International Journal of
Molecular Sciences

MDPI

Article

A Pipeline towards the Biochemical Characterization of the *Arabidopsis* GT14 Family

Lingling Xuan [1,†], Jie Zhang [1,†], Weitai Lu [1], Pawel Gluza [2], Berit Ebert [2], Toshihisa Kotake [3], Mengzhu Lu [1], Yuan Zhang [1], Mads H. Clausen [4], Kim L. Johnson [1,5], Monika S. Doblin [1,5], Joshua L. Heazlewood [2], Antony Bacic [1,5], Lili Song [1,*] and Wei Zeng [1,*]

1 Sino-Australia Plant Cell Wall Research Centre, The State Key Laboratory of Subtropical Silviculture, College of Forestry and Biotechnology, Zhejiang A & F University, Lin'an 311300, China; linglingxuan@stu.zafu.edu.cn (L.X.); jiezhang@zafu.edu.cn (J.Z.); weitailu@stu.zafu.edu.cn (W.L.); lumz@zafu.edu.cn (M.L.); yz192814@ohio.edu (Y.Z.); K.Johnson@latrobe.edu.au (K.L.J.); M.Doblin@latrobe.edu.au (M.S.D.); T.Bacic@latrobe.edu.au (A.B.)
2 School of BioSciences, The University of Melbourne, Parkville, VIC 3010, Australia; pgluza@student.unimelb.edu.au (P.G.); berit.ebert@unimelb.edu.au (B.E.); jheazlewood@unimelb.edu.au (J.L.H.)
3 Division of Life Science, Saitama University, Saitama 338-8642, Japan; kotake@mail.saitama-u.ac.jp
4 Center for Nanomedicine and Theranostics, Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark; mhc@kemi.dtu.dk
5 La Trobe Institute for Agriculture and Food, School of Life Sciences, Department of Animal, Plant, and Soil Sciences, AgriBio, La Trobe University, Bundoora, VIC 3086, Australia
* Correspondence: lilisong@zafu.edu.cn (L.S.); zengw@zafu.edu.cn (W.Z.); Tel.: +86-571-6383-9132 (L.S. & W.Z.)
† These authors contributed equally to this work.

check for updates

Citation: Xuan, L.; Zhang, J.; Lu, W.; Gluza, P.; Ebert, B.; Kotake, T.; Lu, M.; Zhang, Y.; Clausen, M.H.; Johnson, K.L.; et al. A Pipeline towards the Biochemical Characterization of the *Arabidopsis* GT14 Family. *Int. J. Mol. Sci.* **2021**, *22*, 1360. https://doi.org/10.3390/ijms22031360

Academic Editor: Stefano Benini
Received: 23 December 2020
Accepted: 25 January 2021
Published: 29 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Glycosyltransferases (GTs) catalyze the synthesis of glycosidic linkages and are essential in the biosynthesis of glycans, glycoconjugates (glycolipids and glycoproteins), and glycosides. Plant genomes generally encode many more GTs than animal genomes due to the synthesis of a cell wall and a wide variety of glycosylated secondary metabolites. The *Arabidopsis thaliana* genome is predicted to encode over 573 GTs that are currently classified into 42 diverse families. The biochemical functions of most of these GTs are still unknown. In this study, we updated the JBEI *Arabidopsis* GT clone collection by cloning an additional 105 GT cDNAs, 508 in total (89%), into Gateway-compatible vectors for downstream characterization. We further established a functional analysis pipeline using transient expression in tobacco (*Nicotiana benthamiana*) followed by enzymatic assays, fractionation of enzymatic products by reversed-phase HPLC (RP-HPLC) and characterization by mass spectrometry (MS). Using the GT14 family as an exemplar, we outline a strategy for identifying effective substrates of GT enzymes. By addition of UDP-GlcA as donor and the synthetic acceptors galactose-nitrobenzodiazole (Gal-NBD), β-1,6-galactotetraose (β-1,6-Gal_4) and β-1,3-galactopentose (β-1,3-Gal_5) to microsomes expressing individual GT14 enzymes, we verified the β-glucuronosyltransferase (GlcAT) activity of three members of this family (AtGlcAT14A, B, and E). In addition, a new family member (AT4G27480, 248) was shown to possess significantly higher activity than other GT14 enzymes. Our data indicate a likely role in arabinogalactan-protein (AGP) biosynthesis for these GT14 members. Together, the updated *Arabidopsis* GT clone collection and the biochemical analysis pipeline present an efficient means to identify and characterize novel GT catalytic activities.

Keywords: glycosyltransferase; *Arabidopsis*; plant cell wall; glycosylation; AGP; CAZy

1. Introduction

Glycosylation is one of the most important post-translational modifications (PTMs) of biological molecules. Most classes of organic macromolecules including polysaccharides, proteins, lipids, and nucleic acids as well as small molecules are composed of and/or can

be modified with sugars [1]. Moreover, the cell surface of most organisms is coated with polysaccharides; the cell wall surrounding plant and fungal cells, the extracellular matrix of animal cells, bacterial cell envelopes, and the chitin exoskeleton of insects [2,3]. With few exceptions, glycosylation is an enzymatically catalyzed reaction that forms glycosidic bonds by transferring glycosyl residues from an activated donor, such as a nucleotide sugar or a lipid-linked intermediate, to a specific acceptor molecule [4]. This process requires glycosyltransferases (GTs) which can comprise 1–2% of the genome and consumes a significant proportion of cellular metabolism to maintain.

Information relating to GTs has been deposited in the Carbohydrate-Active Enzyme (CAZy) Database (www.cazy.org). This database was created in 1998 and has since been expanded to include other sugar-related enzymes [5]. As of November 2020, the CAZy database has collected 768,132 genes from bacteria, viruses, archaea and 278 eukaryotic species into defined families and non-classified modules. However, only 2065 genes (less than 0.3%) have been functionally characterized. The model plant *Arabidopsis thaliana* is predicted to encode 573 GTs (42 GT families; approx. 2% of genome) compared to over 200 GTs encoded in the human genome (44 GT families; approx. 0.7% of total genes) [6,7]. The abundance of GTs in plants is attributed to a diversity of glycosylated secondary metabolites and a complex polysaccharide-rich cell wall. To facilitate functional approaches, a plant GT clone resource containing 403 CAZy-defined *Arabidopsis* GTs (http://gt.heazleome.org/) was developed [8], and these 573 *Arabidopsis* GTs are from the 42 defined CAZy families.

Small molecule glycosylation is mostly catalyzed by enzymes of the GT1 family [9]. Polysaccharides, such as cellulose the fibrillar component of the cell wall, are synthesized at the plasma membrane by cellulose synthase complexes (CSC) consisting of a "rosette" of cellulose synthase (CESA) catalytic sub-units belonging to the GT2 superfamily [10,11]. By contrast, most non-cellulosic polysaccharides are synthesized by GTs localized in the Golgi apparatus (GA) [12]. Xyloglucan, the dominant hemicellulose in dicot primary cell walls is synthesized by a suite of GTs including Cellulose synthase-like C glucan synthases (CSLCs, GT2) [13,14], xylosyltransferases (XTs, GT34) [15], galactosyltransferases (GalTs, GT47) [16], and fucosyltransferases (FUTs, GT37) [6,17–19]. Heteroxylan is the most dominant hemicellulose in the primary cell wall of the commelinid monocots (e.g., grasses) and the secondary cell walls of both monocots and dicots [20]. The xylan backbone synthesis process, although still not completely understood, involves members of the GT43 (IRX9), GT47 (IRX10, IRX14, IRX7/FRA8) and GT8 (IRX8/GAUT12, PARVUS/GATL1) families [21–26]. Glucuronosyltransferases (GUX, GT8) and arabinosyltransferases (AraT, GT61 in grasses) participate in xylan sidechain decoration [27,28]. Heteromannans, found in minor proportions in the primary vegetative walls of all land plants and as major components of the walls of some seed endosperms as well as the secondary walls of gymnosperms, are synthesized by CSLAs, a sub-class of the GT2 superfamily [29].

Acidic pectins are rich in galacturonic acid (GalA) and include homogalacturonan (HG), xylogalacturonan (XG), rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II) [30]. Only a few GTs involved in pectin biosynthesis have been identified to date, including HG GalATs from the GT8 family [31,32], RG-I RhaTs (GT106) [33], RG-II XylT (GT77) and side chain galactan synthases (GT92) [34–37] and arabinan synthases (GT47) [38,39]. RG-II is considered one of the most complex polysaccharides in nature [40]. However, we only have rudimentary knowledge of the GTs responsible for RG-II biosynthesis. Functionally important glycoproteins such as arabino-galactan-proteins (AGPs) and extensins (EXTs) are also glycosylated in the endoplasmic reticulum/GA endomembrane system [41]. The GT31 family is responsible for the addition of the first galactose (Gal) to the hydroxyproline (Hyp) residues on AGPs and the synthesis of the β-1,3-galactan backbones of AGPs [38,42,43]. GT14 [44,45] and GT37 members [19] have been described that add two of the terminal sugars on AGPs, β-1,6-glucuronic acid (GlcA), and α-1,2-fucose (Fuc), respectively. The glycosylation of EXTs is catalyzed by members of other GT families:

GT77 enzymes add α-1,3/1,5-arabinose (Ara) units [46]. The biochemically characterized *Arabidopsis* cell wall biosynthesis GTs are summarized in Supplementary Table S1.

The catalytic and acceptor specificities of most plant GTs are still unresolved and, therefore, an enzymatic characterization pipeline would greatly assist in this endeavor. The high-throughput characterization strategy for plant GTs could incorporate either genetic analyses or enzymatic assays or a combination of both. Reverse genetics of *Arabidopsis*, the study of the effect of gene disruption on plant phenotype, is a popular approach to determine gene function [47]. However, mutant phenotypes can often be either pleiotropic, redundant, or dependent on certain biotic or abiotic environmental conditions, thereby making the identification of the catalytic specificity of an enzyme challenging [48]. Hence, the biological and biochemical functions of the majority of these GTs is unresolved.

Heterologous protein expression combined with high-throughput enzyme assays can provide direct and robust evidence of a catalytic function. An effective GT enzymatic assay requires the right combination of nucleotide sugar donor, acceptor (derivatized sugar, oligosaccharide, glycolipid, or protein) and the active GT. Bioinformatics analysis of putative GTs is a valuable first step in determining which donors and acceptors to test in enzyme assays. Practical considerations such as availability then follow. Often, the exact acceptor is unknown and potentially complex and/or heterogeneous if derived from natural sources so a synthetic glycoconjugate is utilized instead. Choosing an effective heterologous protein expression system is vital to the success of the GT assay. It is generally recommended that a variety of heterologous expression systems, including prokaryotic (bacterial) and eukaryotic (e.g., yeast, animal or insect cells, or plants), are utilized to express, purify, and test the enzyme activities of plant GTs. The prokaryotic *Escherichia coli* expression system is a convenient and economical system and usually used to characterize GT1 members; however, due to the lack of post-translational modifications and/or appropriate folding, plant cell wall biosynthetic GTs expressed in *E. coli* are often inactive. Eukaryotic cells are most effective at expressing active plant GTs because they replicate both folding mechanisms and post-translational modifications. The yeast expression system *Pichia pastoris* has been utilized to confirm the catalytic activity of several plant GTs, including those involved in AGP [49], xylan [23], or heteromannan [50] biosynthesis (summarized in Supplementary Table S1). Animal cells can also be utilized to express plant GTs to obtain evidence of their functionality including *Arabidopsis* xyloglucan fucosyltransferase [51] and pectin GalAT (GAUT1) [31,32]. A plant expression system has significant technical advantages over non-plant systems. Previously, an active xylan synthase complex consisting of three proteins—IRX9, IRX14, (both GT43) and IRX10 (GT47)—was successfully expressed using *Agrobacterium*-mediated transformation of *Nicotiana benthamiana* [23]. The same plant expression system has also been widely used to identify the catalytic characteristics of GTs involved in the synthesis of other cell wall polysaccharides, such as (1,3;1,4)-β-glucan (mixed-linkage glucan) synthases [52–54], AGP hydroxyproline *O*-galactosyltransferases [43], and pectin RG-I rhamnosyltransferases [33].

Considering that a GT enzyme reaction requires the right combination of nucleotide sugar donor, appropriate acceptor and active GTs, identification of novel GT activity is challenging. Hence, a systematic platform to characterize the biochemical functions of plant GTs is needed to speed up functional assignment. In this study, we expanded the previous *Arabidopsis* GT clone collection [8] and characterized the catalytic activities of four GT14 family members to illustrate our approach. High-throughput subcellular localization of *Arabidopsis* GT14 members was highlighted in the previous GT collection study [8]. Some members of the GT14 family have previously been shown to exhibit glucuronosyltransferase (GlcAT) activity and therefore it was considered a good example to test our discovery pipeline [44,45]. In this study, all 11 members of the *Arabidopsis* GT14 family were transiently expressed in tobacco leaves and assayed using eight activated nucleotide sugar donors and three synthetic acceptor substrates mimicking AG chains on AGPs. Using this approach, an additional GT14 member was identified as a functionally active GlcAT using the acceptors galactose-nitrobenzodiazole (Gal-NBD) and galacto-

oligosaccharides (β-1,6-galactotetraose (β-1,6-Gal_4) and β-1,3-galactopentose (β-1,3-Gal_5)). Our data provide further evidence that this enzyme family is involved in AGP biosynthesis and demonstrates that this strategy can be utilized to characterize GT function.

2. Results

2.1. Arabidopsis GT Collection

In the previous collection, 403 *Arabidopsis* GTs were cloned into a Gateway-compatible binary vector for expression and functional analysis in plants [8]. We have updated the *Arabidopsis* GT collection in this study. A total of 565 *Arabidopsis* GTs is currently included in the CAZy database (www.cazy.org). Eight additional genes have been predicted to be putative GTs by the TAIR10 annotation (www.arabidopsis.org) or predicted to be non-classified GTs [55,56] and therefore have been added to our collection. A total of 573 predicted GTs and GT-like proteins is listed in Table S2. These include 492 GTs distributed across 42 families and 81 non-classified GTs (GTnc). In summary, 508 (89%) full-length GTs have been successfully cloned into the Gateway-compatible pDONR223 vector and are ready for downstream characterization (Figure 1). A total of 105 newly cloned GTs were added to the *Arabidopsis* collection reported by Lao et al. (2014) [8]. The coding sequence (CDS) of these genes are shown in the nucleotide (NT) sequence column of Table S2. The updated collection includes 65 GTnc (80%); proteins with a domain of unknown function (DUF) 266 (22), DUF 707 (11), DUF616 (6), DUF23 (4), and DUF288 (2). The remaining 20 GTnc members are listed without any specific classification. The 34 members of the DUF246 family are now re-classified as the GT106 family, with members seemingly being involved in pectin biosynthesis. Some have been identified as pectin RG-I rhamnosyltransferases [33] and others are involved with the biosynthesis of pectic arabinogalactans [57]. Additionally, four genes from the GT14-like family (DUF266) were cloned and compared with those in the previous collection [8].

2.2. Phylogenetic and Collinearity Analysis of Arabidopsis GT Families

To examine the phylogenetic relationships of *Arabidopsis* GT sequences, we constructed a maximum likelihood phylogenetic tree based on multiple sequences alignments of the full-length amino acid sequences of all 573 GTs from Supplementary Table S1 (see Figure 2). As expected, most GT family members clustered together forming either one or a few master blocks. However, some members showed scattered patterns in the following GT families: GT1, GT10, GT28, GT61, GT106, GT90, GT47, GT 75, GT77, GT95, GT32, GT2, GT37, GT41, GT66, GT34, GT92, GT17, GT48, GT35, GT14, and GT20. GT families 30, 19, 50, 13, 76, 59, 24, 16, 33, 58, and 96 have only one member. In addition, members in GT families 29, 8, 43, 31, 57, 22, 64, 5, and 4 are distributed into several different clades across the tree, indicating that diversified biochemical functions of their members are more likely. There are more inverting GTs (30 families, 369 genes) than retaining GTs (12 families, 123 genes) as shown in Figure 2. Since the biochemical functions of GTnc are not characterized, the inverting/retaining mechanisms of GTnc are not assigned.

If the order of a series of at least five genes in one block of a chromosome is identical to that of homologous genes in another part of the genome, the homologous genes are considered to share collinearity [58]. Hence, these genes are more likely to have similar functions compared to genes that are not collinear. Out of 52,270 predicted *Arabidopsis* proteins according to the TAIR10 database, 6409 proteins have been shown to share collinearity, accounting for 12.26% of all proteins, shown by light grey lines in Figure 3. Among all the collinearity blocks, 84 pairs belong to the GT superfamily (Table S3), indicating the potential functional similarity of these enzymes. The gene density on each chromosome correlates with the number of GTs (Figure 3). The GT distribution of each chromosome is shown in Figure S1.

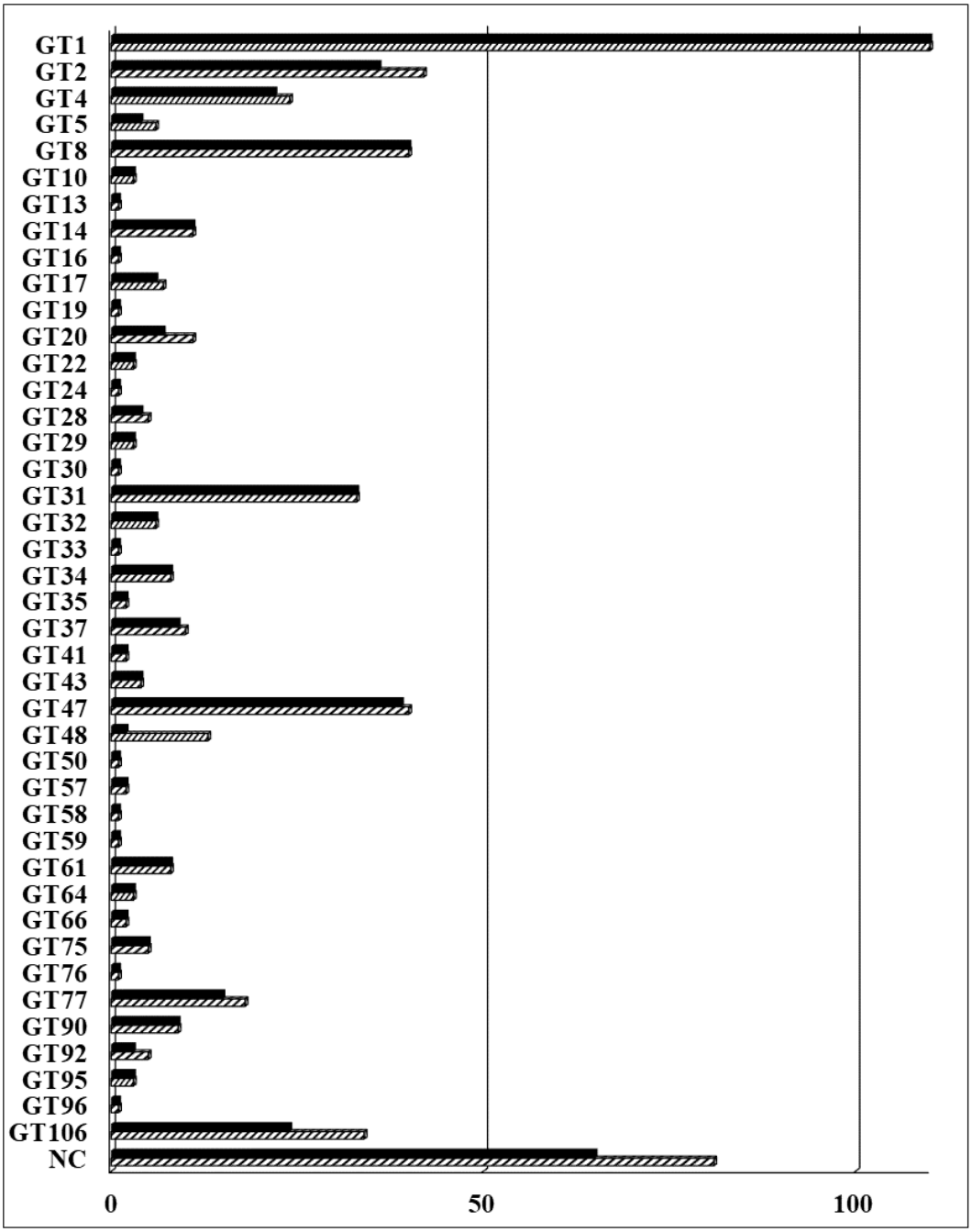


Figure 1. The collection of glycosyltransferases (GTs) from *Arabidopsis thaliana*. A total of 573 predicted GTs including 81 non-classified (NC) proteins are distributed in the 42 listed families; 508 genes have been cloned into pDONR223 vector for downstream analysis. The striped bars represent the total number of genes and the solid bars represent the number of cloned genes in each family. Detailed information of *Arabidopsis* GTs is listed in Supplementary Table S2.

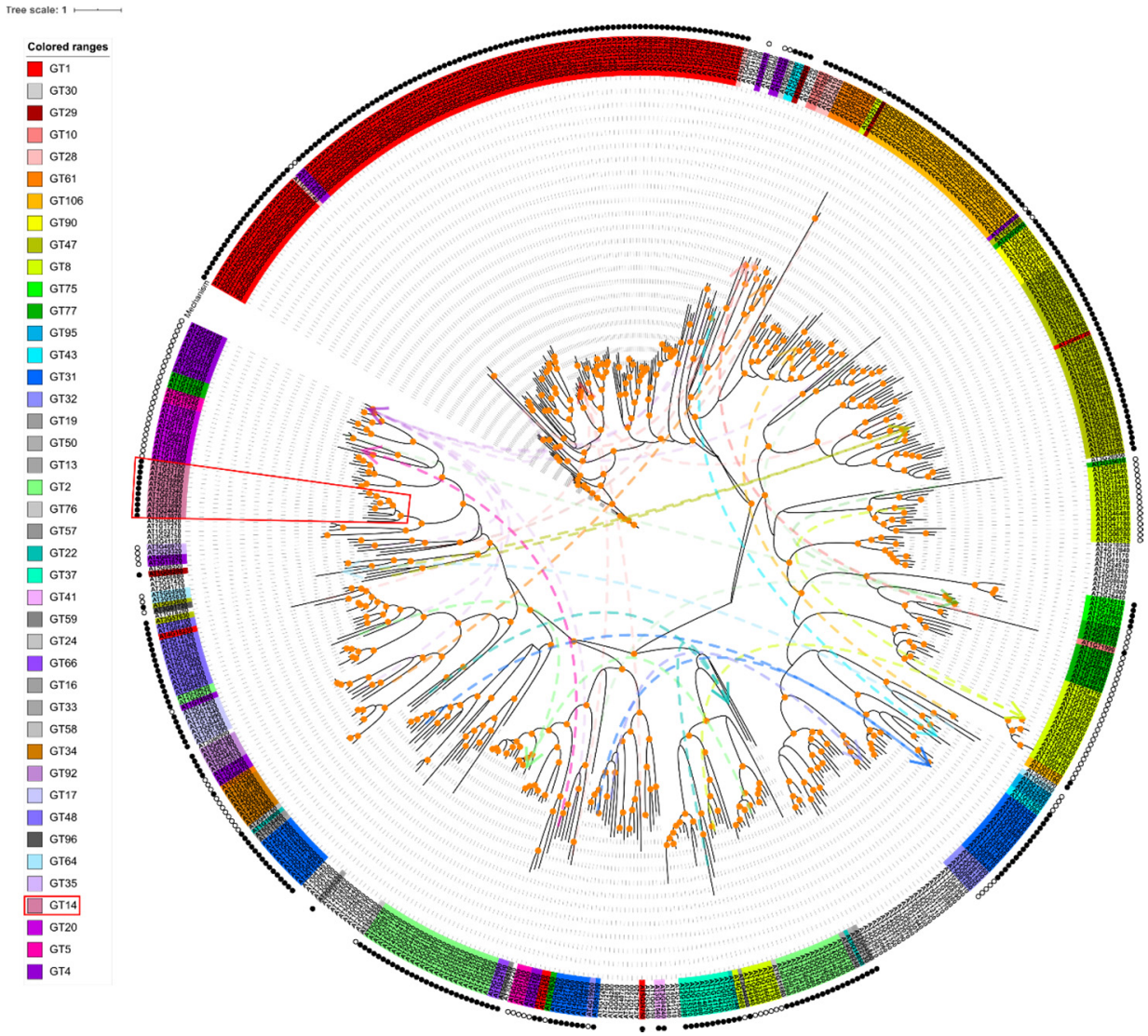


Figure 2. Phylogenetic analysis of the amino acid sequences of *Arabidopsis* GTs. Each GT family has a specific color, the genes without colored ranges are GTnc family, and GT14s are marked with red boxes. The legends with family-specific colors (family colors in all figures are hereby specified) are arranged in the order of family occurrence in the phylogenetic tree. Solid and hollow circles on the periphery represent inverting and retaining mechanism of each GT family, respectively. The branch points with orange circles indicate the bootstrap values over 60 and considered credible. Distant members of the same families on the tree are connected by dotted lines of family-specific colors.

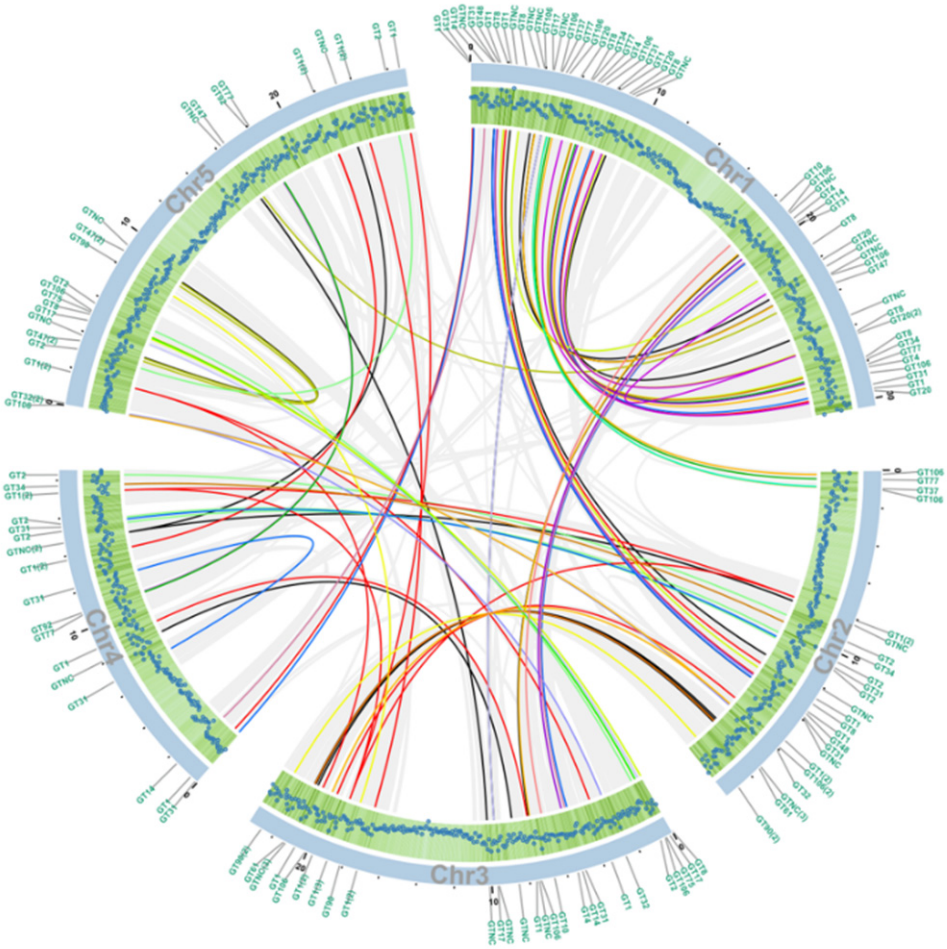


Figure 3. Collinearity analysis of the GT genes between pairs of *Arabidopsis* chromosomes. The scales on the outside of chromosome circle segments in blue represent the length of the chromosome. Each blue dot in the green shaded area shows the gene density at 100 kb intervals, the further towards the outside the greater the density. Collinear blocks of non-GT genes are linked by grey lines, GT family genes are linked by lines of family-specific colors indicated in Figure 2. The GT family to which the gene belongs is marked on the perimeter.

2.3. In Silico Analysis of the Arabidopsis GT14 Family

A maximum likelihood phylogenetic tree of the *Arabidopsis* GT14 family was constructed to obtain further insight into their relatedness and potential function (Figure 4A). The 11 genes (named 240–250) are distributed into 3 clades: AT1G3520 (240) and AT4G03340 (247) form clade A (orange); AT3G03690 (244), AT2G37585 (243), AT1G71070 (242), AT3G24040 (246, AtGlcAT14D) form clade B (yellow); AT5G15050 (249, AtGlcAT14B), AT5G39990 (250, AtGlcAT14A), AT4G27480 (248), AT1G53100 (241), and AT3G15350 (245, AtGlcAT14E) form clade C (pink).

The expression pattern of each GT14 gene in different tissues are shown in a heat map in Figure 4B. The genes AT1G71070 and AT5G15050 have high expression across all tissues and AT4G03340 has the lowest expression across most tissues. AT1G53100 shows highest expressions in senescent organs such as old silique and rosette. The gene structure (intron, exons, 5′ and 3′ UTRs) and protein motif distribution of GT14 members are shown in Figure 4C,D, respectively. Notably, all of the GT14 family genes have a similar exon-intron pattern with three introns and four exons with clade members showing more similarity to each other than with other clade sequences (Figure 4C). Similarly, the GT14 proteins share most of the 10 predicted protein motifs identified with MEME. Motif 4 is missing in AT2G37585, motif 5 missing in AT5G15050, motif 7 is missing in AT3G03690, motif 8 is

missing in AT2G37585 and AT3G03690. Interestingly, motif 10 exists in only one protein (AT4G03340) of clade A and all proteins of clade C except AT3G15350.

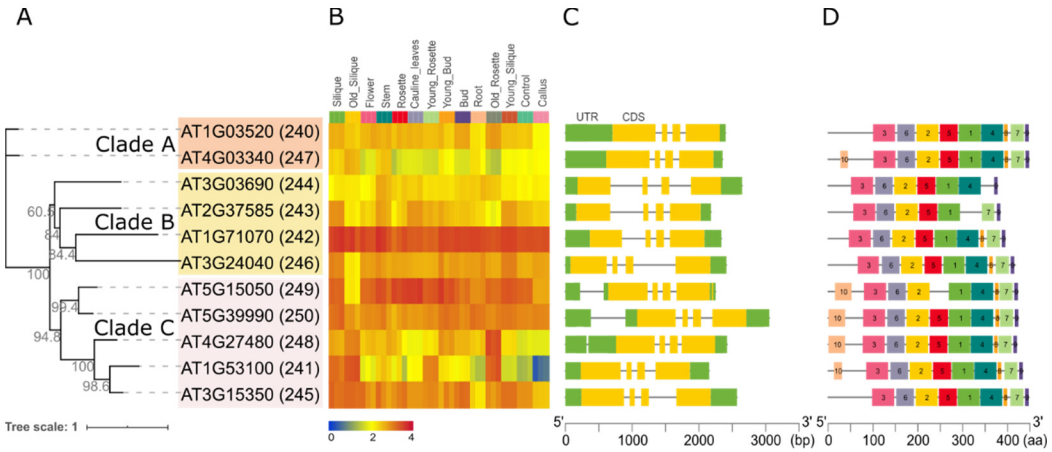


Figure 4. Bioinformatic analyses of the 11 *Arabidopsis* GT14 family genes and their translated proteins. (**A**) ML phylogenetic tree with bootstrap values of 1000 (SH-aLRT test). The sequences were divided into three distinct clades (A, B, and C) shaded in orange, yellow and pink, respectively. (**B**) AtGT14 family genes expression in different tissues during the growth and development processes of *A. thaliana*. The expression data were taken from GSE34188 [59] in the NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). (**C**) Gene structure analysis. The green blocks represent UTR, the yellow blocks represent coding sequence (CDS), the grey lines represent introns. (**D**) Protein sequence motif analysis using MEME with setting the maximum number of motifs = 10. The motifs mean the approximate sequence pattern in a group of homologous genes.

2.4. Heterologous Expression of Arabidopsis GT14 Family Proteins in N. benthamiana

For function assays, C-terminal YFP-tagged versions of the 11 AtGT14 family members were individually expressed in a transient manner in *N. benthamiana* leaves and microsomal membranes (MMs) prepared 3 days post-inoculation. A GFP construct was used as negative control. The level of expression of each fusion protein was detected with anti-GFP antibody using western blotting (Figure S2). Each of the tagged GT14 proteins was expressed and was of the expected size of around 75 kDa (between 44–52 kDa + YFP (approx. 27 kDa), Figure S2). The GFP negative control showed a band of the expected molecular weight of approx. 27 kDa. This confirmed the successful expression of all 11 AtGT14 family proteins in tobacco leaves.

2.5. Identification of Arabidopsis GT14 Family Members as Glucuronosyltransferases (GlcATs)

Previous studies have shown that AtGlcAT14A (AT5G39990, 250), AtGlcAT14B (AT5G15050, 249), and ATGlcAT14C (AT2G37585, 243) have GlcAT activity and are involved in the biosynthesis of type II arabinogalactan (AG), a glycan decoration found on AGPs [44,45]. Recently, AtGlcAT14D (AT3G24040, 246) and AtGlcAT14E (AT3G15350, 245) were also shown to be essential in the glucuronidation of AGs [60]. Here, we sought to investigate the putative GlcAT activity of other members of the AtGT14 family using a suite of synthetic acceptors designed to mimic endogenous AGP acceptors. We utilized MMs prepared from *N. benthamiana* leaves expressing each of the 11 AtGT14 enzymes or the negative control (GFP) in enzymatic assays. Initially, UDP-GlcA and β-galactose-nitrobenzoxadiazole (β-Gal-NBD) were used as donor substrate and acceptor, respectively. β-Gal-NBD has been demonstrated to act as a synthetic mimic of Gal-containing oligosaccharides observed in arabino-3,6-galactan (type II AG) sidechains of AGPs [38]. The chemical structure of β-Gal-NBD is shown in Figure 5A and includes an eight-carbon spacer between β-Gal and NBD to minimize any negative impacts of the tag on enzyme activity, a commonly applied strategy to address steric hinderance issues [61]. Using RP-HPLC to separate the products of the assay, β-Gal-NBD was found to elute at 11.5 min in all samples. A significant peak

was also observed to elute at 8.1 min amongst the assay products of three GT14 member samples (AT3G15350 (245), AT4G27480 (248), and AT5G15050 (249)), with AT3G15350 (245) showing the highest activity. These results demonstrate that Gal-NBD is able to be utilized as acceptor in the GlcAT assay (Figure 5B). The fraction between 7 and 10 min, including the 8.1 min peak of the enzymatic reaction of AT3G15350 (245), was collected and further analyzed by ESI-MS. LC-MS analysis revealed the presence of glucuronosylation of β-Gal-NBD from protonated acceptor (Gal-NBD hydrogen adduct with m/z 552.24 and sodium adduct with m/z 574.22, Figure 5C) with the hydrogen adduct being the most abundant pseudomolecular ion (GlcA-Gal-NBD hydrogen adduct with m/z 728.27 and sodium adduct $[M + Na]^+$ with m/z 750.26) (Figure 5D). This result confirms that AtGlcAT14B and AtGlcAT14E have GlcAT activity and a new GT14 member AT4G27480 (248) was also found to share this activity.

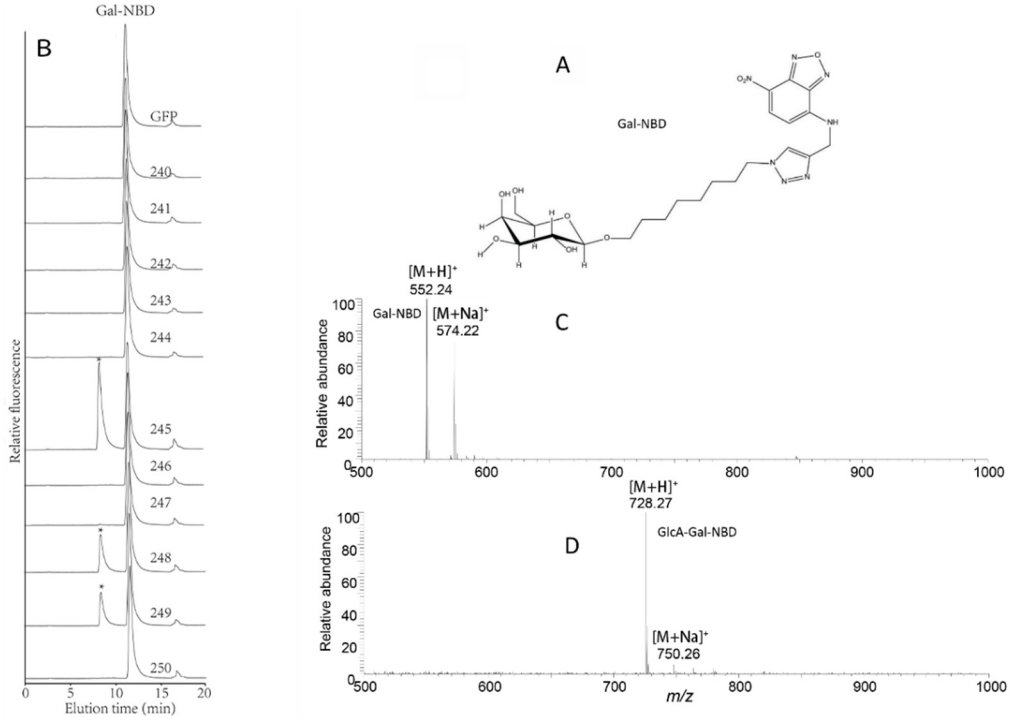


Figure 5. Catalytic activity of 11 *Arabidopsis* GT14 family members using UDP-GlcA as the donor and β-Gal-NBD as the acceptor. (**A**) The molecular structure of the acceptor β-Gal-NBD. (**B**) The microsomal membranes of the negative control (GFP) and the 11 recombinant AtGT14 family genes (240–250) expressed in *Nicotiana benthamiana* as enzymes were incubated with UDP-GlcA donor and β-Galactose-nitrobenzodiazole (β-Gal-NBD) acceptor for 2 h to test glucuronosyltransferase (GlcAT) activity. The reaction products were analyzed by RP-HPLC with fluorescence detection. The product peaks were labeled with an asterisk (*). The 11 recombinant AtGT14 family genes numbered as 240–250 (240: AT1G03520, 241: AT1G53100, 242: AT1G71070, 243: AT2G37585 (AtGlcAT14C), 244: AT3G03690, 245: AT3G15350 (AtGlcAT14E), 246: AT3G24040 (AtGlcAT14D), 247: AT4G03340, 248: AT4G27480, 249: AT5G15050 (AtGlcAT14B), 250: AT5G39990 (AtGlcAT14A)). The fluorescent tag NBD is detected by a fluorescence detector with λex = 470 nm, λem = 540 nm. (**C**) ESI-MS spectrum of the acceptor Gal-NBD. (**D**) The LC-ESI-MS spectrum of the reaction products catalyzed by the recombination AT3G15350 (245) microsomes.

In order to further characterize the biochemical activity of GT14 enzymes, the galacto-oligosaccharides β-1,6-Gal_4 and β-1,3-Gal_5 were fluorescently tagged with anthranilic acid (AA) and then used as acceptors in similar enzymatic assays using the GT14-containing MMs as enzyme source. RP-HPLC chromatograms showed that the AT3G15350 (245), AT4G27480 (248), and AT5G15050 (249) MMs had additional peaks compared to the negative control suggesting that the β-1,6-Gal_4-AA acceptor was able to be utilized by these AtGT14 enzymes, with MM extracts containing AT3G15350 (245) showing significantly

higher GlcAT activity than AT4G27480 (248) and AT5G15050 (249) (Figure 6A), similar to the pattern observed with the Gal-NBD acceptor (Figure 5B). The products of enzymatic assays in which β-1,3-Gal_5-AA was used as acceptor were analyzed in the same way, with AT3G15350 (245), AT4G27480 (248), AT5G15050 (249), and AT5G39990 (250) MM extracts shown to have GlcAT activity with the detectable glucuronosylated products eluting at 11.6 min and 14.1 min (Figure 6B). Thus, *Arabidopsis* GT14 family members including AT5G39990 (AtGlcAT14A), AT5G15050 (AtGlcAT14B), AT3G15350 (AtGlcAT14E), and AT4G27480 show the GlcAT activity using the three different acceptors, thereby verifying the previously reported activities of AtGlcAT14A, B and E and demonstrating that the family-wide approach led to the identification of another GT14 family member (AT4G27480) having similar GlcAT activity.

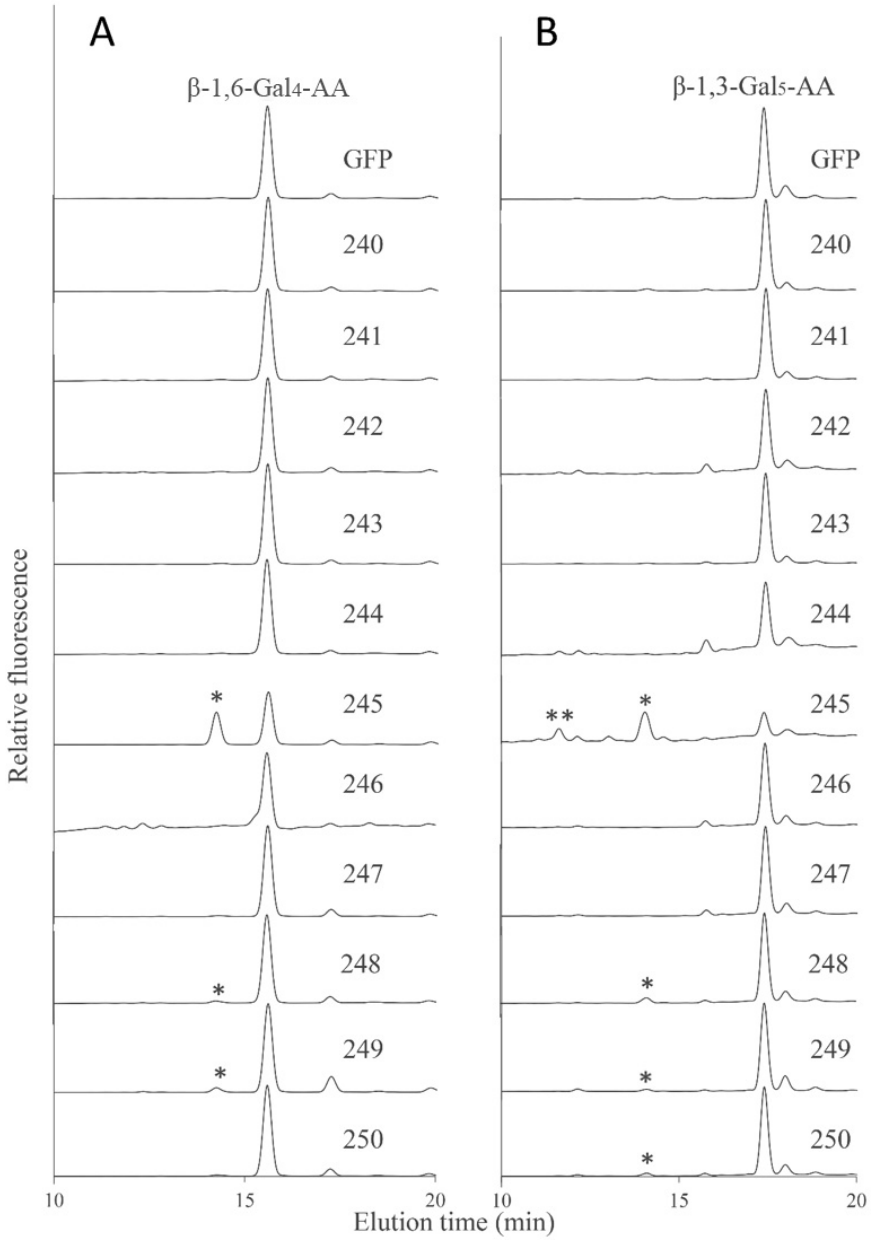


Figure 6. Glucuronosyltransferase (GlcAT) activity of the 11 recombinant AtGT14 family genes (240–250) using β-1,6-Gal_4-AA (**A**) and β-1,3-Gal_5-AA (**B**) as acceptors. The reaction products were analyzed by RP-HPLC using fluorescence detection (λex = 320 nm; λem = 420 nm) and different products eluting at 11.6 min and 14.1 min labeled with asterisks (* and **), respectively. GFP is the negative control. These results indicate that the gene 245 (AT3G15350) catalyzes an addition of GlcA residues to both β-1,3- and β-1,6-linked galactans.

2.6. Characterization of the Donor Substrate Specificity of the AtGT14 Recombinant Proteins

Among the 11 *Arabidopsis* GT14s, AT3G15350 (245, ATGlcAT14E) was demonstrated to have the highest GlcAT activity using Gal-containing acceptors and was therefore used in nucleotide sugar donor specificity assays. Seven additional nucleotide sugar donors (UDP-GalA, UDP-Glc, UDP-Gal, UDP-Ara*p*, GDP-Fuc, UDP-Rha, and UDP-Xyl) were tested in the presence of the β-1,3-Gal_5-AA acceptor. However, only UDP-GlcA was an active substrate donor, while other donors did not show any glycosylation activity (Figure 7A). The fractions from 10.6 to 12.6 min and 13 to 15 min were collected for analysis by LC ESI-MS. The peaks at 14.1 min and 11.6 min with the m/z values of 1126.37 (Figure 7B) and 1302.40 (Figure 7C) were the predicted pseudomolecular ions $[M + H]^+$ of GlcA-Gal_5-AA and $GlcA_2$-Gal_5-AA, respectively. This result indicates that up to two GlcA residues were incorporated by AtGlcAT14E onto the β-1,3-Gal_5-AA acceptor whereas addition of only one GlcA residue was incorporated using the β-1,6-Gal_4-AA acceptor.

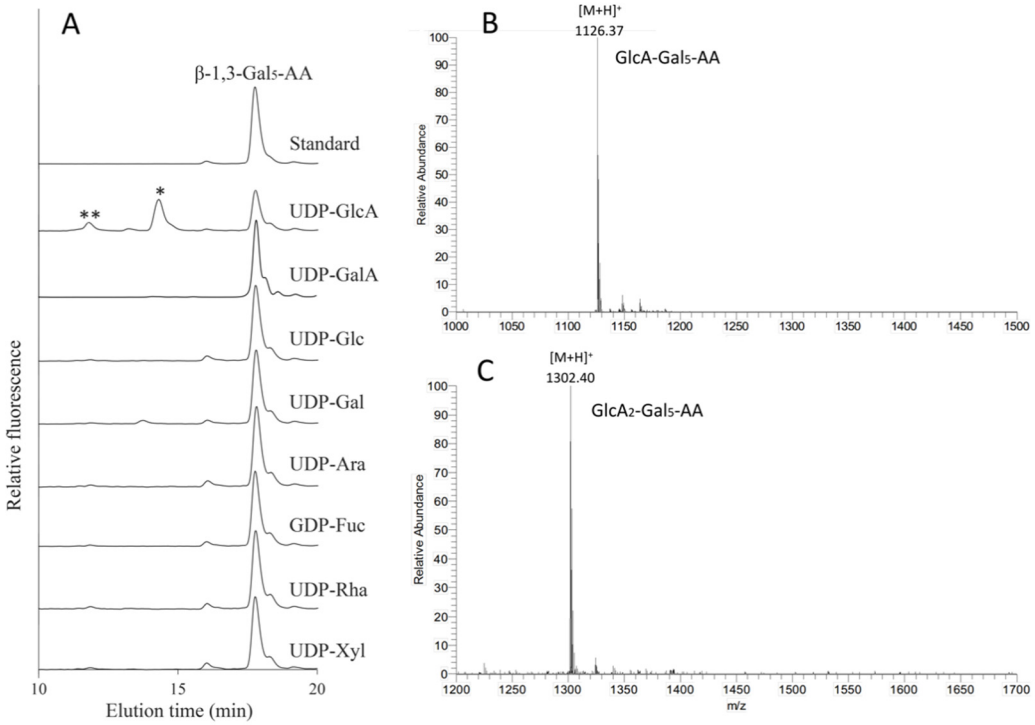


Figure 7. Donor specificity of AT3G15350 (245). (**A**) Eight NDP-sugars were tested for GT activity using *N. benthamiana* microsomes expressing AT3G15350 in the presence of the β-1,3-Gal_5-AA acceptor. The reaction products were analyzed by RP-HPLC using fluorescence detection (λex = 320 nm; λem = 420 nm) and are labeled with asterisks (* and **). The fraction between 13 and 15 min (*) (**B**) and the other fraction between 10.6 and 12.6 min (**). (**C**) collected for analysis by LC ESI-MS. From the molecular ion species we concluded that the peaks at 14.1 min and 11.6 min with the m/z values of 1126.37 (**B**) and 1302.40 (**C**) were the predicted pseudo-molecular ions $[M + H]^+$ of GlcA-Gal_5-AA and $GlcA_2$-Gal_5-AA, respectively.

3. Discussion

3.1. Plant GT Families Are Highly Diverse

Despite significant progress being made in the area of glycobiology, many questions still remain unresolved, especially in the plant field. Among the GTs classified into families, only 151 GTs have been annotated as characterized in *Arabidopsis* according to the CAZy database. In the human genome, only one GT is assigned as non-classified in comparison to 81 in *Arabidopsis*, hence indicating that glycosylation is a far more complex process in plants and further work is required to determine the biochemical activities of these GTs.

Among the *Arabidopsis* non-classified GTs, the DUF266 proteins are plant-specific. These proteins are distantly related to the GT14 family [62] and therefore they were named

as the GT14-like family [63]. Both DUF266 and GT14 enzymes share the similarity of a core 2 β-1,6-N-acetylglucosaminyltransferase C2/4GnT (C2GnT-2) from animal GTs, indicating that the GT14 and DUF266 families may have evolved from the same ancestral gene/s [63,64]. A few studies have used genetic tools to uncover the biological functions of DUF266 proteins. For example, the rice DUF266-containing protein BC10 regulates cell wall biosynthesis as the mutant shows a brittle stem phenotype [65] and another rice DUF266 protein RSE1 impacts leaf senescence [66]. Future analysis and biochemical characterization of members of the DUF266 family will provide further information on the functions of these proteins.

3.2. Evolution of GTs Correlated with Their Enzymatic Functions

The phylogenetic relationships of each *Arabidopsis* GT are reflected in Table S3 and Figure 2. While most GT families are clustered together forming one or more master blocks in the phylogenetic tree, there are few exceptions where several isolated genes from the same family showed a scattered pattern (GT1, 10, 28, 61, 106, 90, 47, 75, 77, 95, 32, 2, 37, 41, 66, 34, 92, 17, 48, 35, 14, 20) (Figure 2) [56,62,67,68].

Among these GT families, the GT1 family is the largest and the most ancient and its members are known to generally glycosylate small hydrophobic molecules. For example, UGT72Es are capable of conjugating lignin monomers and related metabolites (such as aldehydes, coniferyl, and sinapyl alcohols) [69]; UGT76C1, UGT76C2 and UGT85A1 are cytokinin-specific glycosyltransferases [70]; UGT78K1 and UGT78D2 are flavonoid 3-O-glucosyltransferases [71]; UGT80A2 and UGT80B1 are enzymes that catalyze the synthesis of steryl glycosides [72]. It should be noted that two GT28 members (AT1G73740 and AT4G16740) and two GT4 members (AT1G75420 and AT1G19710) are clustered with GT1 (Figure 3, Table S3) [73]. There are three genes in the GT10 family. AT1G49710 and AT3G19280 form a collinearity gene block encoding enzymes with the same biochemical function of α-1,3-L-FucTs, whereas the third gene (AT1G71990) encodes α-1,4-L-FucT (Figure 3, Table S3) [74,75], indicating the gene pairs sharing collinearity may have similar biological functions. Members in the GT106 family are annotated as O-FucTs and RG-I RhaTs. The functions of GT47 genes are annotated as exostosin (α-N-acetylglucosaminyltransferase) family proteins, except for AT5G03800 (Embryo Defective) [75] and AT3G57630 (Extensin Arabinose Deficient Transferase) [76]. The GT77 family can be divided into two main clades, Reduced Residual Arabinoses (RRAs) and Rhamnogalacturonan Xylosyltransferases (RGXT1), except for AT2G02060 (Homeodomain-Like Superfamily Protein) and AT2G35610 (Extensin Arabinosyltransferase) [76]. Noticeably, two 'orphan' GTs—one from the GT77 family AT1G70630 (Reduced Arabinose Yariv1, RAY1) [46] and the GT32 protein (AT4G19900)—are in clades separated from the rest of the family. The GT2 family is divided into two main clades with the exception of AT1G20575 [77] and AT2G39630 (Dolichol Phosphate Synthases), indicating the distinct origin of these GTs. The GT14 family will be discussed in detail below which is shown in Figure 2 (prominently in pink and marked with a red box) and in Figure 4A with three clades (A, B, and C).

For GT29, 8, 43, 31, 57, 22, 64, 5 and 4, the differences in the function of their members are predicted from protein annotation [56,62,68,78] and in part experimentally verified (Figure 2). The three members (AT1G08280 (AtGALT29A), AT1G08660 (MGD2), AT3G48820) of the GT29 family are separated on the phylogenetic tree, suggesting they have functionally diversified. Indeed, the annotation of three enzymes in the GT29 family are different according to the CAZy database although their exact biochemical functions are to be investigated. The GT8 family has 40 members with various glycosylation activities including xylan biosynthesis, sphingolipid glucuronosylation, and HG synthesis. The GT43 family is divided into two clades, including IRX 9 (AT1G27600, AT2G37090) and 14 (AT4G36890, AT5G67230) [79] both involved in xylan backbone biosynthesis with distinct roles [23]. The GT31 family has been characterized as GalTs and falls into three clades, consistent with the phylogenetic analysis [41].

3.3. The Expression System and Enzyme Assay for GT Biochemical Characterization

There are numerous expression systems that have been utilized to characterize GTs, including bacteria, the yeast *Pichia*, and the human cell line HEK 293 [80–82]. There is not one ideal expression system to characterize the biochemical functions of all GTs. However, many non-plant expression systems result in a lack of detectable enzymatic activity that may either be due to inappropriate post-translational modification and/or the absence of appropriate endogenous co-factors. In this study, we took advantage of the *N. benthamiana* (tobacco) transient expression system, which is one of the most efficient systems for expressing plant proteins [83]. YFP-tagged versions of the 11 AtGT14 family members were all able to be expressed at detectable levels in tobacco leaves using *Agrobacterium*-mediated transformation (Figure S2). Other GTs could also be expressed in the same system and verified with western blot to accumulate to detectable levels for the downstream of GT biochemical assays. The MMs prepared from infiltrated leaves were used directly in enzyme assays as crude enzyme preparations but if necessary, can be further purified using GFP-trap beads for activity assays and other purposes, such as identifying interacting proteins. The tobacco expression system can easily accommodate co-infiltration of multiple *Agrobacterium* vectors without the need to construct a complex multigene expression vector [23]. Therefore, the tobacco expression system can be used as a convenient and efficient method to screen the enzymatic activity of plant GTs.

The success of the assay in defining GT specificity depends on the precise combination of an appropriate nucleotide sugar donor, a specific molecule such as an oligosaccharide as an acceptor and an active GT preparation as an enzyme source. A particular bottleneck for the characterization of GTs lies in obtaining the right acceptors. Plant-derived carbohydrate oligosaccharides are usually complex and difficult to use directly in enzyme assays. For example, β-Galactosyl Yariv specifically recognizes AGPs and can be used in their purification [84]. However, as highly heterologous and complex macromolecules, AGPs are difficult to be directly used as acceptors for enzyme assays. Chemical synthesis provides a solution to this problem, although this approach is not straightforward. In future, the chemical synthesis of oligosaccharide fragments mimicking polysaccharides will be of great benefit for the identification of plant cell wall GTs [6].

In this study, we chose the *Arabidopsis* GT14 family to test the pipeline as multiple Gal-containing acceptors were available, were utilized in previous GT14 enzyme assays, and were shown to act as acceptors for multiple GT14 enzymes [44].

3.4. Members of the AtGT14 Family Display GlcAT Activity

AGPs are proteoglycans within the hydroxyproline-rich glycoprotein (HRGP) superfamily with type II arabino-3,6-galactans comprising up to 90% (*w*/*w*) of the molecule [85]. As a major part of AGPs, the type II AG moieties play a pivotal role in various plant growth and developmental processes and the biosynthesis of these complex structures requires multiple GTs [86]. AG sugar chains have a backbone of β-1,3-galactan, with the sidechains of β-1,6-galactan that have Ara, Rha, Fuc, and GlcA as branch terminal residues. Additionally, some GlcA residues are attached directly to the β-1,3-galactan backbone [86]. The GT31 family has been shown to be associated with the synthesis of the AGP galactan backbone and the synthesis of sidechains has been attributed to members of the GT14 (GlcAT), GT29 (β-(1,6)-GalT), GT37 (FucT), and GT77 families (AraT) [87]. Here, we provide evidence that at least one new member, AT4G27480 from the GT14 family, is a functional GlcAT and can add GlcA to both β-1,3- and β-1,6-galacto-oligosaccharides (Figure 6A,B, respectively) in addition to the Gal-NBD acceptor (Figure 5B). AT5G39990 (AtGlcAT14A) only incorporates GlcA onto β-1,3- but not β-1,6-galacto-oligosaccharides. Longer β-1,6-galacto-oligosaccharides could also be tested in the future if available. Notably, the MM extracts from plants expressing AtGlcAT14E had a greater activity than those from plants expressing either AtGlcAT14A, AtGlcAT14B, or AtGlcAT14C enzymes previously identified to possess GlcAT activity (Figure 6) [44,45]. This result suggested that AtGlcAT14E is a GlcAT, catalyzing the addition of GlcA residues onto β-1,3- and β-1,6-

linked galactans of AGPs. MM extracts from plants expressing AT4G27480 also showed some GlcAT activity with the same acceptors (Figure 6A,B), indicating it too is likely a GlcAT. Thus, our data expand the number of AtGT14 family members with demonstrated GlcAT activity to six.

In a previous study, AtGlcAT14C was shown to function as a GlcAT [45], a finding that is inconsistent with our results (Figure 6A). However, in vitro enzyme activity may vary between laboratories, expression systems, and enzyme assays including the availability of acceptors. Four proteins of clade C (including previously characterized AtGlcAT14A, B and E) demonstrated GlcAT activity using galactosyl-oligosaccharides as acceptors. Neither clade A nor clade B (including AtGlcAT14C and AtGlcAT14D) show any GlcAT activity, indicating enzymes from GT14 clade A and B may be involved in other glycosyltransferase activities. We have also expressed nine poplar GT14 genes and their enzymatic activity is consistent with the current *Arabidopsis* study (data not shown). It is somewhat surprising that no GlcAT activity was detected for AT1G53100 given the similarity of the sequences within clade C of the GT14 family (Figure 4A). A possible explanation may be that the level of this recombinant GT14 enzyme was insufficient to generate a detectable amount of reaction product or it needs additional co-factors such as other AGP biosynthesis protein complex components that are missing in tobacco leaf cells. Under the enzymatic conditions used in this study, clade A and B from the GT14 family did not demonstrate any GlcAT activity using the Gal-containing acceptors. It is possible that enzymes from these two clades of GT14 are involved in other biochemical functions that may require additional more complex acceptors to be tested. Alternatively, expressing these GTs without the C-terminal YFP tag may make these enzymes more active due to enhanced active site accessibility; it would, however, negate the ability to rapidly purify (and localise) these heterologously expressed proteins. Another possible strategy is to use a combination of GTs, e.g., co-expression of a GT31 GalT together with either a single or multiple GT14 members, as a means to reveal additional enzyme activities.

An AGP-Ca^{2+} model has been proposed whereby the negatively charged GlcA residues on AGPs are potential intramolecular Ca^{2+} binding sites on the plant cell surface that regulate cell wall integrity and are the source for intracellular Ca^{2+} signals required for the processes of plant growth and development [88]. Ca^{2+} has since been shown to bind to GlcA residues within the side chains of type II AG and that AG directly binds and releases Ca^{2+} in a pH-dependent manner [89]. Loss of function of the *GlcAT* genes *AtGlcAT14A*, *AtGlcAT14B*, and *AtGlcAT14C* led to significant reductions in Ca^{2+} binding on AGPs, which provided the first direct experimental evidence for the AGP-Ca^{2+} model [90]. Recently, Lopez-Hernandez et al. (2020) [60] analyzed higher-order mutant plants of four *Arabidopsis* GT14 family members—AtGlcAT14A, AtGlcAT14B, AtGlcAT14D, and AtGlcAT14E—to explore the impact on glucuronidation. The authors showed that in mutant plants, the GlcA decoration on AGs was similarly reduced and they bound less Ca^{2+} impacting their growth and development. Their findings suggest that the interaction between AG and Ca^{2+} is essential for several developmental processes in *Arabidopsis* and that GT14 family members play vital roles in intracellular Ca^{2+} signaling [60]. Our studies provide further key biochemical evidence for a role of a clade of GT14 family in glucuronidation of AGPs.

In conclusion, we have provided an update of the *Arabidopsis* GT clone collection and leveraged this resource and a plant expression system for the biochemical characterization of novel GTs using the GT14 family as the exemplar. This analysis pipeline can accelerate the functional analysis of classified and non-classified GTs and help scientists to unravel the complex glycosylation processes that occur in plants.

4. Materials and Methods

4.1. Materials

The activated nucleotide sugar donors: UDP-glucuronic acid (UDP-GlcA), GDP-fucose (GDP-Fuc), UDP-glucose (UDP-Glc), UDP-galactose (UDP-Gal) were purchased from Sigma Aldrich (www.sigmaaldrich.com). UDP-Galacturonic acid (UDP-GalA), UDP-

xylose (UDP-Xyl), and UDP-arabinopyranose (UDP-Ara*p*) were provided by CarboSource (Complex Carbohydrate Research Center (CCRC), Athens, GA, USA). UDP-Rhamnose (UDP-Rha) was enzymatically synthesized according to Rautengarten et al. (2014) [91]. The synthetic acceptor galactose-nitrobenzodiazole (Gal-NBD) was chemically synthesized according to McGill and Williams (2009) [61]. The galacto-oligosaccharides, β-1,6-galactotetraose (β-1,6-Gal_4) and β-1,3-galactopentose (β-1,3-Gal_5), were chemically synthesized according to Andersen et al. (2017) [92].

Nicotiana benthamiana (tobacco) was cultivated in a controlled environment growth room at 24 °C with a 16 h photoperiod. The leaves of 4–6 weeks old tobacco seedlings were used for the transient expression of GTs.

4.2. Amino Acid Sequence Alignment and Phylogenetic Analysis

Multiple alignments for the full-length amino acid sequences of GTs in the published *Arabidopsis* genome Araport11 [78] were performed using MAFFT (V7) [67,93] with default settings. Maximum likelihood (ML) phylogenetic trees were constructed using IQ-TREE [94] on XSEDE (V1.6.10) [95] with Model Finder [96] and 1000 bootstrap replicates for support values of SH-aLRT, local bootstrap, parametric aLRT, aBayes, and ultrafast bootstrap [97]. The trees were annotated using iTOL [98].

4.3. Analysis of Genome Collinearity and Chromosomal Localizations of GT Family Genes

The genome collinear blocks were analyzed using MCScanX [58] and TBtools (V1.0) [99], treating at least five genes as a collinear block. Each GT gene was mapped to a chromosome. Genome collinearity and chromosomal locations were drawn using TBtools and Inkscape (http://www.inkscape.org/).

4.4. Publicly Available Microarray Data and Sequence Architecture Analysis for the Arabidopsis GT14 Family

The public microarray data for various tissues are available at NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The series accession numbers GSE34188 [59] of *Arabidopsis* from Agilent platform includes 14 different tissue samples representing three biological replicates were used for the tissue-specific expression analysis. Hybridization intensities in arrays were normalized among different arrays by quantile normalization followed by denary logarithm transformation. The data of probe set IDs corresponding to *Arabidopsis* GT14 family genes were extracted for further analyses. The GT14 gene structures were obtained using the Bath Web CD-Search Tool [100] on NCBI and the protein motifs (novel and ungapped) were found using MEME [58]. The motifs refer to the approximate sequence pattern that occurs in a group of homolog genes [100]. The gene expression and structure profiles were annotated using TBtools [99], iTOL [98], and Inkscape (http://www.inkscape.org/).

4.5. Arabidopsis GT Cloning

Additional GT coding DNA sequences (CDS) were cloned from *Arabidopsis thaliana* (At, Col-0) into pDONR223 according to Lao et al. (2014) [8]. An LR reaction was used to subclone each CDS into the Gateway-compatible pEarleyGate 101 binary vector containing the cauliflower mosaic virus 35S promoter and upstream enhancer and C-terminal YFP and HA tags [101]. All recombinant vectors were sequenced to verify correct assembly.

4.6. Heterologous Expression of GT14 Family Proteins in Tobacco

The eleven pEarleyGate 101 binary vector constructs containing each of the AtGT14 cDNAs were individually transformed into *Agrobacterium tumefaciens* strain AGL1 via electroporation. Single colonies were then cultured for transient plant expression. The abaxial surface of *N. benthamiana* leaves was coinfiltrated with a mix of *Agrobacterium* cells carrying either the GT14 construct or the P19 gene from the tomato bushy stunt virus to suppress gene silencing. The final OD600 of each strain was adjusted to 0.8 in

resuspension buffer (0.01 M $MgCl_2$ and 0.8 mM acetosyringone in 50 mM MES buffer pH 5.3) prior to the infiltration. After 3 days, the infected leaf areas were excised, and the fluorescence confirmed under a confocal microscope (Zeiss LSM 880, Jena, Germany). Microsomal membranes (MMs) were extracted as described by Geshi et al. (2002) [102]. In brief, about 5 g of leaves was ground using a mortar and pestle in 10 mL of chilled extraction buffer (0.1 M HEPES/KOH pH 6.8, 0.4 M sucrose, 5 mM $MgCl_2$, 5 mM $MnCl_2$, 1 mM phenylmethylsulfonyl fluoride; Roche EDTA-free complete protease inhibitor) and quartz sand. The homogenate was filtered through two layers of Miracloth and the filtrate was centrifuged at 3000× *g* for 20 min at 4 °C. The supernatant was then centrifuged at 30,000× *g* for 30 min at 4 °C. The MM pellets were solubilized in extraction buffer to a final concentration of 5 mg/mL and used in biochemical assays. Protein concentrations were determined with a Bradford assay using Protein Assay Dye Reagent (Bio-Rad, Hercules, CA, USA).

4.7. Protein Analysis

For western blot analysis, 5 µL of microsomal proteins (around 25 µg) were denatured by mixing with 15 µL 2× Laemmli sample buffer (Bio-Rad) containing 50% (*v*/*v*) β-mercaptoethanol and incubating at 70 °C for 20 min. The protein samples were then subjected to SDS-PAGE (Bio-Rad) and electroblotted onto a PVDF immobilon membrane (Merck Millipore, Billerica, MA, USA) using a Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked for 15 min at room temperature using TBST (Tris-buffered saline with 0.1% (*v*/*v*) Tween 20) containing 0.5% (*w*/*v*) skim milk powder filtered through a 0.22 µm filter and incubated with an anti-GFP primary antibody (MBL, Nagoya, Japan) at a 1:1000 dilution followed by a horseradish peroxidase (HRP)-conjugated secondary goat anti-rabbit IgG antibody (Proteintech, Rosemont, IL, USA) at a 1:1000 dilution. Blocking and antibody binding were performed using the SNAP i.d.® 2.0 Protein Detection System. Afterwards, WesternBright ECL HRP substrate (Advansta, San Jose, CA, USA) was added to the membranes for 5 min in the dark. The HRP signal was detected by chemiluminescence using a ChemiDoc MP imaging system (Bio-Rad) and processed using the Image Lab software (Bio-Rad). *N. benthamiana* leaves infiltrated with a GFP construct were used as negative control.

4.8. Anthranilic Acid Labeling of Galacto-Oligosaccharides

The synthetic acceptor β-1,6-Gal_4 and β-1,3-Gal_5 were labeled with anthranilic acid (AA) at their reducing ends according to the method described by Alwael et al. [103]. Excess derivatization reagent AA was removed with diethyl ether three times and the AA-labeled galacto-oligosaccharides (β-1,6-Gal_4-AA and β-1,3-Gal_5-AA) in the lower layer separately purified using a Sep-Pak C18 cartridge (Waters, Milford, MA, USA).

4.9. Enzyme Assays

The *N. benthamiana* MMs were pretreated with 1% (*v*/*v*) Triton X-100 on ice prior to use them in biochemical activity assays. The reaction was performed in a total volume of 50 µL containing 5 mg/mL MMs, containing either 5 µM β-Gal-NBD or AA-labeled acceptor (β-1,6-Gal_4-AA or β-1,3-Gal_5-AA), 0.4 mM nucleotide sugar as a donor, and the reaction incubated at 25 °C for 2 h under constant shaking. Reactions were terminated by adding 1 µL glacial acetic acid and 1 µL 0.5 M EDTA. The reaction mixture was then centrifuged for 1 min at 13,040× *g* and the supernatant filtered using a 0.22 µm Ultrafree-MC filter (Merck Millipore, Billerica, MA, USA). The filtrate was separated on a reversed-phase (RP)-HPLC (Agilent HC-C18 Analytical 4.6 × 250 mm, 5 µm) with a fluorescence detector, using a gradient of acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA): 30% ACN, 15 min; 80% ACN, 5 min; 100% ACN, 5 min; 30% ACN, 5 min for re-equilibrating at a constant flow rate (0.5 mL/min) for NBD-labeled samples (λex = 470 nm, λem = 540 nm), and a different gradient of ACN/0.1% TFA: 5–10% ACN, 10 min; 10–45% ACN, 5 min; 45% ACN,

15 min; 100% ACN, 5 min; 5% ACN, 10 min for re-equilibrating at a constant flow rate (0.5 mL/min) for AA-labeled samples (λex = 320 nm, λem = 420 nm).

4.10. Analysis of the Enzymatic Products by Mass Spectrometry

The GlcAT assay products with β-Gal-NBD and β-1,3-Gal_5-AA as acceptors were fractionated and collected by RP-HPLC. The products were concentrated by a rotation vacuum concentrator (Christ RVC 2–25 CDplus, Osterode, Germany) and introduced into the ultra-performance liquid chromatography-time-of-flight mass spectrometer (Synapt-G2-Si, Waters, Milford, MA, USA) equipped with electro-spray ionization (ESI) source. The samples were directly infused through 20% mobile phase A (water, 0.1% formic acid) and 80% mobile phase B (ACN, 0.1% formic acid) at 0.2 mL/min. ESI-MS/MS spectra were operated in positive ion mode with a high voltage of 4000 V for ionization, 40 V and for sampling cone and 80 V for source offset, respectively. The source temperature was set to 100 °C and the samples were dried with high nitrogen under the temperature of 400 °C and the flow rate of 900 L/h. For MS analysis, the samples were subjected to a full scan from m/z 400 to 2000 Da and from 100 Da to 1500 Da in MS^2 test with collision energy fixed at 32 eV.

Supplementary Materials: The following are available online at https://www.mdpi.com/1422-0067/22/3/1360/s1, Figure S1: Chromosome location of *Arabidopsis* GT family genes. Genes from different GT families are located on chromosomes, using family-specific colors defined in Figure 2. The number of genes contained within each annotated block is indicated after the decimal point. The color of the chromosome strip shows the gene density at 100 kb intervals, lower for cool colors and higher for warm colors; Figure S2: Western-blotting analysis of the 11 recombinant AtGT14 family genes expressed in *Nicotiana benthamiana*. Lane GFP, 5 mg/mL MMs extracted from *N. benthamiana* expressing empty pEarley101 vector with GFP tag (27 kDa, 25 µg total protein). Lanes 240–250, 25 µg microsomal samples of the 11 recombinant AtGT14 family genes (240: AT1G03520, 241: AT1G53100, 242: AT1G71070, 243: AT2G37585, 244: AT3G03690, 245: AT3G15350, 246: AT3G24040, 247: AT4G03340, 248: AT4G27480, 249: AT5G15050, 250: AT5G39990); Table S1: Selected examples of *Arabidopsis* cell wall biosynthetic GTs whose catalytic specificities have been analyzed following expression in heterologous systems; Table S2: The collection of *Arabidopsis* glycosyltransferases according to CAZy (http://www.cazy.org/) and cross-referenced with the *Arabidopsis* sequences from TAIR10 (http://www.Arabidopsis.org/), Table S3: GT family genes belonging to collinearity blocks. The gene pairs are divided into two groups colored yellow and blue, and they correspond to each other in the co-linearity block. The rightmost column shows the RGB colors of their connecting lines in Figures 2 and 3. All *A. thaliana* co-linearity protein pairs are displayed in second worksheet in this table.

Author Contributions: Conceived and designed the experiments: W.Z., L.X., and J.Z.; Performed the experiments: L.X., J.Z., and W.L.; Analyzed data: L.X., W.L., and W.Z., Writing—original draft preparation: W.Z., L.X., J.Z., W.L., A.B., M.S.D., and J.L.H.; Writing—review and editing: P.G., B.E., Y.Z., T.K., M.L., M.H.C., K.L.J. and L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant no. 31971619) and the Overseas Expertise Introduction Project for Discipline Innovation (111 Project D18008) and by grants from Australia Research Council to the ARC Centre of Excellence in Plant Cell Walls (CE1101007) to A.B., M.S.D., K.L.J. and W.Z. M.H.C. acknowledges funding from the Novo Nordisk Foundation (grant nos. NNF18OC0053048 and NNF20OC0065094) and the Carlsberg Foundation (grant no. CF19-0072).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors would like to thank B.E. who is supported by an Australian Research Council Future Fellowship FT160100276 and the Instrumentation and Service Center for Molecular Sciences at Westlake University for technical assistance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hansen, S.F.; Bettler, E.; Rinnan, A.; Engelsen, S.B.; Breton, C. Exploring genomes for glycosyltransferases. *Mol. Biosyst.* **2010**, *6*, 1773–1781. [CrossRef] [PubMed]
2. Lampugnani, E.R.; Khan, G.A.; Somssich, M.; Persson, S. Building a plant cell wall at a glance. *J. Cell Sci.* **2018**, *131*, jcs207373. [CrossRef] [PubMed]
3. Merzendorfer, H.; Zimoch, L. Chitin metabolism in insects: Structure, function and regulation of chitin synthases and chitinases. *J. Exp. Biol.* **2003**, *206*, 4393–4412. [CrossRef] [PubMed]
4. Jozwiak, A.; Sonawane, P.D.; Panda, S.; Garagounis, C.; Papadopoulou, K.K.; Abebie, B.; Massalha, H.; Almekias-Siegl, E.; Scherf, T.; Aharoni, A. Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery. *Nat. Chem. Biol.* **2020**, *16*, 740–748. [CrossRef] [PubMed]
5. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef]
6. Ruprecht, C.; Bartetzko, M.P.; Senf, D.; Lakhina, A.; Smith, P.J.; Soto, M.J.; Oh, H.; Yang, J.Y.; Chapla, D.; Varon Silva, D.; et al. A glycan array-based assay for the identification and characterization of plant glycosyltransferases. *Angew. Chem. Int. Ed.* **2020**, *59*, 12493–12498. [CrossRef]
7. Hansen, L.; Lind-Thomsen, A.; Joshi, H.J.; Pedersen, N.B.; Have, C.T.; Kong, Y.; Wang, S.; Sparso, T.; Grarup, N.; Vester-Christensen, M.B.; et al. A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology* **2015**, *25*, 211–224. [CrossRef]
8. Lao, J.; Oikawa, A.; Bromley, J.R.; McInerney, P.; Suttangkakul, A.; Smith-Moritz, A.M.; Plahar, H.; Chiu, T.Y.; González Fernández-Niño, S.M.; Ebert, B.; et al. The plant glycosyltransferase clone collection for functional genomics. *Plant J.* **2014**, *79*, 517–529. [CrossRef]
9. Caputi, L.; Malnoy, M.; Goremykin, V.; Nikiforova, S.; Martens, S. A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* **2012**, *69*, 1030–1042. [CrossRef]
10. Polko, J.K.; Kieber, J.J. The regulation of cellulose biosynthesis in plants. *Plant Cell* **2019**, *31*, 282–296. [CrossRef]
11. Speicher, T.L.; Li, P.Z.; Wallace, I.S. Phosphoregulation of the plant cellulose synthase complex and cellulose synthase-like proteins. *Plants* **2018**, *7*, 52. [CrossRef] [PubMed]
12. Driouich, A.; Follet-Gueye, M.L.; Bernard, S.; Kousar, S.; Chevalier, L.; Vicré-Gibouin, M.; Lerouxel, O. Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. *Front. Plant Sci.* **2012**, *3*, 79. [CrossRef] [PubMed]
13. Cocuron, J.C.; Lerouxel, O.; Drakakaki, G.; Alonso, A.P.; Liepman, A.H.; Keegstra, K.; Raikhel, N.; Wilkerson, C.G. A gene from the cellulose synthase-like C family encodes a β-1,4 glucan synthase. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8550–8555. [CrossRef] [PubMed]
14. Kim, S.J.; Chandrasekar, B.; Rea, A.C.; Danhof, L.; Zemelis-Durfee, S.; Thrower, N.; Shepard, Z.S.; Pauly, M.; Brandizzi, F.; Keegstra, K. The synthesis of xyloglucan, an abundant plant cell wall polysaccharide, requires CSLC function. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 20316–20324. [CrossRef] [PubMed]
15. Faik, A.; Price, N.J.; Raikhel, N.V.; Keegstra, K. An *Arabidopsis* gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7797–7802. [CrossRef] [PubMed]
16. Madson, M.; Dunand, C.; Li, X.; Verma, R.; Vanzin, G.F.; Caplan, J.; Shoue, D.A.; Carpita, N.C.; Reiter, W.D. The MUR3 gene of *Arabidopsis* encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins. *Plant Cell* **2003**, *15*, 1662–1670. [CrossRef]
17. Liang, Y.; Basu, D.; Pattathil, S.; Xu, W.L.; Venetos, A.; Martin, S.L.; Faik, A.; Hahn, M.G.; Showalter, A.M. Biochemical and physiological characterization of *fut4* and *fut6* mutants defective in arabinogalactan-protein fucosylation in *Arabidopsis*. *J. Exp. Bot.* **2013**, *64*, 5537–5551. [CrossRef] [PubMed]
18. Tryfona, T.; Theys, T.E.; Wagner, T.; Stott, K.; Keegstra, K.; Dupree, P. Characterisation of FUT4 and FUT6 α-(1 → 2)-fucosyltransferases reveals that absence of root arabinogalactan fucosylation increases *Arabidopsis* root growth salt sensitivity. *PLoS ONE* **2014**, *9*, e93291. [CrossRef]
19. Wu, Y.; Williams, M.; Bernard, S.; Driouich, A.; Showalter, A.M.; Faik, A. Functional identification of two nonredundant *Arabidopsis* α(1,2)fucosyltransferases specific to arabinogalactan proteins. *J. Biol. Chem.* **2010**, *285*, 13638–13645. [CrossRef]
20. Scheller, H.V.; Ulvskov, P. Hemicelluloses. *Annu. Rev. Plant Biol.* **2010**, *61*, 263–289. [CrossRef]
21. Wu, A.M.; Hornblad, E.; Voxeur, A.; Gerber, L.; Rihouey, C.; Lerouge, P.; Marchant, A. Analysis of the *Arabidopsis* IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveals critical contributions to biosynthesis of the hemicellulose glucuronoxylan. *Plant Physiol.* **2010**, *153*, 542–554. [CrossRef] [PubMed]
22. Jensen, J.K.; Johnson, N.R.; Wilkerson, C.G. *Arabidopsis thaliana* IRX10 and two related proteins from psyllium and *Physcomitrella patens* are xylan xylosyltransferases. *Plant J.* **2014**, *80*, 207–215. [CrossRef] [PubMed]
23. Zeng, W.; Lampugnani, E.R.; Picard, K.L.; Song, L.; Wu, A.M.; Farion, I.M.; Zhao, J.; Ford, K.; Doblin, M.S.; Bacic, A. *Asparagus* IRX9, IRX10, and IRX14A are components of an active xylan backbone synthase complex that forms in the Golgi apparatus. *Plant Physiol.* **2016**, *171*, 93–109. [CrossRef] [PubMed]
24. Keppler, B.D.; Showalter, A.M. IRX14 and IRX14-LIKE, two glycosyl transferases involved in glucuronoxylan biosynthesis and drought tolerance in *Arabidopsis*. *Mol. Plant* **2010**, *3*, 834–841. [CrossRef] [PubMed]

25. Jensen, J.K.; Busse-Wicher, M.; Poulsen, C.P.; Fangel, J.U.; Smith, P.J.; Yang, J.Y.; Pena, M.J.; Dinesen, M.H.; Martens, H.J.; Melkonian, M.; et al. Identification of an algal xylan synthase indicates that there is functional orthology between algal and plant cell wall biosynthesis. *New Phytol.* **2018**, *218*, 1049–1060. [CrossRef]
26. Rennie, E.A.; Ebert, B.; Miles, G.P.; Cahoon, R.E.; Christiansen, K.M.; Stonebloom, S.; Khatab, H.; Twell, D.; Petzold, C.J.; Adams, P.D.; et al. Identification of a sphingolipid α-glucuronosyltransferase that is essential for pollen function in *Arabidopsis*. *Plant Cell* **2014**, *26*, 3314–3325. [CrossRef]
27. Rennie, E.A.; Hansen, S.F.; Baidoo, E.E.; Hadi, M.Z.; Keasling, J.D.; Scheller, H.V. Three members of the *Arabidopsis* glycosyltransferase family 8 are xylan glucuronosyltransferases. *Plant Physiol.* **2012**, *159*, 1408–1417. [CrossRef]
28. Anders, N.; Wilkinson, M.D.; Lovegrove, A.; Freeman, J.; Tryfona, T.; Pellny, T.K.; Weimar, T.; Mortimer, J.C.; Stott, K.; Baker, J.M.; et al. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 989–993. [CrossRef]
29. Liepman, A.H.; Cavalier, D.M. The CELLULOSE SYNTHASE-LIKE A and CELLULOSE SYNTHASE-LIKE C families: Recent advances and future perspectives. *Front. Plant Sci.* **2012**, *3*, 109. [CrossRef]
30. Ridley, B.L.; O'Neill, M.A.; Mohnen, D. Pectins: Structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **2001**, *57*, 929–967. [CrossRef]
31. Sterling, J.D.; Atmodjo, M.A.; Inwood, S.E.; Kumar Kolli, V.S.; Quigley, H.F.; Hahn, M.G.; Mohnen, D. Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5236–5241. [CrossRef] [PubMed]
32. Atmodjo, M.A.; Sakuragi, Y.; Zhu, X.; Burrell, A.J.; Mohanty, S.S.; Atwood, J.A.; Orlando, R.; Scheller, H.V.; Mohnen, D. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan: Galacturonosyltransferase complex. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20225–20230. [CrossRef] [PubMed]
33. Takenaka, Y.; Kato, K.; Ogawa-Ohnishi, M.; Tsuruhama, K.; Kajiura, H.; Yagyu, K.; Takeda, A.; Takeda, Y.; Kunieda, T.; Hara-Nishimura, I.; et al. Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nat. Plants* **2018**, *4*, 669–676. [CrossRef] [PubMed]
34. Laursen, T.; Stonebloom, S.H.; Pidatala, V.R.; Birdseye, D.S.; Clausen, M.H.; Mortimer, J.C.; Scheller, H.V. Bifunctional glycosyltransferases catalyze both extension and termination of pectic galactan oligosaccharides. *Plant J.* **2018**, *94*, 340–351. [CrossRef] [PubMed]
35. Liwanag, A.J.; Ebert, B.; Verhertbruggen, Y.; Rennie, E.A.; Rautengarten, C.; Oikawa, A.; Andersen, M.C.; Clausen, M.H.; Scheller, H.V. Pectin biosynthesis: GALS1 in *Arabidopsis thaliana* is a β-1,4-galactan β-1,4-galactosyltransferase. *Plant Cell* **2012**, *24*, 5024–5036. [CrossRef]
36. Ebert, B.; Birdseye, D.; Liwanag, A.J.M.; Laursen, T.; Rennie, E.A.; Guo, X.; Catena, M.; Rautengarten, C.; Stonebloom, S.H.; Gluza, P.; et al. The three members of the *Arabidopsis* glycosyltransferase family 92 are functional β-1,4-galactan synthases. *Plant Cell Physiol.* **2018**, *59*, 2624–2636. [CrossRef]
37. Egelund, J.; Petersen, B.L.; Motawia, M.S.; Damager, I.; Faik, A.; Olsen, C.E.; Ishii, T.; Clausen, H.; Ulvskov, P.; Geshi, N. *Arabidopsis thaliana* RGXT1 and RGXT2 encode Golgi-localized (1,3)-α-D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan-II. *Plant Cell* **2006**, *18*, 2593–2607. [CrossRef]
38. Suzuki, T.; Narciso, J.O.; Zeng, W.; van de Meene, A.; Yasutomi, M.; Takemura, S.; Lampugnani, E.R.; Doblin, M.S.; Bacic, A.; Ishiguro, S. KNS4/UPEX1: A type II arabinogalactan β-(1,3)-galactosyltransferase required for pollen exine development. *Plant Physiol.* **2016**, *173*, 183–205. [CrossRef]
39. Harholt, J.; Jensen, J.K.; Verhertbruggen, Y.; Sogaard, C.; Bernard, S.; Nafisi, M.; Poulsen, C.P.; Geshi, N.; Sakuragi, Y.; Driouich, A.; et al. ARAD proteins associated with pectic Arabinan biosynthesis form complexes when transiently overexpressed in planta. *Planta* **2012**, *236*, 115–128. [CrossRef]
40. Ndeh, D.; Rogowski, A.; Cartmell, A.; Luis, A.S.; Baslé, A.; Gray, J.; Venditto, I.; Briggs, J.; Zhang, X.; Labourel, A.; et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **2017**, *544*, 65–70. [CrossRef]
41. Qu, Y.; Egelund, J.; Gilson, P.R.; Houghton, F.; Gleeson, P.A.; Schultz, C.J.; Bacic, A. Identification of a novel group of putative *Arabidopsis thaliana* β-(1,3)-galactosyltransferases. *Plant Mol. Biol.* **2008**, *68*, 43–59. [CrossRef] [PubMed]
42. Basu, D.; Tian, L.; Wang, W.; Bobbs, S.; Herock, H.; Travers, A.; Showalter, A.M. A small multigene hydroxyproline-*O*-galactosyltransferase family functions in arabinogalactan-protein glycosylation, growth and development in *Arabidopsis*. *BMC Plant Biol.* **2015**, *15*, 295. [CrossRef] [PubMed]
43. Ogawa-Ohnishi, M.; Matsubayashi, Y. Identification of three potent hydroxyproline *O*-galactosyltransferases in *Arabidopsis*. *Plant J.* **2015**, *81*, 736–746. [CrossRef]
44. Knoch, E.; Dilokpimol, A.; Tryfona, T.; Poulsen, C.P.; Xiong, G.; Harholt, J.; Petersen, B.L.; Ulvskov, P.; Hadi, M.Z.; Kotake, T.; et al. A β-glucuronosyltransferase from *Arabidopsis thaliana* involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth. *Plant J.* **2013**, *76*, 1016–1029. [CrossRef] [PubMed]
45. Dilokpimol, A.; Geshi, N. *Arabidopsis thaliana* glucuronosyltransferase in family GT14. *Plant Signal Behav.* **2014**, *9*, e28891. [CrossRef] [PubMed]
46. Gille, S.; Sharma, V.; Baidoo, E.E.; Keasling, J.D.; Scheller, H.V.; Pauly, M. Arabinosylation of a Yariv-precipitable cell wall polymer impacts plant growth as exemplified by the *Arabidopsis* glycosyltransferase mutant *ray1*. *Mol. Plant* **2013**, *6*, 1369–1372. [CrossRef] [PubMed]

47. Reski, R. *Physcomitrella* and *Arabidopsis*: The David and Goliath of reverse genetics. *Trends Plant Sci.* **1998**, *3*, 209–210. [CrossRef]
48. Kondou, Y.; Higuchi, M.; Matsui, M. High-throughput characterization of plant gene functions by using gain-of-function technology. *Annu. Rev. Plant Biol.* **2010**, *61*, 373–393. [CrossRef]
49. Basu, D.; Liang, Y.; Liu, X.; Himmeldirk, K.; Faik, A.; Kieliszewski, M.; Held, M.; Showalter, A.M. Functional identification of a hydroxyproline-*O*-galactosyltransferase specific for arabinogalactan protein biosynthesis in *Arabidopsis*. *J. Biol. Chem.* **2013**, *288*, 10132–10143. [CrossRef]
50. Voiniciuc, C.; Dama, M.; Gawenda, N.; Stritt, F.; Pauly, M. Mechanistic insights from plant heteromannan synthesis in yeast. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 522–527. [CrossRef]
51. Perrin, R.M.; DeRocher, A.E.; Bar-Peled, M.; Zeng, W.; Norambuena, L.; Orellana, A.; Raikhel, N.V.; Keegstra, K. Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall biosynthesis. *Science* **1999**, *284*, 1976–1979. [CrossRef] [PubMed]
52. Wilson, S.M.; Ho, Y.Y.; Lampugnani, E.R.; Van de Meene, A.M.; Bain, M.P.; Bacic, A.; Doblin, M.S. Determining the subcellular location of synthesis and assembly of the cell wall polysaccharide (1,3; 1,4)-β-D-glucan in grasses. *Plant Cell* **2015**, *27*, 754–771. [CrossRef] [PubMed]
53. Kim, S.J.; Zemelis, S.; Keegstra, K.; Brandizzi, F. The cytoplasmic localization of the catalytic site of CSLF6 supports a channeling model for the biosynthesis of mixed-linkage glucan. *Plant J.* **2015**, *81*, 537–547. [CrossRef] [PubMed]
54. Jobling, S.A. Membrane pore architecture of the CslF6 protein controls (1-3,1-4)-β-glucan structure. *Sci. Adv.* **2015**, *1*, e1500069. [CrossRef]
55. Nikolovski, N.; Rubtsov, D.; Segura, M.P.; Miles, G.P.; Stevens, T.J.; Dunkley, T.P.; Munro, S.; Lilley, K.S.; Dupree, P. Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. *Plant Physiol.* **2012**, *160*, 1037–1051. [CrossRef]
56. Hansen, S.F.; Harholt, J.; Oikawa, A.; Scheller, H.V. Plant glycosyltransferases beyond CAZy: A perspective on DUF families. *Front. Plant Sci.* **2012**, *3*, 59. [CrossRef]
57. Stonebloom, S.; Ebert, B.; Xiong, G.; Pattathil, S.; Birdseye, D.; Lao, J.; Pauly, M.; Hahn, M.G.; Heazlewood, J.L.; Scheller, H.V. A DUF-246 family glycosyltransferase-like gene affects male fertility and the biosynthesis of pectic arabinogalactans. *BMC Plant Biol.* **2016**, *16*, 90. [CrossRef]
58. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef]
59. Hanada, K.; Higuchi-Takeuchi, M.; Okamoto, M.; Yoshizumi, T.; Shimizu, M.; Nakaminami, K.; Nishi, R.; Ohashi, C.; Iida, K.; Tanaka, M.; et al. Small open reading frames associated with morphogenesis are hidden in plant genomes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2395–2400. [CrossRef]
60. Lopez-Hernandez, F.; Tryfona, T.; Rizza, A.; Yu, X.L.; Harris, M.O.B.; Webb, A.A.R.; Kotake, T.; Dupree, P. Calcium binding by arabinogalactan polysaccharides is important for normal plant development. *Plant Cell* **2020**, *32*, 3346–3369. [CrossRef]
61. McGill, N.W.; Williams, S.J. 2,6-Disubstituted benzoates as neighboring groups for enhanced diastereoselectivity in β-galactosylation reactions: Synthesis of β-1,3-linked oligogalactosides related to arabinogalactan proteins. *J. Org. Chem.* **2009**, *74*, 9388–9398. [CrossRef] [PubMed]
62. Berardini, T.Z.; Reiser, L.; Li, D.; Mezheritsky, Y.; Muller, R.; Strait, E.; Huala, E. The *Arabidopsis* information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* **2015**, *53*, 474–485. [CrossRef] [PubMed]
63. Ye, C.Y.; Li, T.; Tuskan, G.A.; Tschaplinski, T.J.; Yang, X. Comparative analysis of GT14/GT14-like gene family in *Arabidopsis*, *Oryza*, *Populus*, *Sorghum* and *Vitis*. *Plant Sci.* **2011**, *181*, 688–695. [CrossRef]
64. Beum, P.V.; Basma, H.; Bastola, D.R.; Cheng, P.W. Mucin biosynthesis: Upregulation of core 2 beta 1,6 N-acetylglucosaminyltransferase by retinoic acid and Th2 cytokines in a human airway epithelial cell line. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2005**, *288*, L116–L124. [CrossRef] [PubMed]
65. Zhou, Y.; Li, S.; Qian, Q.; Zeng, D.; Zhang, M.; Guo, L.; Liu, X.; Zhang, B.; Deng, L.; Liu, X.; et al. BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (*Oryza sativa L.*). *Plant J.* **2009**, *57*, 446–462. [CrossRef]
66. Lee, S.; Kim, M.H.; Lee, J.H.; Jeon, J.; Kwak, J.M.; Kim, Y.J. Glycosyltransferase-Like RSE1 negatively regulates leaf senescence through salicylic acid signaling in *Arabidopsis*. *Front. Plant Sci.* **2020**, *11*, 551. [CrossRef]
67. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* **2019**, *20*, 1160–1166. [CrossRef]
68. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]
69. Lim, E.K.; Jackson, R.G.; Bowles, D.J. Identification and characterisation of *Arabidopsis* glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. *FEBS Lett.* **2005**, *579*, 2802–2806. [CrossRef]
70. Šmehilová, M.; Dobrůšková, J.; Novák, O.; Takáč, T.; Galuszka, P. Cytokinin-specific glycosyltransferases possess different roles in cytokinin homeostasis maintenance. *Front Plant Sci.* **2016**, *7*, 1264. [CrossRef]
71. Kovinich, N.; Saleem, A.; Arnason, J.T.; Miki, B. Functional characterization of a UDP-glucose:flavonoid 3-*O*-glucosyltransferase from the seed coat of black soybean (*Glycine max* (L.) Merr.). *Phytochemistry* **2010**, *71*, 1253–1263. [CrossRef] [PubMed]

72. DeBolt, S.; Scheible, W.R.; Schrick, K.; Auer, M.; Beisson, F.; Bischoff, V.; Bouvier-Navé, P.; Carroll, A.; Hematy, K.; Li, Y.; et al. Mutations in UDP-Glucose:sterol glucosyltransferase in *Arabidopsis* cause transparent testa phenotype and suberization defect in seeds. *Plant Physiol.* **2009**, *151*, 78–87. [CrossRef] [PubMed]
73. Fäldt, J.; Arimura, G.; Gershenzon, J.; Takabayashi, J.; Bohlmann, J. Functional identification of AtTPS03 as (E)-β-ocimene synthase: A monoterpene synthase catalyzing jasmonate- and wound-induced volatile formation in *Arabidopsis thaliana*. *Planta* **2003**, *216*, 745–751. [CrossRef] [PubMed]
74. Rips, S.; Frank, M.; Elting, A.; Offenborn, J.N.; von Schaewen, A. Golgi α1,4-fucosyltransferase of *Arabidopsis thaliana* partially localizes at the nuclear envelope. *Traffic* **2017**, *18*, 646–657. [CrossRef] [PubMed]
75. Cushing, D.A.; Forsthoefel, N.R.; Gestaut, D.R.; Vernon, D.M. *Arabidopsis* emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* **2005**, *221*, 424–436. [CrossRef]
76. Møller, S.R.; Yi, X.; Velásquez, S.M.; Gille, S.; Hansen, P.L.M.; Poulsen, C.P.; Olsen, C.E.; Rejzek, M.; Parsons, H.; Yang, Z.; et al. Identification and evolution of a plant cell wall specific glycoprotein glycosyl transferase, ExAD. *Sci. Rep.* **2017**, *7*, 45341. [CrossRef]
77. Jadid, N.; Mialoundama, A.S.; Heintz, D.; Ayoub, D.; Erhardt, M.; Mutterer, J.; Meyer, D.; Alioua, A.; Van Dorsselaer, A.; Rahier, A.; et al. DOLICHOL PHOSPHATE MANNOSE SYNTHASE1 mediates the biogenesis of isoprenyl-linked glycans and influences development, stress response, and ammonium hypersensitivity in *Arabidopsis*. *Plant Cell* **2011**, *23*, 1985–2005. [CrossRef]
78. Cheng, C.Y.; Krishnakumar, V.; Chan, A.P.; Thibaud-Nissen, F.; Schobel, S.; Town, C.D. Araport11: A complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **2017**, *89*, 789–804. [CrossRef]
79. Lee, C.; Teng, Q.; Huang, W.; Zhong, R.; Ye, Z.H. The *Arabidopsis* family GT43 glycosyltransferases form two functionally nonredundant groups essential for the elongation of glucuronoxylan backbone. *Plant Physiol.* **2010**, *153*, 526–541. [CrossRef]
80. Basu, D.; Wang, W.; Ma, S.; DeBrosse, T.; Poirier, E.; Emch, K.; Soukup, E.; Tian, L.; Showalter, A.M. Two hydroxyproline galactosyltransferases, GALT5 and GALT2, function in arabinogalactan-protein glycosylation, growth and development in *Arabidopsis*. *PLoS ONE* **2015**, *10*, e0125624. [CrossRef]
81. Lalonde, M.E.; Durocher, Y. Therapeutic glycoprotein production in mammalian cells. *J. Biotechnol.* **2017**, *251*, 128–140. [CrossRef] [PubMed]
82. Perrin, R.; Wilkerson, C.; Keegstra, K. Golgi enzymes that synthesize plant cell wall polysaccharides: Finding and evaluating candidates in the genomic era. *Plant Mol. Biol.* **2001**, *47*, 115–130. [CrossRef] [PubMed]
83. Yamamoto, T.; Hoshikawa, K.; Ezura, K.; Okazawa, R.; Fujita, S.; Takaoka, M.; Mason, H.S.; Ezura, H.; Miura, K. Improvement of the transient expression system for production of recombinant proteins in plants. *Sci. Rep.* **2018**, *8*, 4755. [CrossRef] [PubMed]
84. Kitazawa, K.; Tryfona, T.; Yoshimi, Y.; Hayashi, Y.; Kawauchi, S.; Antonov, L.; Tanaka, H.; Takahashi, T.; Kaneko, S.; Dupree, P.; et al. β-galactosyl Yariv reagent binds to the β-1,3-galactan of arabinogalactan proteins. *Plant Physiol.* **2013**, *161*, 1117–1126. [CrossRef] [PubMed]
85. Su, S.; Higashiyama, T. Arabinogalactan proteins and their sugar chains: Functions in plant reproduction, research methods, and biosynthesis. *Plant Reprod.* **2018**, *31*, 67–75. [CrossRef]
86. Ma, Y.; Zeng, W.; Bacic, A.; Johnson, K. AGPs through time and space. *Annual Plant Reviews* **2018**, *1*, 1–38. [CrossRef]
87. Knoch, E.; Dilokpimol, A.; Geshi, N. Arabinogalactan proteins: Focus on carbohydrate active enzymes. *Front. Plant Sci.* **2014**, *5*, 198. [CrossRef]
88. Lamport, D.T.A.; Várnai, P. Periplasmic arabinogalactan glycoproteins act as a calcium capacitor that regulates plant growth and development. *New Phytol.* **2013**, *197*, 58–64. [CrossRef]
89. Lamport, D.T.A.; Varnai, P.; Seal, C.E. Back to the future with the AGP-Ca2+ flux capacitor. *Ann. Bot.* **2014**, *114*, 1069–1085. [CrossRef]
90. Zhang, Y.; Held, M.A.; Showalter, A.M. Elucidating the roles of three β-glucuronosyltransferases (GLCATs) acting on arabinogalactan-proteins using a CRISPR-Cas9 multiplexing approach in *Arabidopsis*. *BMC Plant Biol.* **2020**, *20*, 221. [CrossRef]
91. Rautengarten, C.; Ebert, B.; Moreno, I.; Temple, H.; Herter, T.; Link, B.; Donas-Cofre, D.; Moreno, A.; Saez-Aguayo, S.; Blanco, F.; et al. The Golgi localized bifunctional UDP-rhamnose/UDP-galactose transporter family of *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11563–11568. [CrossRef] [PubMed]
92. Andersen, M.C.F.; Boos, I.; Ruprecht, C.; Willats, W.G.T.; Pfrengle, F.; Clausen, M.H. Synthesis and application of branched type II arabinogalactans. *J. Org. Chem.* **2017**, *82*, 12066–12084. [CrossRef] [PubMed]
93. Kuraku, S.; Zmasek, C.M.; Nishimura, O.; Katoh, K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* **2013**, *41*, W22–W28. [CrossRef] [PubMed]
94. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
95. Miller, M.A.; Pfeiffer, W.T.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8. [CrossRef]
96. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [CrossRef] [PubMed]
97. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [CrossRef]

98. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [CrossRef]
99. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef]
100. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1994**, *2*, 28–36.
101. Earley, K.W.; Haag, J.R.; Pontes, O.; Opper, K.; Juehne, T.; Song, K.; Pikaard, C.S. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **2006**, *45*, 616–629. [CrossRef]
102. Geshi, N.; Pauly, M.; Ulvskov, P. Solubilization of galactosyltransferase that synthesizes 1,4-β-galactan side chains in pectic rhamnogalacturonan I. *Physiol. Plant* **2002**, *114*, 540–548. [CrossRef] [PubMed]
103. Alwael, H.; Connolly, D.; Paull, B. Liquid chromatographic profiling of monosaccharide concentrations in complex cell-culture media and fermentation broths. *Anal. Methods* **2011**, *3*, 62–69. [CrossRef] [PubMed]