L) and supplemented with Acetosynringone (0.1mM). Root explants were cocultivated with *A. tumefacient* for three days at 25°C under dim light (2,000 lux, 16:8 L:D). The infected root explants were then washed with Timentin-supplemented (100mg/L) B5 liquid media four times and plated onto CIM media supplemented with antibiotics (50 mg/L Kanamycin and 100mg/L Timentin). After 14 days, the explants were transferred to fresh CIM media containing Kanamycin and Timentin. Incipient calli resistant to Kanamycin were observed. Subsequent subcultures occurred every 14 days.

CRISPR editing analysis of the transgenic callus mutants

Callus tissue was sampled from each independent transformation event, flash frozen in liquid nitrogen, and milled with metallic beads using the TissueLyzer (Qiagen) for subsequent genomic DNA (gDNA) extraction. Genomic DNA was extracted from the callus using QIAGEN DNeasy Plant Mini Kit®. Primers that amplified the genomic region of the guide RNA (gRNA) target sites were used to genotype the material (**Supplementary Table 3**). PCR amplicons were column-purified using the Zymo Research DNA Clean & Concentrator® kit, and Sanger sequenced by

Azenta (https://www.azenta.com/). The Decodr v3.0 software (https://decodr.org/) was used to deconvolute the sequenced amplicon for mutation analysis.

GIPC purification and analysis by TLC

GIPC analysis by TLC was performed as described by Jing et al. [39]. Lyophilized callus tissue (~0.5 g dry weight) were ground by metallic beads using TissueLyzer (Qiagen). Sphingolipids were extracted from the grounded tissue using the lower layer (10 mL) of the isopropanol/hexane/water (55:20:25 v/v/v) followed by 33% methylamine (4 mL) treatment as previously described [76]. To further enrich GIPCs, the dried crude extracts were first resuspended and incubated in 1 mL of enrichment buffer (chloroform/ethanol/ammonia/water [10:60:6:24 v/v/v/v]) overnight at room temperature with agitation (2 rpm), and then purified using a weak anion exchange SPE cartridge (Strata X-AW; Phenomenex) as previously described [27]. GIPCs bound to the cartridge was eluted with 3 mL of enrichment buffer and dried under nitrogen or clean dry air stream. The dried GIPCs were resuspended in buffer A {chloroform/methanol/[4M ammonium hydroxide in 1.8M ammonium acetate] (9:7:2 v/v/v)} and separated by thin layer chromatography (TLC). The separations were performed on High-performance TLC Silica gel on glass plates (Merck) with buffer A as the mobile phase. GIPCs were stained by primuline solution (10 mg in 100 mL acetone/water [8:2 v/v]), and visualized by fluorescence blue light (460 nm) [77]. The water used in this experiment is HPLC-grade only.

GIPC glycan profiling by HPAEC-PAD

Purified GIPC extracts was incubated in 1 mL of the 2M trifluoroacetic acid (TFA) for 1 hour at 120°C, cooled down on the ice, and dried by speed-vac overnight. Following resuspension and proper dilution (2-5 folds) with HPLC grade water, the liberated monosaccharides from the GIPC glycan headgroups were analyzed by high-performance anion-exchange chromatography and a pulsed amperometric detector (HPAEC-PAD) using the elution profile as previously described [28].

Preparation of cell walsl as alcohol insoluble residue (AIR)

AIR was extracted as described in [27]. Briefly, lyophilized callus was milled into a fine powder and resuspended in ethanol (96% v/v). Following incubation at 70°C for 30 minutes, the suspension was cooled, and centrifuged (15 mins, 18,500 rcf). The pellet was then resuspended in ethanol (96% v/v), vortexed, and then centrifuged again (15 mins, 18,500 rcf). The pellet was then resuspeneded in methanol:chloroform (2v:3v) and incubated at 4 °C with agitation (300 rpm) overnight. Following centrifugation (15 min, 18,500 rcf), the pellet was incubated in methanol:chloroform (2v:3v) (1h, 300 rpm, 4 °C). Following centrifugation (15 min, 18,500 rcf), the retained pellet was washed in the following ethanol series: 100% (v/v), 65% (v/v), 70% (v/v), 80% (v/v), 100% (v/v). Each wash involved a 5-minute centrifugation (18,500 rcf) and removal of the supernatant in between, with a thorough vortex to resuspend the pellet. The resulting Alcohol Insoluble Residue (AIR) was dried in a vacuum concentrator or oven at 50 °C overnight.

For starch removal, AIR (~10 mg) was first incubated with α-amylase (3 U/mL) in MOPS buffer (50 mM, pH 7.3) for 10 mins (75°C with agitation at 300 rpm). Pullulanase (1 U/mL) and amyloglucosidase (3 U/mL) in the sodium acetate solution (pH 4.5) were then added, and the samples were incubated for a further two hours (40°C, with agitation at 300 rpm). Following centrifugation, the pellet was washed in the following ethanol series: 100% (v/v), 70% (v/v), 70% (v/v). The destarched AIR was dried in the vacuum concentrator at 30°C overnight.

Cell wall monosaccharide composition analysis

The analysis of the major plant cell wall monosaccharides were performed as previously described [28]. Non-cellulosic sugars were liberated from de-starched AIR by Trifluoroacetic Acid (TFA) hydrolysis at 120°C for one hour. The supernatant containing the liberated monosaccharides was collected after centrifugation. Residual monosaccharides in the remaining pellet were further extracted using water and combined into the former supernatant. The combined supernatant was dried in a vacuum concentrator overnight at 30°C. The dried supernatant was resuspended in HPLC-grade water and filtered through a nitrocellulose filter (0.45 µM). The non-cellulosic sugars in the filtrate were properly diluted (4-5 folds) with water and filtered through a nitrocellulose filter (0.45 µM) for HPAEC-PAD analysis.

The TFA-insoluble AIR pellet was further treated by Saeman hydrolysis to liberate cellulosic sugars. The pellet was resuspended in 72% sulfuric acid and incubated at room temperature for one hour. Following the addition of water that further diluted sulfuric acid to 1M, the samples were incubated at 100°C for three hours. The samples were then neutralized by barium carbonate and centrifuged. The supernatant containing the cellulose-derived sugars was dried in the sample concentrator overnight at 30°C. The dried sugar pellet was resuspended in water and filtered through the nitrocellulose filter (0.45 µM). The cellulosic sugars in the filtrate were properly diluted with water and filter purified again for HPAEC-PAD analysis. The water used in this experiment is HPLC-grade only.

Isolation of RG-II from AIR of controls and the *rckt1* mutants

As described in [40], the callus AIR (x g) were first resuspended in x mL of NaOAc (50 mM, pH 5.2), and then treated with 5 U/g endopolygalacturonase (EPG) at 30 °C for 24 hrs, shaking at 150 rpm. The EPG solublized AIR and the insoluble residue were separated by filtration through a nylon mesh (100 µm pre size). The insoluble material collected on the mesh were retreated with EPG and then refiltered as before. The EPG-soluble material were obtained by combining the filtrates after the two treatments. The x mL solution in total were then fractionated by preparative size-exclusion chromotography (SEC) on a column (1 m x 4 cm) of Sephadex G-75 fine (Cytiva, USA) by elution with NaOAc (50 mM, pH 5) at 1 mL/min. Fractions enriched in RG-II were pooled, dialyzed (3500 Dalton MWCO) against deionized water, and freeze-dried.

RG-II glycosyl residue composition analyses

To specifically detect Kdo and Dha, the glycosyl residue of RG-II were analyzed as their trimethylsilyl methyl glycodide (TMS) derivatives by Gas-liquid chromatography with electron impact mass spectrometry (GLC-EI-MS) as described in [40]. In general, RG-II (x µg) was suspended in x µL methanolic HCl (1 M) in screw-top glass tubes secured with Teflon-lined caps and heated at 80 °C for 18 hrs. After cooling to room temperature, the solutions were concentrated to dry pellets under the nitrogen stream. The released methyl glycosides and methyl glycoside methyl-esters were then reacted for 30 min at 80 °C with Tri-Sil® (ThermoFisher, USA). GLC-EI-MS analysis of the TMS methyl glycosides was performed on an Agilent 7890A GC interfaced to an Agilent 5975C mass selective detector, with a Supelco Equity-1 fused silica capillary column (30 m × 0.25 mm ID).

Preparation of RG-II monomer for NMR spectroscopic analysis

The RG-II dimers were converted to monomers following the protocol described in [40]. Briefly, the RG-II enriched fraction from SEC was first treated with HCl (0.1 M) at room temperature for 1 hr. The solution was then dialyzed (3500 Dalton MWCO) against deionized water and freeze-dried. The resulting RG-II monomers were de-esterified by NaOH (0.1 M) at 4 °C for 16 hrs. Lastly, the solution was neutralized with acetic acid, dialyzed (3500 Dalton MWCO) against water, and freeze-dried.

1D ^{1}H-NMR spectroscopy of RG-II monomer

As described in [40], RG-II samples were dissolved in D2O (0.2 mL, 99.9%; Cambridge Isotope Laboratories, Tewksbury, MA, USA) and placed in a 3 mm NMR tube. The samples were analyzed by a Varian NMR spectrometer (Agilent Technologies) operating at 600 MHz using a 5 mm cold probe, with the ^{1}H-NMR spectra recorded by standard Varian pulse programs. Chemical shifts of the RG-II glycosyl residue were measured relative to internal DMSO (δH 2.721) or acetone (δH 2.225). Data were processed using MestReNova software (Mestrelab Research S.L., Spain).

2D TOCSY NMR analysis of RG-II monomer

Statistical analysis

One-way ANOVA and Student's t-test were used. P value < 0.05 is indicated by different lettering or by *.

Accession Numbers

Sequence data from this article can be found at The Arabidopsis Information Resource (TAIR) under accession number At3g55830 (*AtGMT1*), At1g08660 (*AtRCKT1*), AT3G48820 (*AtSIA2*), and AT1G08280 (*AtGALT29A*).

Figures and Tables (Max 6)

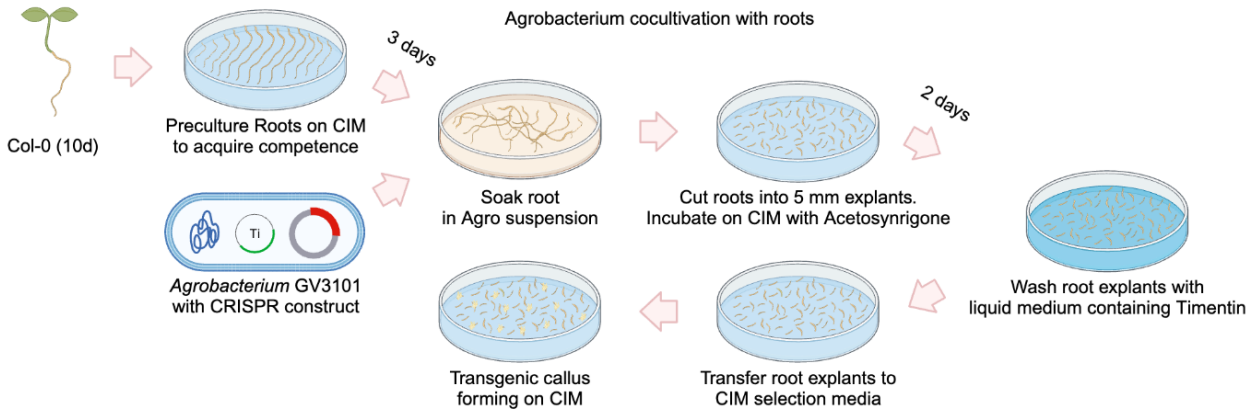

Fig. 1| Agrobacterium-mediated Root Transformation of Arabidopsis. Schematic illustration of the transformation procedure of Arabidopsis root system to generate transgenic callus. Primary roots are first taken from 10-day-old Arabidopsis seedlings and pre-cultured on callus induction media (CIM). After three days, the primary roots are infected with Agrobacteria carrying the T-DNA vector and then cut into root explants. The infected root explants are incubated on co-cultivation media for two days, and then washed using liquid media supplemented with Timentin (100 mg/L) to neutralize the infecting Agrobacteria. The explants are then incubated on CIM media supplemented with selection antibiotics (Kanamycin) only allowing transgenic callus with stable T-DNA integration to survive after about one month.

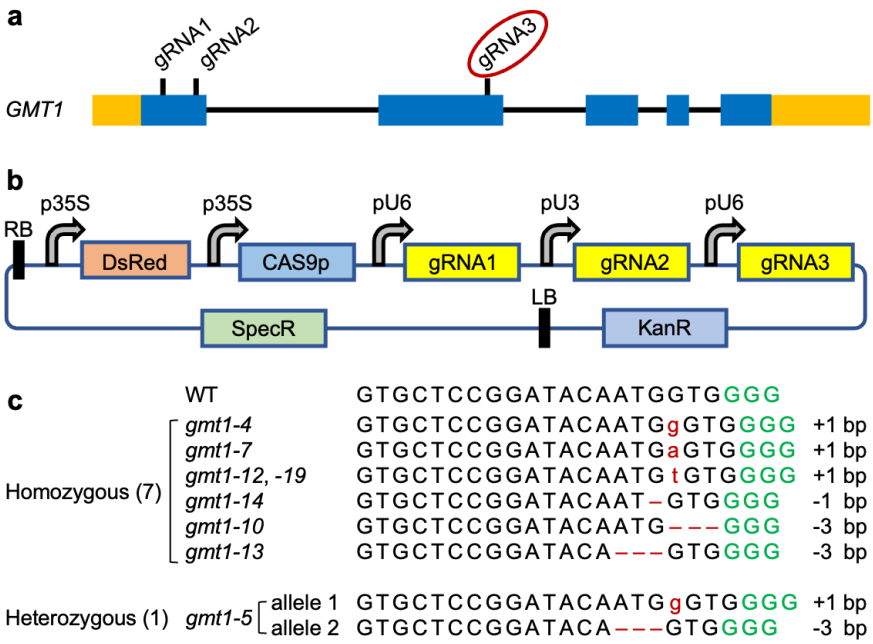

Fig. 2| Creation of CRISPR/Cas9-edited Arabidopsis callus lines with different mutations on *GMT1*. a, Schematic diagram of the *GMT1* gene structure with gRNA targeting sites labeled. The gRNAs were designed using the web-based tool CRISPR-PLANT v2 (http://omap.org/crispr2/) and validated by two additional web-based tools, CRISPR-GE (http://skl.scau.edu.cn/) and CRISPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/). Only the *GMT1* site targeted by gRNA3 has successful CRISPR/Cas9 editing in all lines. **b**, Illustration of the T-DNA transformation vector used for CRISPR editing of *GMT1*. Arabidopsis U6 (AtU6) promoter is used to express gRNA1 and gRNA3, while Arabidopsis U3 (AtU3) is used to express gRNA2. **c**, Mutation analysis of the gRNA3 target site of the *GMT1* gene. Seven homozygous callus mutants were identified: five are *gmt1* knockout mutants with the frameshifted and truncated protein of *GMT1*, and two are mutants with the loss of a single amino acid from the GMT1 protein. One heterozygous callus mutant was identified with two mutant alleles, one with a 1 bp insertion causing premature termination of the GMT1 protein, and the other with a 3 bp deletion resulting in the loss of one amino acid from the GMT1 protein.

Table 1. Mannose content of an enriched GIPC fraction isolated from the calli of Col-0 wild-type and *gmt1* lines.				
µg/mg DW	Col-0	*gmt1-3*	NE control	*gmt1-4*
Man	0.007	**0.001**	0.006	**0.002**
An enriched GIPC fraction from *gmt1* and wild-type callus was hydrolyzed with TFA. The monosaccharides released were quantified by HPAEC-PAD; n=1. Bold indicates a significant difference from the wild-type. DW, dry weight.				

Table 1. Mannose content of the GIPCs isolated from Col-0 wild-type and *gmt1* callus line. GIPCs were extracted from the calli of Col-0 wild-type, NE (non-edited) control, *gmt1-4* and *gmt1-3* lines, subsequently hydrolyzed by TFA to liberate monosaccharides in their glycan headgroups. The derived monosaccharides were quantified by HPAEC-PAD.

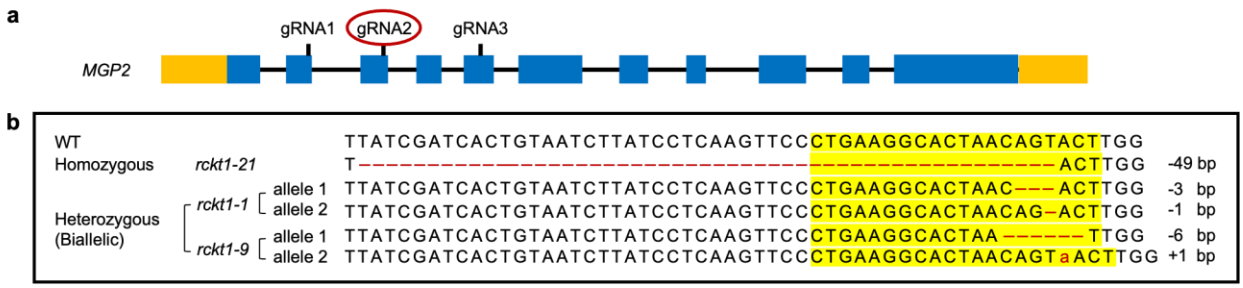

Fig. 3| Creation of CRISPR/Cas9-edited Arabidopsis callus lines with different mutations on *RCKT1*. a, Schematic diagram of the *RCKT1* gene structure with gRNA targeting sites labeled. Only the *RCKT1* site targeted by gRNA2 has successful CRISPR/Cas9 editing in all lines. **b**, Mutation analysis of the gRNA2 target site of the *RCKT1* gene. The homozygous knockout mutant *rckt1-21* has a 49 bp deletion within the *RCKT1* gene, creating a frameshift and truncated protein. The *rckt1-1* heterozygous line has two mutant alleles, one with a 3 bp deletion causing the loss of one amino acid, and the other with a 1 bp deletion resulting in a frameshift. The *rckt1-9* heterozygous line has two mutant alleles, one with a 6 bp deletion causing the loss of two amino acids, and the other with a 1 bp insertion resulting in a frameshift.

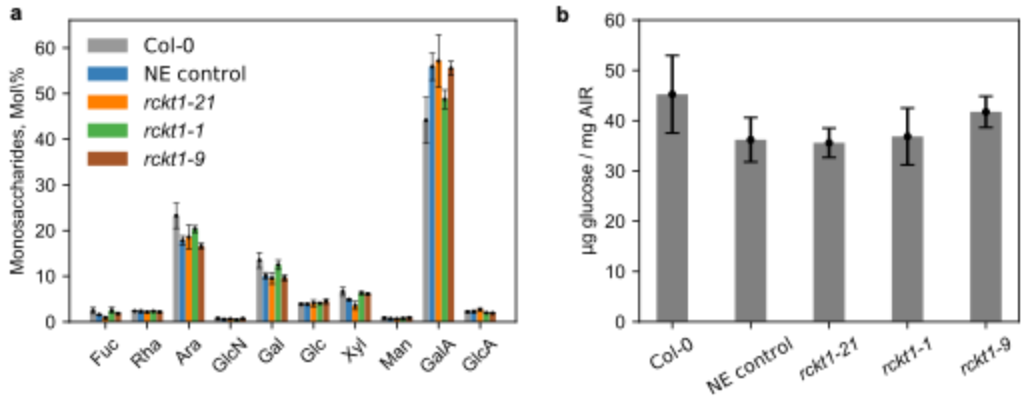

Fig. 4| Monosaccharide composition of *rckt1* callus cell walls. a, Monosaccharide composition of non-cellulosic cell wall polysaccharides **b**, Cellulose-derived glucose. Data are means of three biological replicates ± SD. No statistically significant difference (ANOVA test) was observed between wild-type, Non-edited Control, and the homozygous and heterozygous *rckt1* mutants.

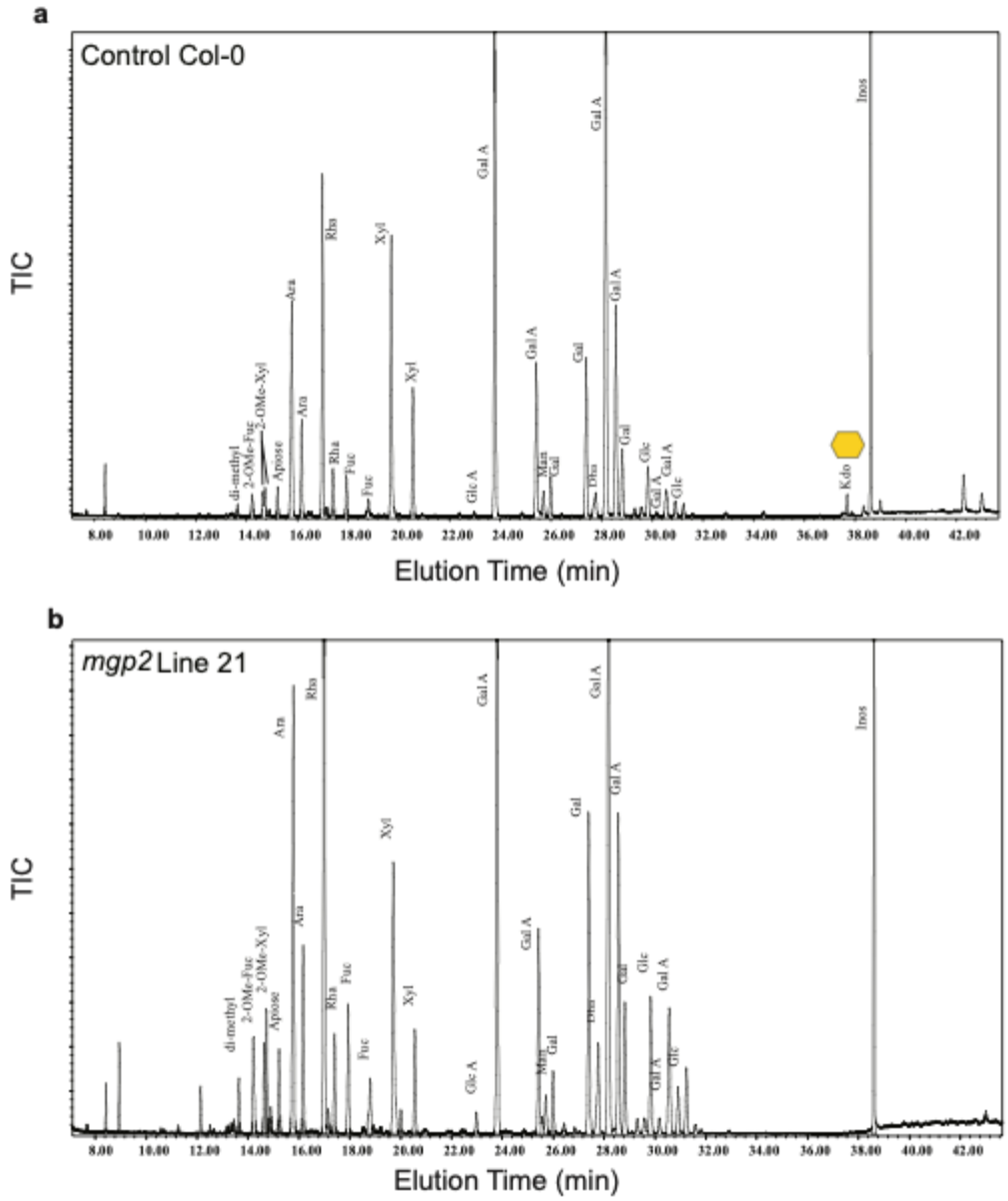

Fig. 5| GLC-EI-MS total ion current profiles of the trimethylsilyl methyl-ester methyl glycoside derivatives of the monosaccharides generated from RG-II. The GLC-EI-MS total ion current (TIC) profile of the TMS derivatives generated from **a**, Col-0 wild-type and **b**, *rckt1*-21 mutant RG-II. The identity of the monosaccharide derivative in each peak is shown. The peak eluting at ~ 39 min is the TMS derivative of Myo-inositol used as an internal standard.

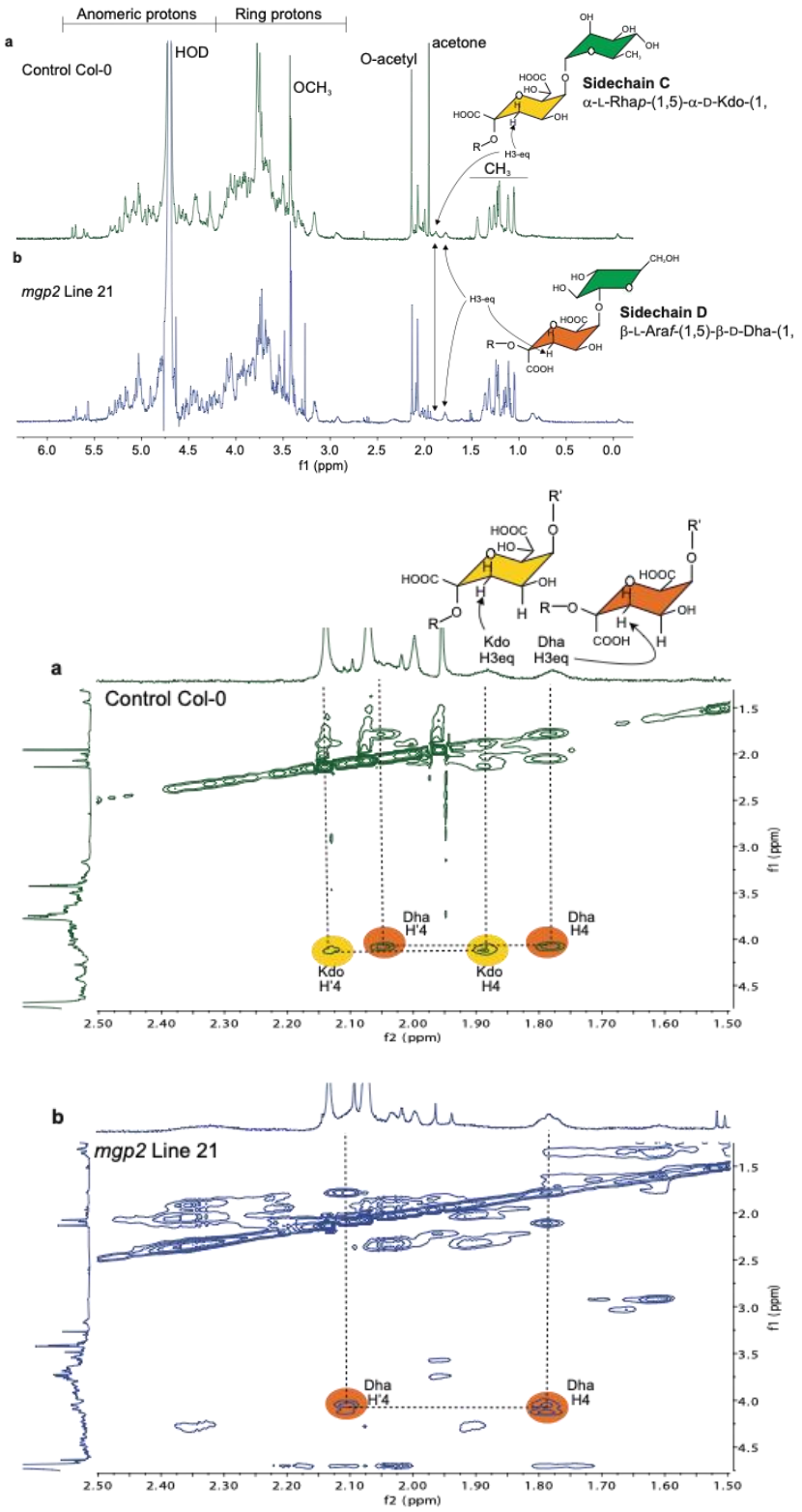

Fig 6| NMR analysis of Arabidopsis RG-II monomer. a, ^{1}H-NMR analysis and **b**, 2D TOCSY NMR analysis of Arabidopsis RG-II monomer isolated from Col-0 wildtype (green) and *rckt1-21* mutant (blue).

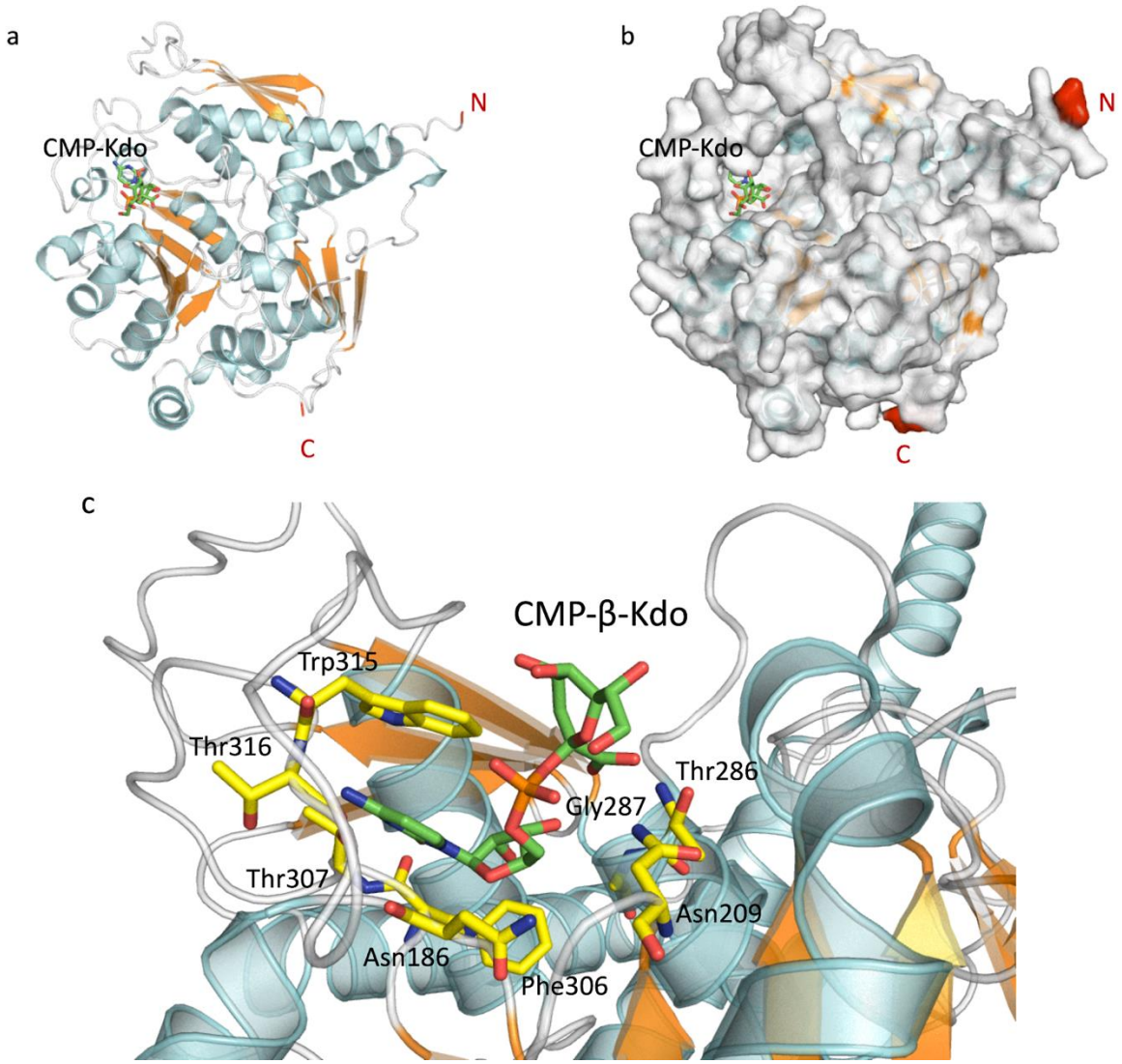

Fig. 7| Structure of Arabidopsis RCKT1 predicted by AlphaFold modeling. a, Cartoon representation of general architecture of RCKT1 model. The predicted N-terminal transmembrane domain correspondent from residue 1 to 36 is not shown. The nucleotide donor substrate, CMP-β-Kdo, is positioned atop a central β-sheet comprising seven parallel β-strands surrounded by 14 α-helices and 2 small β-sheets. **b,** Surface representation of RCKT1 structure in complex with CMP-β-Kdo. It reveals the potential CMP-β-Kdo binding site and the surrounding binding cleft, which is a likely location for the binding of the acceptor substrate. **c**, Zoom in on the CMP-β-Kdo binding site showing all the residues interacting via hydrogen bond and aromatic interactions between RCKT1 and CMP-β-Kdo ligand.

Data availability

The raw data underpinning these experiments is contained in the supplementary data files.

Reference

1. Cosgrove, D. J. Building an extensible cell wall. *Plant Physiol.* **189**, 1246–1277 (2022).
2. Ulvskov, P. & Scheller, H. V. Cell walls have a new family. *Nat. Plants* **4**, 635–636 (2018).
3. Ishii, T. & Matsunaga, T. Isolation and characterization of a boron-rhamnogalacturonan-II complex from cell walls of sugar beet pulp. *Carbohydr. Res.* **284**, 1–9 (1996).
4. Ishii, T., Matsunaga, T. & Hayashi, N. Formation of rhamnogalacturonan II-borate dimer in pectin determines cell wall thickness of pumpkin tissue. *Plant Physiol.* **126**, 1698–1705 (2001).
5. Matsunaga, T. *et al.* Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implications for the evolution of vascular plants. *Plant Physiol.* **134**, 339–351 (2004).
6. Funakawa, H. & Miwa, K. Synthesis of borate cross-linked rhamnogalacturonan II. *Front. Plant Sci.* **6**, 223 (2015).
7. Mohnen, D. Pectin structure and biosynthesis. *Curr. Opin. Plant Biol.* **11**, 266–277 (2008).
8. O'Neill, M. A., Ishii, T., Albersheim, P. & Darvill, A. G. Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu. Rev. Plant Biol.* **55**, 109–139 (2004).
9. Kobayashi, M., Matoh, T. & Azuma, Ji. Two Chains of Rhamnogalacturonan II Are Cross-Linked by Borate-Diol Ester Bonds in Higher Plant Cell Walls. *Plant Physiol.* **110**, 1017–1020 (1996).
10. Yang, Y. & Anderson, C. T. Biosynthesis, localisation, and function of pectins in plants. in *Pectin: technological and physiological properties* (ed. Kontogiorgos, V.) 1–15 (Springer International Publishing, 2020). doi:10.1007/978-3-030-53421-9_1.
11. Bar-Peled, M., Urbanowicz, B. R. & O'Neill, M. A. The Synthesis and Origin of the Pectic Polysaccharide Rhamnogalacturonan II - Insights from Nucleotide Sugar Formation and Diversity. *Front. Plant Sci.* **3**, 92 (2012).
12. González-Fontes, A. Why boron is an essential element for vascular plants. *New Phytol.* **226**, 1228–1230 (2020).
13. O'Neill, M. A., Eberhard, S., Albersheim, P. & Darvill, A. G. Requirement of borate cross-linking of cell wall rhamnogalacturonan II for Arabidopsis growth. *Science* **294**, 846–849 (2001).
14. Pabst, M. *et al.* Rhamnogalacturonan II structure shows variation in the side chains monosaccharide composition and methylation status within and across different plant species. *Plant J.* **76**, 61–72 (2013).
15. Rautengarten, C. *et al.* The Arabidopsis Golgi-localized GDP-L-fucose transporter is required for plant development. *Nat. Commun.* **7**, 12119 (2016).
16. Qi, T. *et al.* GDP-D-mannose epimerase regulates male gametophyte development, plant growth and leaf senescence in Arabidopsis. *Sci. Rep.* **7**, 10309 (2017).
17. Sechet, J. *et al.* Suppression of Arabidopsis GGLT1 affects growth by reducing the L-galactose content and borate cross-linking of rhamnogalacturonan-II. *Plant J.* **96**, 1036–1050 (2018).
18. Reboul, R. *et al.* Down-regulation of UDP-glucuronic acid biosynthesis leads to swollen plant cell walls and severe developmental defects associated with changes in pectic polysaccharides. *J. Biol. Chem.* **286**, 39982–39992 (2011).
19. Delmas, F. *et al.* The synthesis of the rhamnogalacturonan II component 3-deoxy-D-manno-2-octulosonic acid (Kdo) is required for pollen tube growth and elongation. *J. Exp. Bot.* **59**, 2639–2647 (2008).

20. Voxeur, A. *et al.* Altered lignification in mur1-1 a mutant deficient in GDP-L-fucose synthesis with reduced RG-II cross linking. *PLoS ONE* **12**, e0184820 (2017).
21. Chatterjee, M. *et al.* The boron efflux transporter ROTTEN EAR is required for maize inflorescence development and fertility. *Plant Cell* **26**, 2962–2977 (2014).
22. Biswal, A. K. *et al.* Sugar release and growth of biofuel crops are improved by downregulation of pectin biosynthesis. *Nat. Biotechnol.* **36**, 249–257 (2018).
23. Egelund, J. *et al.* Arabidopsis thaliana RGXT1 and RGXT2 encode Golgi-localized (1,3)-alpha-D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan-II. *Plant Cell* **18**, 2593–2607 (2006).
24. Egelund, J. *et al.* Functional characterisation of a putative rhamnogalacturonan II specific xylosyltransferase. *FEBS Lett.* **582**, 3217–3222 (2008).
25. Petersen, B. L. *et al.* Assay and heterologous expression in Pichia pastoris of plant cell wall type-II membrane anchored glycosyltransferases. *Glycoconj. J.* **26**, 1235–1246 (2009).
26. Liu, X.-L. *et al.* Male gametophyte defective 4 encodes a rhamnogalacturonan II xylosyltransferase and is important for growth of pollen tubes and roots in Arabidopsis. *Plant J.* **65**, 647–660 (2011).
27. Mortimer, J. C. *et al.* Abnormal glycosphingolipid mannosylation triggers salicylic acid-mediated responses in Arabidopsis. *Plant Cell* **25**, 1881–1894 (2013).
28. Fang, L. *et al.* Loss of inositol phosphorylceramide sphingolipid mannosylation induces plant immune responses and reduces cellulose content in arabidopsis. *Plant Cell* **28**, 2991–3004 (2016).
29. Ishikawa, T. *et al.* GLUCOSAMINE INOSITOLPHOSPHORYLCERAMIDE TRANSFERASE1 (GINT1) Is a GlcNAc-Containing Glycosylinositol Phosphorylceramide Glycosyltransferase. *Plant Physiol.* **177**, 938–952 (2018).
30. Singh, S. K., Eland, C., Harholt, J., Scheller, H. V. & Marchant, A. Cell adhesion in Arabidopsis thaliana is mediated by ECTOPICALLY PARTING CELLS 1--a glycosyltransferase (GT64) related to the animal exostosins. *Plant J.* **43**, 384–397 (2005).
31. Jarvis, M. C., Briggs, S. P. H. & Knox, J. P. Intercellular adhesion and cell separation in plants. *Plant Cell Environ.* **26**, 977–989 (2003).
32. *Plant cell separation and adhesion*. (Blackwell Publishing Ltd, 2007). doi:10.1002/9780470988824.
33. Daskalova, S. M., Pah, A. R., Baluch, D. P. & Lopez, L. C. The Arabidopsis thaliana putative sialyltransferase resides in the Golgi apparatus but lacks the ability to transfer sialic acid. *Plant Biol (Stuttg)* **11**, 284–299 (2009).
34. Angata, T. & Varki, A. Chemical Diversity in the Sialic Acids and Related α-Keto Acids: An Evolutionary Perspective. *Chem. Rev.* **102**, 439–470 (2002).
35. Paccalet, T. *et al.* Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthesizing enzymes. *Plant Biotechnol. J.* **5**, 16–25 (2007).
36. Vergunst, A. C., de Waal, E. C. & Hooykaas, P. J. Root transformation by Agrobacterium tumefaciens. *Methods Mol. Biol.* **82**, 227–244 (1998).
37. Lowder, L. G. *et al.* A crispr/cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol.* **169**, 971–985 (2015).
38. Bloh, K. *et al.* Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data. *The CRISPR Journal* **4**, 120–131 (2021).
39. Jing, B. *et al.* The Arabidopsis thaliana nucleotide sugar transporter GONST2 is a functional homolog of GONST1. *Plant Direct* **5**, e00309 (2021).
40. Barnes, W. J. *et al.* Protocols for isolating and characterizing polysaccharides from plant cell walls: a case study using rhamnogalacturonan-II. *Biotechnol. Biofuels* **14**, 142 (2021).

41. Reuhs, B. L. *et al.* l-Galactose replaces l-fucose in the pectic polysaccharide rhamnogalacturonan II synthesized by the l-fucose-deficient mur1 Arabidopsis mutant. *Planta* **219**, 147–157 (2004).
42. Amos, R. A. & Mohnen, D. Critical review of plant cell wall matrix polysaccharide glycosyltransferase activities verified by heterologous protein expression. *Front. Plant Sci.* **10**, 915 (2019).
43. Welner, D. H. *et al.* Plant cell wall glycosyltransferases: High-throughput recombinant expression screening and general requirements for these challenging enzymes. *PLoS ONE* **12**, e0177591 (2017).
44. Wen, L., Zheng, Y., Li, T. & Wang, P. G. Enzymatic synthesis of 3-deoxy-d-manno-octulosonic acid (KDO) and its application for LPS assembly. *Bioorg. Med. Chem. Lett.* **26**, 2825–2828 (2016).
45. Sugai, T., Lin, C. H., Shen, G. J. & Wong, C. H. CMP-KDO synthetase: overproduction and application to the synthesis of CMP-KDO and analogs. *Bioorg. Med. Chem.* **3**, 313–320 (1995).
46. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
47. Rao, F. V. *et al.* Structural insight into mammalian sialyltransferases. *Nat. Struct. Mol. Biol.* **16**, 1186–1188 (2009).
48. Chiu, C. P. C. *et al.* Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog. *Nat. Struct. Mol. Biol.* **11**, 163–170 (2004).
49. Caffall, K. H., Pattathil, S., Phillips, S. E., Hahn, M. G. & Mohnen, D. Arabidopsis thaliana T-DNA mutants implicate GAUT genes in the biosynthesis of pectin and xylan in cell walls and seed testa. *Mol. Plant* **2**, 1000–1014 (2009).
50. Sterling, J. D. *et al.* Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proc Natl Acad Sci USA* **103**, 5236–5241 (2006).
51. Johnson, K. L., Kibble, N. A. J., Bacic, A. & Schultz, C. J. A fasciclin-like arabinogalactan-protein (FLA) mutant of Arabidopsis thaliana, fla1, shows defects in shoot regeneration. *PLoS ONE* **6**, e25154 (2011).
52. Meng, X. *et al.* Construction of a Genome-Wide Mutant Library in Rice Using CRISPR/Cas9. *Mol. Plant* **10**, 1238–1241 (2017).
53. Liang, Y. *et al.* A screening method to identify efficient sgRNAs in Arabidopsis, used in conjunction with cell-specific lignin reduction. *Biotechnol. Biofuels* **12**, 130 (2019).
54. Grützner, R. *et al.* High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns. *Plant Commun.* **2**, 100135 (2021).
55. Debernardi, J. M. *et al.* A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat. Biotechnol.* **38**, 1274–1279 (2020).
56. Braybrook, S. A. & Peaucelle, A. Mechano-chemical aspects of organ formation in Arabidopsis thaliana: the relationship between auxin and pectin. *PLoS ONE* **8**, e57813 (2013).
57. Levesque-Tremblay, G., Pelloux, J., Braybrook, S. A. & Müller, K. Tuning of pectin methylesterification: consequences for cell wall biomechanics and development. *Planta* **242**, 791–811 (2015).
58. *Pectin: technological and physiological properties*. (Springer International Publishing, 2020). doi:10.1007/978-3-030-53421-9.
59. Caffall, K. H. & Mohnen, D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr. Res.* **344**, 1879–1900 (2009).
60. Fleischer, A., O'Neill, M. A. & Ehwald, R. The Pore Size of Non-Graminaceous Plant Cell Walls Is Rapidly Decreased by Borate Ester Cross-Linking of the Pectic Polysaccharide Rhamnogalacturonan II. *Plant Physiol.* **121**, 829–838 (1999).

61. Zhou, T. *et al.* Low-Boron Tolerance Strategies Involving Pectin-Mediated Cell Wall Mechanical Properties in Brassica napus. *Plant Cell Physiol.* **58**, 1991–2005 (2017).
62. Moremen, K. W. *et al.* Expression system for structural and functional studies of human glycosylation enzymes. *Nat. Chem. Biol.* **14**, 156–162 (2018).
63. Urbanowicz, B. R., Peña, M. J., Moniz, H. A., Moremen, K. W. & York, W. S. Two Arabidopsis proteins synthesize acetylated xylan in vitro. *Plant J.* **80**, 197–206 (2014).
64. Urbanowicz, B. R. *et al.* Structural, mutagenic and in silico studies of xyloglucan fucosylation in Arabidopsis thaliana suggest a water-mediated mechanism. *Plant J.* **91**, 931–949 (2017).
65. Voiniciuc, C. *et al.* Identification of key enzymes for pectin synthesis in seed mucilage. *Plant Physiol.* **178**, 1045–1064 (2018).
66. Amos, R. A. *et al.* A two-phase model for the non-processive biosynthesis of homogalacturonan polysaccharides by the GAUT1:GAUT7 complex. *J. Biol. Chem.* **293**, 19047–19063 (2018).
67. Soto, M. J. *et al.* AtFUT4 and AtFUT6 Are Arabinofuranose-Specific Fucosyltransferases. *Front. Plant Sci.* **12**, 589518 (2021).
68. Engle, K. A. *et al.* Multiple Arabidopsis galacturonosyltransferases synthesize polymeric homogalacturonan by oligosaccharide acceptor-dependent or de novo synthesis. *Plant J.* **109**, 1441–1456 (2022).
69. Amos, R. A. *et al.* Polymerization of the backbone of the pectic polysaccharide rhamnogalacturonan I. *Nat. Plants* **8**, 1289–1303 (2022).
70. Gronow, S., Brabetz, W. & Brade, H. Comparative functional characterization in vitro of heptosyltransferase I (WaaC) and II (WaaF) from Escherichia coli. *Eur. J. Biochem.* **267**, 6602–6611 (2000).
71. Ovchinnikova, O. G. *et al.* Bacterial β-Kdo glycosyltransferases represent a new glycosyltransferase family (GT99). *Proc Natl Acad Sci USA* **113**, E3120-9 (2016).
72. Lanz, N. D. *et al.* Characterization of the β-KDO Transferase KpsS, the Initiating Enzyme in the Biosynthesis of the Lipid Acceptor for Escherichia coli Polysialic Acid. *Biochemistry* (2021) doi:10.1021/acs.biochem.1c00088.
73. Czako, M. & Marton, L. In.* Arabidopsis: The Compleat Guide*(D. Flanders, and C. Dean eds). AFRC PMB Programme. (1992).
74. Prime, T. A., Sherrier, D. J., Mahon, P., Packman, L. C. & Dupree, P. A proteomic analysis of organelles from Arabidopsis thaliana. *ELECTROPHORESIS* (2000).
75. Krishnakumar, V. *et al.* Thalemine: A warehouse for arabidopsis data integration and discovery. *Plant Cell Physiol.* **58**, e4 (2017).
76. Markham, J. E. & Jaworski, J. G. Rapid measurement of sphingolipids from Arabidopsis thaliana by reversed-phase high-performance liquid chromatography coupled to electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **21**, 1304–1314 (2007).
77. Skipski, V. P. [47] Thin-layer chromatography of neutral glycosphingolipids. in *Lipids Part B* vol. 35 396–425 (Elsevier, 1975).
78. Ndeh, D. *et al.* Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70 (2017).
79. Waese, J. *et al.* ePlant: Visualizing and Exploring Multiple Levels of Data for Hypothesis Generation in Plant Biology. *Plant Cell* **29**, 1806–1821 (2017).
80. Schmid, M. *et al.* A gene expression map of Arabidopsis thaliana development. *Nat. Genet.* **37**, 501–506 (2005).
81. Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y. & Nambara, E. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* **41**, 697–709 (2005).

Acknowledgements

This work was conducted as part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract xxx between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. This work was also in part supported by an establishment grant to JCM from the University of Adelaide. Part of this work was conducted by the Complex Carbohydrate Research Center, at the University of Georgia, and supported by xxx under Contract No. xxx. We thank Yu Gao and Yi-Chun Chen for their helpful support with HPAEC-PAD experiments. We thank Jutta Dalton and Mi Yeon Lee for their helpful support with growth chamber maintenance. We thank Ramana Pidatala for his advice on plant tissue culture. We thank Ariel Orellana for the insightful discussion on the putative CMP-Kdo transporter, homologous to CMP-sialic acid transporter identified in mammals.

Contributions

Y.Z. and J.C.M. conceived and designed the study. Y.Z. conducted most of the experiments, assisted by N.D., Y.L., and F.D. RG-II characterization was conducted mostly by D.S., K.T., M.O., and B.U.. J.H.P. and P.A. performed structural analysis of RCKT1 by AlphaFold modeling. Data analysis and interpretation were conducted by Y.Z., H.V.S. D.S., B.U., M.O., J.H.P., and J.C.M. The paper was written by Y.Z. and J.C.M. with revisions from all authors.