Annual Plant Reviews (2011) **41**, 167–212
doi: 10.1002/9781444391015.ch6

http://onlinelibrary.wiley.com

Chapter 6

GLYCOSYLTRANSFERASES OF THE GT8 FAMILY

Yanbin Yin[1,3,4], Debra Mohnen[2,3,4], Ivana Gelineo-Albersheim[2,3], Ying Xu[1,3,4] and Michael G. Hahn[2,3,5]

[1] *Computational Systems Biology Lab, Institute of Bioinformatics, The University of Georgia, Athens, GA 30602, USA*
[2] *Complex Carbohydrate Research Center, The University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA*
[3] *BioEnergy Science Center, The University of Georgia, Athens, GA 30602, USA*
[4] *Department of Biochemistry and Molecular Biology, The University of Georgia, Athens, GA 30602, USA*
[5] *Department of Plant Biology, The University of Georgia, Athens, GA 30602, USA*

Manuscript received November 2009

Abstract: The higher plant genomes sequenced to date include numerous genes encoding proteins classified as belonging to CAZy family GT8. The large number and diversity of GT8 proteins in higher plants, which currently constitute more than 65% of the identified eukaryotic GT8 genes, highlight the importance of these proteins in plants. Here we summarize a detailed phylogenetic study of GT8 proteins from three monocot and four dicot plant genomes that clearly divides higher plant GT8 proteins into two distantly related sets of clades, many of which are further divided into statistically well-supported subclades. One set, the GAUT1 (*GA*lact*U*ronosyl*T*ransferase1)-related family, includes the GAUT and *GAUT-L*ike (GATL) proteins, comprising one proven galacturonsyltransferase and multiple additional members strongly implicated in the synthesis of pectins and xylan, two major types of polysaccharides present in plant cell walls. The second set, which includes *P*lant *G*lycogenin-like *S*tarch *I*nitiation *P*roteins (PGSIPs) and *G*alactin*ol S*ynthase*s* (GolSs), appears not to be directly involved in plant cell wall synthesis. The PGSIPs have been suggested to play a role in priming starch biosynthesis, while the GolSs are key enzymes in the synthesis of

Annual Plant Reviews Volume 41: Plant Polysaccharides, Biosynthesis and Bioengineering Edited by P. Ulvskov
 ISBN: 978-1-405-18172-3

the raffinose family of oligosaccharides that play important roles in plant responses to environmental stress. This chapter also summarizes data from biochemical, transcriptional, and mutational studies that provide additional insights into possible functions of higher plant GT8 proteins. However, much work is needed to fully define the roles of GT8 proteins in plants, particularly with respect to their enzymatic function in plant cell wall biosynthesis.

Keywords: biosynthesis; cell wall; dicot; galactinol synthase; galacturonosyltransferase; GATL; GAUT; glycosyltransferase; GolS; GT8; monocot; pectin; PGSIP; retaining; starch; xylan

6.1 Introduction

Glycosyltransferase family 8 (GT8) of the CAZy Carbohydrate Active Enz*y*mes database (Cantarel *et al.* 2009 and see Chapter 4) is presently a family of 973 predicted retaining glycosyltransferases (GTs) with 486 bacterial, 451 eukaryote, 34 viral and 2 unclassified members. Of the eukaryote GT8 members currently listed, 297 are from higher plants with 41 of these from Arabidopsis.

The enzyme activities that have been confirmed biochemically for selected members of this family include multiple bacterial α-glycosyltransferases, many of which synthesize part of the core region of lipopolysaccharide (LPS), as well as three classes of eukaryotic enzymes: glycogenin glucosyltransferases and plant glycogenin-like proteins; galactinol synthases; and a galacturonosyltransferase involved in the synthesis of the plant cell wall pectic polysaccharide, homogalacturonan. As a group, the GT8 enzymes catalyse the transfer of diverse sugars (Glc, Gal, GlcNAC, GalA) onto lipooligosaccharide, protein, inositol, oligosaccharide or polysaccharide acceptors using nucleotide sugar substrates.

After a very brief overview of bacterial GT8 enzymes, this chapter focuses on the eukaryotic enzymes with particular emphasis on the GT8 genes present in plants. The initial overview of bacterial GT8 enzymes, which are involved in lipopolysaccharide biosynthesis, is provided in part because the structure of one of these bacterial GT8 proteins has been determined by X-ray crystallography (see below).

Lipopolysaccharides are amphipathic glycoconjugates located in the outer membrane of Gram-negative bacteria, but whose synthesis begins on the cytoplasmic side of the inner membrane and includes flipping and transport across the periplasm (Raetz & Whitfield 2002). LPS has a protective and permeability barrier function and also provides information for interaction of bacteria with other cells. In this regard, LPS can also serve as a virulence factor. LPS consists of a hydrophobic so-called lipid A domain, a nonrepeating core domain, and a distal polysaccharide domain such as O-specific polysaccharide, enterobacterial common antigen (ECA) or capsular polysaccharide. Many of the bacterial GT8 enzymes, which include α1,2 and

α1,3 glucosyltransferases (GlcTs) and galactosyltransferases (GalTs), synthesize part of the core region of LPS. The LPS core is an oligosaccharide region that contains up to 15 glycosyl residues (Holst 2007). There is structural variation between the core oligosaccharide regions of LPS from different bacteria. Examples of GT8 LPS core GTs include the lipopolysaccharide glucosyltransferase 1 (EC 2.4.1.58) (also known as lipopolysaccharide glucosyltransferase I and UDPglucose:(heptosyl)lipopolysaccharide 1,3-glucosyltransferase) which transfers Glc from UDP-Glc onto the nonreducing end heptose of some LPS inner cores (Müller *et al.* 1972). Another example is the bacterial lipopolysaccharide (LPS) α1,3-galactosyltransferase (EC 2.4.1.44) which catalyses the transfer of Gal from UDP-Gal on to Glc in the partially completed outer LPS core (Holst 2007).

The eukarotic GT8 proteins are divided into three groups based on the glycan synthesized. Proteins from each of these groups are found in plants. These three groups of enzymes are:

1 Glycogenin glucosyltransferase (EC 2.4.1.186) which, in glycogen-producing organisms such as mammals, is an α-glucosyltransferase that first glucosylates itself at a Tyr residue and then elongates the Glc to form an oligosaccharide of 5–13 α1,4-linked glucosyl resides attached to the protein glycogenin (Hurley *et al.* 2005). This oligosaccharide primer is then further glycosylated by glycogen synthase and branching enzyme to yield glycogen, an energy storage polysaccharide.
2 Inositol 1-α-galactosyltransferase (galactinol synthase) (EC 2.4.1.123) is another class of eukaryotic GT8 GTs which galactosylates myoinositol to form UDP and *O*-α-D-galactopyranosyl-[1→1]-L-*myo*-inositol, also known as galactinol. Galactinol serves as the donor for stachyose and raffinose synthesis. As such, galactinol synthase represents the first unique enzyme in the biosynthetic pathway that leads to these oligosaccharides, which are osmoprotectants in specialized plant tissues and serve as sources of energy during the germination of some seeds (Saravitz *et al.* 1987). Sequential transfer of αGal from galactinol on to sucrose yields the trisaccharide (raffinose: αGal-1,6-αGlc-1,2-βFru) or the tetrasaccharide (stachyose: αGal-1,6-αGal-1,6-αGlc-1,2-βFru).
3 The largest of the eukaryotic GT8 groups consists of plant cell wall biosynthetic glycosyltransferases belonging to the GAUT1-related gene family. Only one protein in this group has had its activity biochemically verified and that is the pectin biosynthetic enzyme homogalacturonan α1,4-galacturonosyltransferase 1 (EC 2.4.1.43) (GAUT1) (Sterling *et al.* 2001).

As a group the GT8 enzymes are predicted to be retaining glycosyltransferases and to have a GT-A fold, on the basis of information gained from the only two family members having a resolved crystal structure: UDP-Gal:α1,4-galactosyltransferase (LgtC) from *Neisseria meningitides*, which catalyses the transfer of D-Gal from UDP-Gal on to the Gal in a terminal

lactose portion of LPS in Neisseria (Persson *et al.* 2001), and rabbit muscle glycogenin (Gyg), the self-glucosylating initiation protein that synthesizes the oligosaccharide primer for glycogen synthesis (Gibbons *et al.* 2002). Like all retaining glycosyltransferases, the GT8 GTs synthesize glycoconjugates with an α-anomeric configuration when D-nucleotide sugars (or other activated sugars with an α linkage) are used as donor substrates. Of course, retaining enzymes will synthesize β-linked glycoconjugates when L-nucleotide sugars are the donor substrates.

6.2 Phylogeny of family GT8

A detailed phylogenetic analysis of the complete CAZy Family GT8 has been carried out using genes identified from fully sequenced organisms utilizing the Pfam GT8 domain model (PF01501; http://pfam.sanger.ac.uk/family?entry=PF01501) as the probe (Yin *et al.* 2010). We summarize the results of those analyses here as they pertain to GT8 proteins identified from those higher plants for which complete genome sequences have been published. An E-value cutoff of $1e^{-2}$ was used as a threshold to identify genes encoding GT8 family members. This threshold resulted in the identification of all 41 GT8 proteins previously identified in Arabidopsis (Sterling *et al.* 2006) with no false positives identified. The plants used for the analysis reported here include three monocots (*Oryza sativa* (Yu *et al.* 2002; Goff *et al.* 2002), *Sorghum bicolor* (Paterson *et al.* 2009), *Brachypodium distachyon* (http://www.phytozome.net/) and four dicots (*Arabidopsis thaliana* (Arabidopsis Genome Initiative 2000), *Populus trichocarpa* (Tuskan *et al.* 2006), *Vitis vinifera* (Jaillon *et al.* 2007), *Glycine max* (http://www.phytozome.net/)). During the analysis it became clear that some proteomes, such as those of Arabidopsis and rice, contain protein variants that arise from alternative splicing. For such genes, we kept only one protein sequence for each gene.

The plant GT8 family members identified in these screens are listed in Table 6.1, together with useful information about each gene and its encoded protein, including the plant species, Gene ID, GT8 clade, clade subfamily (where identifiable), the number of ESTs found in GenBank, the presence of signal peptide (as predicted by SignalP v3.0; Bendtsen *et al.* 2004)), the number of transmembrane domains (as predicted by TMHMM v2.0; Krogh *et al.* 2001)), and the protein length. All listed proteins contain the Glyco_transf_8 (PF01501) Pfam domain. The EST matches were found by a BLAST search of the full-length coding region sequences of the plant GT8 proteins against the GenBank EST sequences obtained from PlantGDB (Dong *et al.* 2005). Using an E-value cutoff of $\leq 1e^{-2}$, ESTs more than 98% identical to the query were kept and numerated. These results provide evidence that at least 75% of the GT8 proteins listed are expressed in at least one plant tissue.

Table 6.1 Family GT8 genes identified in seven sequenced higher plant genomes. All encoded proteins contain the Glyco_transf_8 (PF01501) Pfam domain

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
sorghum (sb)	Sb03g033400	XP_002456278	GAUT	GAUT-a	1	N	0	474
rice (os)	LOC_Os01g52710.1	NP_001044121	GAUT	GAUT-a	17	N	0	537
Brachypodium (bd)	Bradi2g48710.1	NA	GAUT	GAUT-a	0	Y	0	539
soybean (gm)	Glyma08g42280.1	NA	GAUT	GAUT-a	15	N	1	526
soybean (gm)	Glyma14g03110.1	NA	GAUT	GAUT-a	0	N	1	525
soybean (gm)	Glyma02g45720.1	NA	GAUT	GAUT-a	0	N	0	446
poplar (pt)	estExt_Genewise1Plus.C_LG_XIV2539	XP_002321071	GAUT	GAUT-a	0	N	1	532
grape (vv)	GSVIVP00018928001	XP_002279893	GAUT	GAUT-a	1	N	1	528
Arabidopsis (at)	AT3G58790.1 (AtGAUT15)	NP_191438	GAUT	GAUT-a	7	Y	1	541
Arabidopsis (at)	AT5G54690.1 (AtGAUT12)	NP_200280	GAUT	GAUT-a	4	N	1	536
poplar (pt)	estExt_fgenesh4_pm.C_LG_XIII0357	XP_002319802	GAUT	GAUT-a	3	N	1	534
poplar (pt)	eugene3.00111083	XP_002317564	GAUT	GAUT-a	0	N	1	533
grape (vv)	GSVIVP00027429001	CAO39049	GAUT	GAUT-a	2	N	1	533
soybean (gm)	Glyma12g34280.1	NA	GAUT	GAUT-a	2	N	1	534
soybean (gm)	Glyma13g36280.1	NA	GAUT	GAUT-a	6	N	1	534
soybean (gm)	Glyma06g41630.1	NA	GAUT	GAUT-a	1	N	1	534
soybean (gm)	Glyma12g16550.1	NA	GAUT	GAUT-a	4	N	1	534
sorghum (sb)	estExt_Genewise1Plus.C_chr_85936	XP_002443426	GAUT	GAUT-a	6	N	1	536
rice (os)	LOC_Os12g38930.1	NP_001067123	GAUT	GAUT-a	8	N	1	555
Brachypodium (bd)	Bradi4g03670.1	NA	GAUT	GAUT-a	0	N	1	536

Table 6.1 *Continued*

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
Brachypodium (bd)	Bradi1g70290.1	NA	GAUT	GAUT-a	2	N	1	564
sorghum (sb)	estExt_fgenesh1_pg.C_chr_14427	XP_002465655	GAUT	GAUT-a	3	N	1	489
rice (os)	LOC_Os03g11330.1	ABF94604	GAUT	GAUT-a	1	N	1	578
grape (vv)	GSVIVP00021044001	CAO15050	GAUT	GAUT-a	3	N	1	533
Arabidopsis (at)	AT5G15470.1 (AtGAUT14)	NP_197051	GAUT	GAUT-a	13	N	1	533
Arabidopsis (at)	AT3G01040.1 (AtGAUT13)	NP_186753	GAUT	GAUT-a	13	N	1	534
poplar (pt)	eugene3.00170460	XP_002324094	GAUT	GAUT-a	0	N	1	529
poplar (pt)	eugene3.00041059	XP_002305374	GAUT	GAUT-a	0	N	1	529
soybean (gm)	Glyma08g26480.1	NA	GAUT	GAUT-a	1	N	1	539
soybean (gm)	Glyma18g49960.1	NA	GAUT	GAUT-a	1	N	1	540
soybean (gm)	Glyma13g05950.1	NA	GAUT	GAUT-a	2	N	1	535
soybean (gm)	Glyma19g03460.1	NA	GAUT	GAUT-a	1	N	1	535
sorghum (sb)	fgenesh1_pg.C_chr_10000882	XP_002438104	GAUT	GAUT-b	0	Y	1	505
rice (os)	LOC_Os06g12280.1	BAD37314	GAUT	GAUT-b	32	Y	1	505
Brachypodium (bd)	Bradi1g45210.1	NA	GAUT	GAUT-b	0	Y	1	501
rice (os)	LOC_Os02g51130.1	NP_001048110	GAUT	GAUT-b	15	Y	1	494
sorghum (sb)	e_gw1.4.16023.1	XP_002454284	GAUT	GAUT-b	0	Y	1	493
Brachypodium (bd)	Bradi3g59370.1	NA	GAUT	GAUT-b	0	Y	1	508
grape (vv)	GSVIVP00038116001	XP_002282423	GAUT	GAUT-b	1	N	1	543
poplar (pt)	eugene3.03090007	XP_002332802	GAUT	GAUT-b	2	N	1	565
poplar (pt)	gw1.123.42.1	XP_002329246	GAUT	GAUT-b	2	N	0	504

Arabidopsis (at)	AT3G02350.1 (AtGAUT9)	NP_566170	GAUT	GAUT-b	39	N	1	562
soybean (gm)	Glyma19g05060.1	NA	GAUT	GAUT-b	4	Y	1	553
soybean (gm)	Glyma13g06990.1	NA	GAUT	GAUT-b	4	N	1	553
poplar (pt)	eugene3.00080075	XP_002310951	GAUT	GAUT-b	0	Y	1	555
soybean (gm)	Glyma18g33210.1	NA	GAUT	GAUT-b	42	N	0	509
soybean (gm)	Glyma08g46210.1	ACU20809	GAUT	GAUT-b	43	N	1	557
poplar (pt)	eugene3.00002521	XP_002301803	GAUT	GAUT-b	5	Y	1	555
Arabidopsis (at)	AT3G25140.1 (AtGAUT8)	NP_189150	GAUT	GAUT-b	64	N	1	560
grape (vv)	GSVIVP00010370001	XP_002273962	GAUT	GAUT-b	15	N	0	137
sorghum (sb)	Sb04g020140	XP_002453869	GAUT	GAUT-b	7	Y	1	535
rice (os)	LOC_Os02g29530.1	NP_001046899	GAUT	GAUT-b	15	Y	1	534
Brachypodium (bd)	Bradi3g43810.1	NA	GAUT	GAUT-b	0	Y	1	524
sorghum (sb)	estExt_fgenesh1_pg.C_chr_62314	XP_002447180	GAUT	GAUT-c	5	Y	1	556
Brachypodium (bd)	Bradi5g23250.1	NA	GAUT	GAUT-c	0	N	1	566
rice (os)	LOC_Os04g54360.1	NP_001054014	GAUT	GAUT-c	9	N	1	557
grape (vv)	GSVIVP00026021001	XP_002279062	GAUT	GAUT-c	0	N	1	533
poplar (pt)	estExt_Genewise1Plus.C_1200026	XP_002329358	GAUT	GAUT-c	0	N	1	535
poplar (pt)	fgenesh4_pm.C_LG_XIII000435	XP_002319923	GAUT	GAUT-c	0	N	1	535
Arabidopsis (at)	AT2G20810.1 (AtGAUT10)	NP_565485	GAUT	GAUT-c	35	N	1	537
soybean (gm)	Glyma13g37650.1	NA	GAUT	GAUT-c	14	N	1	534
soybean (gm)	Glyma12g32820.1	NA	GAUT	GAUT-c	6	N	1	534
Brachypodium (bd)	Bradi1g60010.1	NA	GAUT	GAUT-c	2	N	1	540
sorghum (sb)	estExt_fgenesh1_pm.C_chr_10735	XP_002465134	GAUT	GAUT-c	9	N	1	544

Table 6.1 *Continued*

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
rice (os)	LOC_Os03g30000.1	NP_001050360	GAUT	GAUT-c	6	N	1	542
grape (vv)	GSVIVP00000008001	CAO62433	GAUT	GAUT-c	4	N	1	499
grape (vv)	GSVIVP00000005001	CAO62431	GAUT	GAUT-c	2	N	0	211
Arabidopsis (at)	AT1G18580.1 (AtGAUT11)	NP_564057	GAUT	GAUT-c	19	N	1	538
poplar (pt)	eugene3.01290051	XP_002330135	GAUT	GAUT-c	0	N	1	532
poplar (pt)	grail3.0138001201	XP_002318580	GAUT	GAUT-c	1	N	0	490
soybean (gm)	Glyma05g07410.1	NA	GAUT	GAUT-c	0	N	0	474
soybean (gm)	Glyma17g08910.1	NA	GAUT	GAUT-c	5	N	1	537
soybean (gm)	Glyma06g22730.1	NA	GAUT	GAUT-c	5	N	1	535
soybean (gm)	Glyma04g31770.1	NA	GAUT	GAUT-c	4	N	1	535
soybean (gm)	Glyma05g09200.1	NA	GAUT	GAUT-d	9	N	0	585
poplar (pt)	eugene3.00660198	XP_002328269	GAUT	GAUT-d	0	Y	0	656
Arabidopsis (at)	AT4G38270.1 (AtGAUT3)	NP_195540	GAUT	GAUT-d	10	Y	0	681
grape (vv)	GSVIVP00036268001	CAO47521	GAUT	GAUT-d	3	N	0	618
sorghum (sb)	estExt_Genewise1.C_chr_108333	XP_002439065	GAUT	GAUT-d	1	N	0	513
Brachypodium (bd)	Bradi1g29780.1	NA	GAUT	GAUT-d	0	N	0	1060
rice (os)	LOC_Os10g21890.1	ABB47337	GAUT	GAUT-d	17	N	1	687
rice (os)	LOC_Os06g51160.1	BAD61814	GAUT	GAUT-d	2	Y	0	602
Arabidopsis (at)	AT2G46480.1 (AtGAUT2)	NP_182171	GAUT	GAUT-d	1	N	0	529
Arabidopsis (at)	AT3G61130.1 (AtGAUT1)	NP_191672	GAUT	GAUT-d	41	N	1	674
grape (vv)	GSVIVP00026306001	XP_002272447	GAUT	GAUT-d	14	N	0	228

poplar (pt)	estExt_Genewise1Plus.C_LG_XIV0504	XP_002320745	GAUT	GAUT-d	0	N	1	688
poplar (pt)	eugene3.00021408	XP_002302550	GAUT	GAUT-d	0	N	1	645
soybean (gm)	Glyma09g40260.1	NA	GAUT	GAUT-d	13	Y	1	665
soybean (gm)	Glyma18g45750.1	NA	GAUT	GAUT-d	12	N	0	607
soybean (gm)	Glyma03g02250.1	NA	GAUT	GAUT-d	9	Y	2	845
soybean (gm)	Glyma07g08910.1	NA	GAUT	GAUT-d	14	N	0	613
rice (os)	LOC_Os06g49810.1	EEC81302	GAUT	GAUT-d	23	Y	0	589
Brachypodium (bd)	Bradi1g12180.1	NA	GAUT	GAUT-d	1	Y	0	590
sorghum (sb)	estExt_Genewise1Plus.C_chr_12326	XP_002466644	GAUT	GAUT-d	3	Y	1	589
sorghum (sb)	Sb02g030820	XP_002462717	GAUT	GAUT-d	2	N	1	684
Brachypodium (bd)	Bradi4g36050.1	NA	GAUT	GAUT-d	0	N	1	692
rice (os)	LOC_Os09g36190.1	NP_001063757	GAUT	GAUT-d	23	N	1	698
rice (os)	LOC_Os09g36180.1	NA	GAUT	GAUT-d	3	N	1	1081
sorghum (sb)	estExt_Genewise1Plus.C_chr_27961	XP_002460421	GAUT	GAUT-e	0	Y	1	705
Brachypodium (bd)	Bradi4g33280.1	NA	GAUT	GAUT-e	0	Y	1	697
rice (os)	LOC_Os09g30280.1	NP_001063487	GAUT	GAUT-e	8	Y	1	708
sorghum (sb)	fgenesh1_pg.C_chr_4003251	XP_002454735	GAUT	GAUT-e	0	N	0	371
rice (os)	LOC_Os08g38740.1	BAD10126	GAUT	GAUT-e	2	Y	1	727
Brachypodium (bd)	Bradi3g39270.1	NA	GAUT	GAUT-e	0	Y	1	704
rice (os)	LOC_Os08g23780.1	NP_001061555	GAUT	GAUT-e	51	Y	1	644
sorghum (sb)	Sb07g014890	XP_002444200	GAUT	GAUT-e	3	Y	1	649
Brachypodium (bd)	Bradi3g20550.1	NA	GAUT	GAUT-e	1	Y	1	661
soybean (gm)	Glyma17g00790.1	NA	GAUT	GAUT-e	4	N	0	399
soybean (gm)	Glyma07g40020.1	NA	GAUT	GAUT-e	20	N	0	399

Table 6.1 *Continued*

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
soybean (gm)	Glyma15g12900.1	NA	GAUT	GAUT-e	7	Y	1	658
soybean (gm)	Glyma09g01980.1	NA	GAUT	GAUT-e	14	Y	1	658
grape (vv)	GSVIVP00011843001	CAO23686	GAUT	GAUT-e	22	Y	1	613
Arabidopsis (at)	AT5G47780.1 (AtGAUT4)	NP_568688	GAUT	GAUT-e	46	Y	1	617
poplar (pt)	eugene3.01370031	XP_002329942	GAUT	GAUT-e	4	Y	1	666
poplar (pt)	fgenesh4_pg.C_LG_VI000014	XP_002308736	GAUT	GAUT-e	0	Y	1	649
sorghum (sb)	e_gw1.5.15262.1	XP_002451032	GAUT	GAUT-f	0	N	0	434
Brachypodium (bd)	Bradi4g14910.1	NA	GAUT	GAUT-f	0	Y	1	537
rice (os)	LOC_Os11g37980.1	ABA94533	GAUT	GAUT-f	1	N	1	549
Arabidopsis (at)	AT1G06780.1 (AtGAUT6)	NP_563771	GAUT	GAUT-f	15	N	1	590
Arabidopsis (at)	AT2G30575.1 (AtGAUT5)	NP_850150	GAUT	GAUT-f	5	Y	1	611
grape (vv)	GSVIVP00023858001	XP_002282102	GAUT	GAUT-f	6	Y	1	588
poplar (pt)	eugene3.00051260	XP_002307628	GAUT	GAUT-f	3	N	1	606
poplar (pt)	fgenesh4_pg.C_LG_II000411	NA	GAUT	GAUT-f	0	N	0	365
soybean (gm)	Glyma19g34420.1	NA	GAUT	GAUT-f	21	Y	1	626
soybean (gm)	Glyma03g31590.1	NA	GAUT	GAUT-f	17	Y	1	626
soybean (gm)	Glyma02g15990.1	NA	GAUT	GAUT-f	6	N	1	576
soybean (gm)	Glyma10g03770.1	NA	GAUT	GAUT-f	7	Y	1	586
grape (vv)	GSVIVP00033610001	CAO45341	GAUT	GAUT-g	3	N	0	452
Arabidopsis (at)	AT2G38650.1 (AtGAUT7)	NP_565893	GAUT	GAUT-g	27	N	1	620
poplar (pt)	e_gw1.XVI.562.1	XP_002323701	GAUT	GAUT-g	0	N	1	621

poplar (pt)	estExt_Genewise1Plus.C_281089	XP_002326255	GAUT	GAUT-g	0	N	1	591
soybean (gm)	Glyma09g40610.1	NA	GAUT	GAUT-g	4	N	0	563
soybean (gm)	Glyma18g45230.1	NA	GAUT	GAUT-g	3	N	1	658
soybean (gm)	Glyma16g09420.1	NA	GAUT	GAUT-g	0	N	0	245
sorghum (sb)	Sb01g036430	XP_002465323	GAUT	GAUT-g	2	Y	1	629
rice (os)	LOC_Os03g21250.1	NP_001050007	GAUT	GAUT-g	2	Y	1	632
Brachypodium (bd)	Bradi1g63520.1	NA	GAUT	GAUT-g	0	Y	1	626
sorghum (sb)	estExt_Genewise1.C_chr_211179	XP_002461225	GAUT	GAUT-g	4	Y	1	628
rice (os)	LOC_Os07g48370.1	NP_001060656	GAUT	GAUT-g	7	Y	1	626
Brachypodium (bd)	Bradi1g17570.1	NA	GAUT	GAUT-g	0	Y	1	622
rice (os)	LOC_Os05g40720.1	AAT01328	GAUT	GAUT-g	2	Y	2	668
sorghum (sb)	Sb09g023760	XP_002441280	GAUT	GAUT-g	0	Y	1	639
rice (os)	LOC_Os12g02910.1	EEC68769	GAUT	GAUT-g	2	N	1	660
rice (os)	LOC_Os11g03160.1	NP_001065619	GAUT	GAUT-g	3	N	1	643
sorghum (sb)	fgenesh1_pm.C_chr_8000106	NA	GAUT	GAUT-g	1	N	0	581
Brachypodium (bd)	Bradi3g61120.1	NA	GAUT	GAUT-g	0	N	0	565
Brachypodium (bd)	Bradi4g44070.1	NA	GAUT	GAUT-g	0	Y	1	633
rice (os)	LOC_Os06g13760.1	BAD45664	GATL	GATL-a	2	Y	1	372
Brachypodium (bd)	Bradi1g44470.1	NA	GATL	GATL-a	0	Y	0	358
sorghum (sb)	Sb10g008830	XP_002436791	GATL	GATL-a	0	Y	1	368
sorghum (sb)	estExt_fgenesh1_pg.C_chr_42380	XP_002452580	GATL	GATL-a	15	Y	1	427
Brachypodium (bd)	Bradi3g59760.1	NA	GATL	GATL-a	2	Y	1	380
rice (os)	LOC_Os02g50600.1	NP_001048067	GATL	GATL-a	8	Y	1	374

Table 6.1 *Continued*

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
Arabidopsis (at)	AT3G28340.1 (AtGATL10)	NP_189474	GATL	GATL-a	10	Y	1	366
soybean (gm)	Glyma19g01910.1	ACU19924	GATL	GATL-a	4	Y	1	382
soybean (gm)	Glyma13g04780.1	NA	GATL	GATL-a	7	Y	1	382
grape (vv)	GSVIVP00028923001	CAO40973	GATL	GATL-a	9	Y	0	381
soybean (gm)	Glyma02g03090.1	NA	GATL	GATL-a	3	Y	1	379
soybean (gm)	Glyma01g04460.1	NA	GATL	GATL-a	5	Y	1	379
poplar (pt)	fgenesh4_pm.C_LG_VIII000888	XP_002311853	GATL	GATL-a	0	Y	0	384
poplar (pt)	estExt_Genewise1Plus.C_LG_VIII2617	XP_002311789	GATL	GATL-a	0	Y	1	379
Arabidopsis (at)	AT1G70090.1 (AtGATL9)	NP_564983	GATL	GATL-a	36	Y	1	391
Arabidopsis (at)	AT1G24170.1 (AtGATL8)	NP_173827	GATL	GATL-a	21	Y	1	394
rice (os)	LOC_Os03g18890.1	NP_001049861	GATL	GATL-b	24	Y	0	369
Brachypodium (bd)	Bradi1g64830.1	NA	GATL	GATL-b	0	Y	0	368
sorghum (sb)	estExt_Genewise1.C_chr_19676	XP_002468004	GATL	GATL-b	3	Y	0	372
grape (vv)	GSVIVP00028063001	CAO40071	GATL	GATL-b	6	Y	1	340
soybean (gm)	Glyma10g01960.1	NA	GATL	GATL-b	45	Y	0	360
soybean (gm)	Glyma02g01880.1	NA	GATL	GATL-b	30	Y	0	358
soybean (gm)	Glyma19g40180.1	NA	GATL	GATL-b	46	Y	1	347
soybean (gm)	Glyma03g37560.1	NA	GATL	GATL-b	37	Y	1	347
poplar (pt)	eugene3.00021805	XP_002302739	GATL	GATL-b	4	Y	1	368
poplar (pt)	grail3.0061016301	XP_002320324	GATL	GATL-b	2	Y	1	369

Arabidopsis (at)	AT3G62660.1 (AtGATL7)	NP_191825	GATL	GATL-b	58	Y	0	362
Arabidopsis (at)	AT4G02130.1 (AtGATL6)	NP_192122	GATL	GATL-b	23	Y	0	347
Arabidopsis (at)	AT1G02720.1 (AtGATL5)	NP_171772	GATL	GATL-b	12	Y	0	362
Brachypodium (bd)	Bradi1g12560.1	NA	GATL	GATL-c	0	Y	1	358
sorghum (sb)	estExt_Genewise1.C_chr_12425	XP_002466674	GATL	GATL-c	1	N	0	292
rice (os)	LOC_Os03g47530.1	ABF98189	GATL	GATL-c	24	Y	1	361
rice (os)	LOC_Os10g31650.1	AAP54070	GATL	GATL-c	0	N	0	265
grape (vv)	GSVIVP00014689001	CAO61226	GATL	GATL-c	3	N	0	186
poplar (pt)	estExt_fgenesh4_pm.C_LG_VII0375	XP_002310780	GATL	GATL-c	0	Y	0	353
poplar (pt)	fgenesh4_pg.C_scaffold_57000218	XP_002327710	GATL	GATL-c	0	Y	1	353
Arabidopsis (at)	AT3G50760.1 (AtGATL2)	NP_190645	GATL	GATL-c	4	Y	1	342
soybean (gm)	Glyma01g38520.1	NA	GATL	GATL-c	5	Y	1	352
soybean (gm)	Glyma02g06640.1	NA	GATL	GATL-c	1	N	0	333
grape (vv)	GSVIVP00025008001	CAO22912	GATL	GATL-c	2	Y	1	340
Arabidopsis (at)	AT1G19300.1 (AtGATL1)	NP_564077	GATL	GATL-c	28	Y	0	352
poplar (pt)	eugene3.00400186	ABK94510	GATL	GATL-c	4	Y	1	361
poplar (pt)	estExt_Genewise1Plus.C_LG_II1880	XP_002302469	GATL	GATL-c	1	Y	0	352
soybean (gm)	Glyma06g03770.1	NA	GATL	GATL-c	4	Y	1	367
soybean (gm)	Glyma04g03690.1	NA	GATL	GATL-c	4	N	0	319
soybean (gm)	Glyma17g36650.1	NA	GATL	GATL-c	2	Y	1	353
soybean (gm)	Glyma14g08430.1	NA	GATL	GATL-c	2	Y	1	362
rice (os)	LOC_Os04g44850.1	NP_001053393	GATL	GATL-d	10	Y	0	342
sorghum (sb)	e_gw1.6.14698.1	XP_002446838	GATL	GATL-d	0	Y	0	343

Table 6.1 *Continued*

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
Brachypodium (bd)	Bradi5g16690.1	NA	GATL	GATL-d	0	Y	0	343
Arabidopsis (at)	AT1G13250.1 (AtGATL3)	NP_563925	GATL	GATL-d	14	Y	1	346
grape (vv)	GSVIVP00030290001	XP_002281658	GATL	GATL-d	6	Y	0	345
poplar (pt)	grail3.0006037101	XP_002314884	GATL	GATL-d	0	Y	0	343
poplar (pt)	grail3.0010045901	XP_002312381	GATL	GATL-d	0	Y	0	347
soybean (gm)	Glyma02g11100.1	NA	GATL	GATL-d	12	Y	1	343
soybean (gm)	Glyma01g22480.1	NA	GATL	GATL-d	6	Y	0	339
rice (os)	LOC_Os07g45260.1	NP_001060465	GATL	GATL-e	10	Y	1	378
sorghum (sb)	fgenesh1_pm.C_chr_2000674	XP_002461136	GATL	GATL-e	0	Y	0	335
sorghum (sb)	e_gw1.1.15754.1	XP_002465236	GATL	GATL-e	0	N	0	310
rice (os)	LOC_Os03g24510.1	NP_001050151	GATL	GATL-e	7	Y	1	348
Brachypodium (bd)	Bradi1g61830.1	NA	GATL	GATL-e	0	Y	1	352
grape (vv)	GSVIVP00019872001	XP_002277334	GATL	GATL-e	2	Y	0	352
grape (vv)	GSVIVP00023650001	XP_002271296	GATL	GATL-e	1	Y	0	356
poplar (pt)	grail3.0090012501	XP_002311936	GATL	GATL-e	0	Y	0	348
poplar (pt)	grail3.0106011001	XP_002315420	GATL	GATL-e	0	Y	0	349
Arabidopsis (at)	AT3G06260.1 (AtGATL4)	NP_187277	GATL	GATL-e	2	Y	1	352
soybean (gm)	Glyma07g38430.1	NA	GATL	GATL-e	1	Y	1	351
soybean (gm)	Glyma17g02330.1	NA	GATL	GATL-e	1	Y	1	347
grape (vv)	GSVIVP00022838001	XP_002280832	PGSIP-A		0	Y	1	546
Arabidopsis (at)	AT1G54940.1 (AtPGSIP4)	NP_175891	PGSIP-A		0	N	1	558

Arabidopsis (at)	AT1G08990.1 (AtPGSIP5)	NP_172373	PGSIP-A	1	N	1	567
poplar (pt)	fgenesh4_pg.C_scaffold_70000240	XP_002328440	PGSIP-A	0	N	0	431
poplar (pt)	e_gw1.XIII.1015.1	XP_002319544	PGSIP-A	0	N	0	426
soybean (gm)	Glyma10g29570.1	NA	PGSIP-A	2	N	0	541
soybean (gm)	Glyma19g42380.1	NA	PGSIP-A	0	N	0	517
soybean (gm)	Glyma03g39820.1	NA	PGSIP-A	0	Y	0	434
Brachypodium (bd)	Bradi1g72350.1	NA	PGSIP-A	1	N	1	608
rice (os)	LOC_Os03g08600.1	AAK92624	PGSIP-A	22	Y	1	615
sorghum (sb)	Sb01g044930	XP_002468380	PGSIP-A	3	Y	1	606
poplar (pt)	estExt_Genewise1_v1.C_400718	XP_002326938	PGSIP-A	0	N	0	361
grape (vv)	GSVIVP00014811001	XP_002282762	PGSIP-A	1	N	1	590
Arabidopsis (at)	AT4G33330.1 (AtPGSIP3)	NP_195059	PGSIP-A	4	Y	1	597
soybean (gm)	Glyma04g04080.1	NA	PGSIP-A	4	Y	1	588
soybean (gm)	Glyma14g09070.1	NA	PGSIP-A	0	N	1	598
soybean (gm)	Glyma17g36100.1	NA	PGSIP-A	1	N	1	593
sorghum (sb)	e_gw1.3.18262.1	XP_002458785	PGSIP-A	0	N	1	630
Brachypodium (bd)	Bradi5g27680.1	NA	PGSIP-A	0	N	1	581
Brachypodium (bd)	Bradi3g45800.1	NA	PGSIP-A	0	Y	0	612
sorghum (sb)	fgenesh1_pg.C_chr_4001767	XP_002453984	PGSIP-A	0	N	0	645
rice (os)	LOC_Os02g35020.1	BAD15458	PGSIP-A	6	N	0	655
sorghum (sb)	Sb09g020930	XP_002441128	PGSIP-A	18	N	1	632
Brachypodium (bd)	Bradi2g24740.1	NA	PGSIP-A	0	N	1	634
rice (os)	LOC_Os01g65780.1	NP_001044991	PGSIP-A	88	N	1	636

Table 6.1 *Continued*

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
Brachypodium (bd)	Bradi2g56810.1	NA	PGSIP-A		3	N	1	633
sorghum (sb)	estExt_Genewise1.C_chr_310243	XP_002456736	PGSIP-A		12	N	1	634
grape (vv)	GSVIVP00014288001	CAO60879	PGSIP-A		6	N	0	652
poplar (pt)	e_gw1.V.184.1	XP_002307480	PGSIP-A		0	N	1	630
Arabidopsis (at)	AT1G77130.1 (AtPGSIP2)	NP_177838	PGSIP-A		32	N	1	619
soybean (gm)	Glyma02g40480.1	NA	PGSIP-A		9	N	1	645
soybean (gm)	Glyma14g28370.1	NA	PGSIP-A		1	N	1	542
soybean (gm)	Glyma0214s00200.1	NA	PGSIP-A		1	N	1	591
grape (vv)	GSVIVP00034844001	XP_002275240	PGSIP-A		1	N	1	636
Arabidopsis (at)	AT3G18660.1 (AtPGSIP1)	NP_566615	PGSIP-A		10	N	0	656
poplar (pt)	eugene3.00050013	XP_002307022	PGSIP-A		2	N	0	545
poplar (pt)	eugene3.00070393	XP_002310513	PGSIP-A		0	N	0	631
soybean (gm)	Glyma06g15690.1	NA	PGSIP-A		1	N	0	537
soybean (gm)	Glyma04g39240.1	NA	PGSIP-A		0	N	1	627
soybean (gm)	Glyma05g32370.1	NA	PGSIP-A		5	N	1	641
soybean (gm)	Glyma05g04770.1	NA	PGSIP-A		0	N	1	628
soybean (gm)	Glyma08g15640.1	NA	PGSIP-A		0	N	0	483
rice (os)	LOC_Os02g41520.1	NP_001047475	PGSIP-B		7	Y	5	548
Brachypodium (bd)	Bradi3g49200.1	NA	PGSIP-B		4	Y	7	1189
rice (os)	LOC_Os04g43700.1	CAE05448	PGSIP-B		2	Y	4	480
sorghum (sb)	estExt_fgenesh1_pg.C_chr_61566	XP_002446789	PGSIP-B		14	Y	6	537

Brachypodium (bd)	Bradi5g15780.1	NA	PGSIP-B	0	Y	6	545
Arabidopsis (at)	AT5G18480.1 (AtPGSIP6)	NP_197349	PGSIP-B	43	Y	5	538
grape (vv)	GSVIVP00002209001	XP_002266145	PGSIP-B	14	Y	5	546
poplar (pt)	eugene3.00130489	XP_002319122	PGSIP-B	3	Y	5	547
soybean (gm)	Glyma03g40980.1	NA	PGSIP-B	8	N	5	485
soybean (gm)	Glyma19g43630.1	NA	PGSIP-B	10	Y	6	553
soybean (gm)	Glyma20g37000.1	NA	PGSIP-B	14	Y	5	542
soybean (gm)	Glyma10g30700.1	NA	PGSIP-B	1	Y	5	537
grape (vv)	GSVIVP00009599001	XP_002269578	PGSIP-C	24	Y	5	473
soybean (gm)	Glyma11g03550.1	NA	PGSIP-C	1	N	5	432
soybean (gm)	Glyma05g04630.1	NA	PGSIP-C	4	Y	6	478
Arabidopsis (at)	AT2G35710.1 (AtPGSIP7)	NP_565817	PGSIP-C	45	Y	6	498
Arabidopsis (at)	AT4G16600.1 (AtPGSIP8)	NP_193393	PGSIP-C	0	Y	5	495
rice (os)	LOC_Os04g46750.1	NP_001053509	PGSIP-C	12	N	5	429
sorghum (sb)	Sb06g024740	XP_002448294	PGSIP-C	12	Y	6	476
Brachypodium (bd)	Bradi5g17880.1	NA	PGSIP-C	0	Y	6	467
rice (os)	LOC_Os10g40640.1	NP_001065345	PGSIP-C	23	Y	6	493
sorghum (sb)	Sb01g029400	XP_002467515	PGSIP-C	3	Y	6	487
Brachypodium (bd)	Bradi3g33080.1	NA	PGSIP-C	0	Y	5	488
grape (vv)	GSVIVP00033193001	XP_002280616	GolS	25	N	0	318
grape (vv)	GSVIVP00019669001	XP_002262651	GolS	27	N	0	317
Arabidopsis (at)	AT1G60470.1 (AtGolS4)	NP_176250	GolS	10	N	0	335
Arabidopsis (at)	AT1G60450.1 (AtGolS7)	NP_176248	GolS	4	N	0	333
grape (vv)	GSVIVP00002727001	XP_002265947	GolS	153	N	0	336
poplar (pt)	grail3.0009037801	XP_002311774	GolS	82	N	0	339

Table 6.1 *Continued*

Species	GeneID	GenBankID	GT8 clade	GT8 subclade	No. of ESTs	Signal peptide	No. of transmembrane regions	Protein length (a.a.)
poplar (pt)	estExt_fgenesh4_pg.C_LG_X0618	XP_002314613	GolS		5	N	0	339
rice (os)	LOC_Os07g48830.1	NP_001060697	GolS		38	N	0	329
sorghum (sb)	e_gw1.2.20933.1	XP_002461242	GolS		5	N	0	322
Brachypodium (bd)	Bradi1g17200.1	NA	GolS		1	N	0	345
sorghum (sb)	Sb01g037090	XP_002467954	GolS		14	Y	0	350
rice (os)	LOC_Os03g20120.1	NP_001049939	GolS		55	N	0	342
Brachypodium (bd)	Bradi1g64120.1	NA	GolS		0	N	0	338
poplar (pt)	fgenesh4_pm.C_LG_X000590	XP_002314975	GolS		4	N	0	326
poplar (pt)	fgenesh4_pm.C_LG_VIII000417	XP_002312306	GolS		2	N	0	326
Arabidopsis (at)	AT5G30500.1(AtGolS8)	NP_850902	GolS		0	N	0	329
Arabidopsis (at)	AT4G26250.1(AtGolS6)	NP_567741	GolS		0	N	0	337
Arabidopsis (at)	AT5G23790.1(AtGolS5)	NP_197768	GolS		0	N	0	334
poplar (pt)	gw1.131.196.1	XP_002330017	GolS		1	N	0	306
poplar (pt)	estExt_fgenesh4_pm.C_LG_XIII0025	XP_002319473	GolS		6	N	0	338
poplar (pt)	e_gw1.XIII.549.1	XP_002319472	GolS		1	N	0	335
grape (vv)	GSVIVP00022562001	CAO17390	GolS		0	N	0	332

grape (vv)	GSVIVP00022565001	XP_002281304	GolS	8	Y	0	325
grape (vv)	GSVIVP00022570001	XP_002281369	GolS	13	Y	0	324
grape (vv)	GSVIVP00022561001	XP_002281261	GolS	6	Y	0	325
grape (vv)	GSVIVP00022559001	CAO17387	GolS	2	Y	0	162
Arabidopsis (at)	AT2G47180.1 (AtGolS1)	NP_182240	GolS	28	N	0	345
poplar (pt)	eugene3.00140617	XP_002320958	GolS	2	N	0	337
poplar (pt)	estExt_fgenesh4_pm.C_LG_II0906	XP_002301531	GolS	1	N	0	338
soybean (gm)	Glyma19g40680.1	NA	GolS	5	N	0	336
soybean (gm)	Glyma03g38080.1	NA	GolS	7	N	0	340
grape (vv)	GSVIVP00028165001	XP_002279114	GolS	19	N	0	342
grape (vv)	GSVIVP00028167001	XP_002279157	GolS	9	N	0	340
Arabidopsis (at)	AT1G56600.1 (AtGolS2)	NP_176053	GolS	9	N	0	336
Arabidopsis (at)	AT1G09350.1 (AtGolS3)	NP_172406	GolS	18	N	0	335
soybean (gm)	Glyma20g22700.1	NA	GolS	11	N	0	325
soybean (gm)	Glyma10g28610.1	AAM96867	GolS	21	N	0	329
soybean (gm)	Glyma03g38910.1	NA	GolS	5	N	0	332
soybean (gm)	Glyma19g41550.1	NA	GolS	11	N	0	331

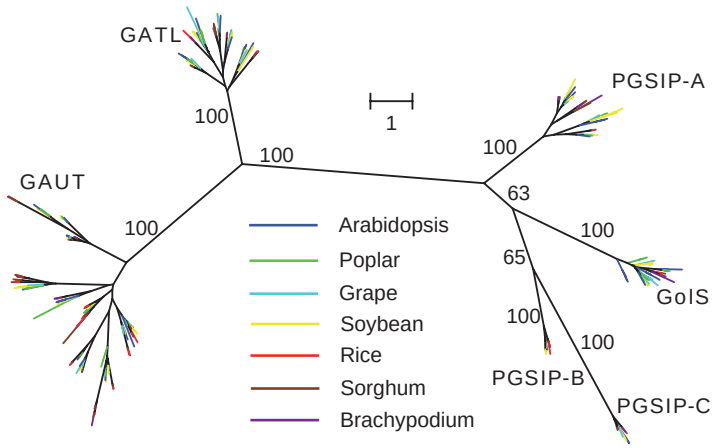

Figure 6.1 The phylogeny of higher plant GT8 proteins. The multiple sequence alignment of 319 full-length plant GT8 protein sequences (Table 6.1) was used to perform a maximum likelihood (ML) phylogeny reconstruction, to which bootstrap analyses were applied 100 times. The statistical bootstrap values are shown beside the branches to show the confidence levels with regard to the clustering of relevant proteins into each group. The colour scheme used for the branches is as follows: Arabidopsis, dark blue; poplar, green; grape, aqua; soybean, yellow; rice, red; sorghum, brown; Brachypodium, purple.

The full-length GT8 protein sequences identified from the screen of fully sequenced higher plant genomes were aligned using MAFFT v6.603 (Katoh *et al.* 2005) and the resulting alignment was used to perform maximum likelihood phylogeny reconstruction using PhyML v2.4.4 (Guindon & Gascuel 2003). The phylogeny of plant GT8 proteins shown in Fig. 6.1 suggests that there are six monophyletic plant GT8 clades that are distributed into two widely divergent groups having only distant evolutionary relatedness. One group contains the GAUT and GATL proteins, some of which have been linked functionally to cell wall polysaccharide biosynthesis. The other group contains the PGSIP and GolS proteins, some of which have been linked to starch and raffinose or stachyose oligosaccharide synthesis, respectively. Detailed phylogenetic studies of GT8 domains from all sequenced organisms suggests that the two broad groups of plant GT8 proteins each arose from distinct ancestral bacterial genes (Yin *et al.* 2010). A more detailed discussion of each of these groups of proteins follows later in this chapter.

Each of the monophyletic clades identified in the phylogenetic reconstruction of the GT8 protein sequences was subjected to further individual analysis to discern subclade structures within each clade (Figs 6.2–6.7). The full-length protein sequences for each clade were realigned using MAFFT, and maximum likelihood phylogenies were built using PhyML as previously. The results of these studies demonstrate that most of the clades can

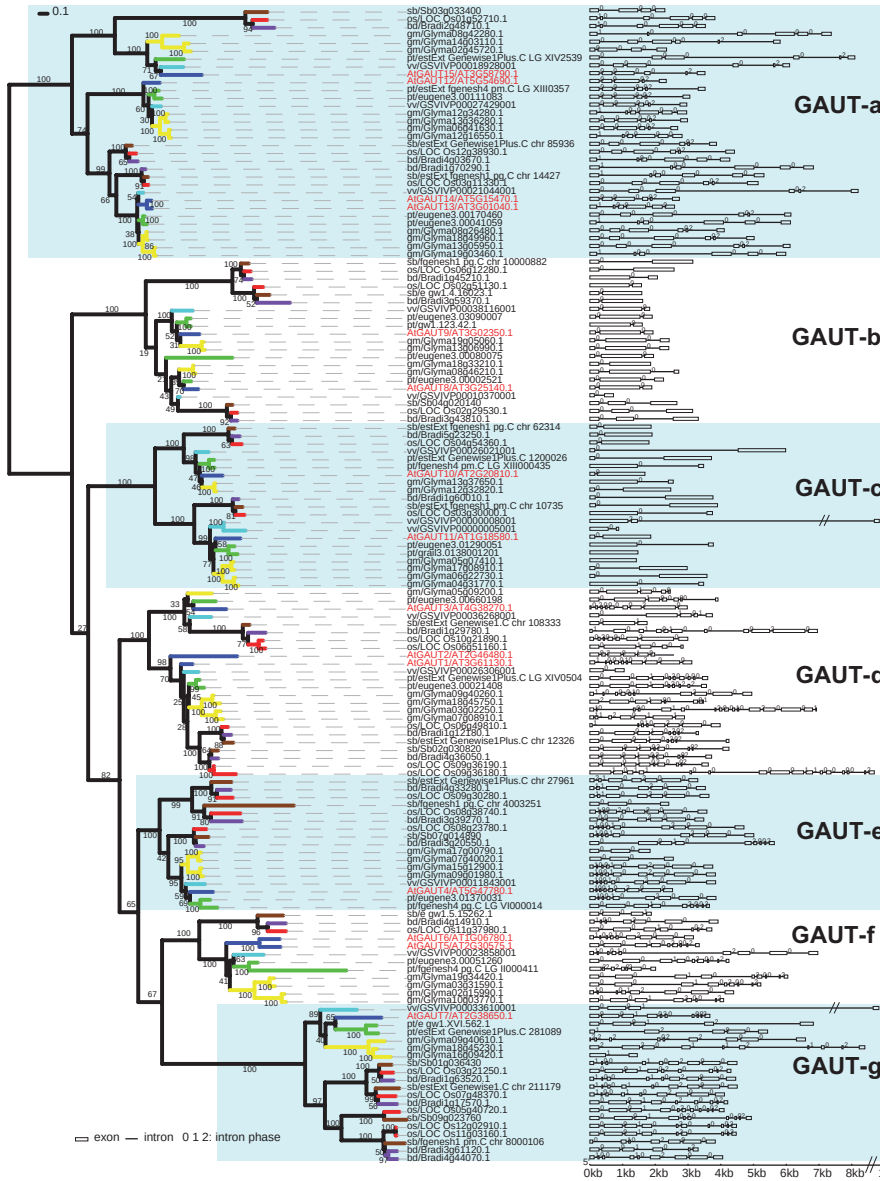

Figure 6.2 The phylogeny of the GAUT clade and structures of the *GAUT* genes. This ML phylogeny was built based on the multiple sequence alignment of full length GAUT protein sequences identified from fully sequenced higher plant genomes. The bootstrap values are shown to indicate the confidence level of each grouping. Subclades are named in the 'GAUT-alphabet' format according to the phylogenetic grouping and gene structure conservation. The colour scheme for the branches is the same as that used in Fig. 6.1. The gene structure was plotted using the GSDS server (Guo *et al.* 2007) with intron–exon coordinate information acquired from the downloaded annotation files of the seven genomes. The intron–exon structure for each gene is shown on the right. The intron phase indicates the position of the intron within a codon. If it is not located within a codon the phase is 0. If it is located within a codon and after the first base of the codon the phase is 1. If it is located within a codon and after the second base of the codon, the phase is 2.

be further classified into distinct well-supported subfamilies. The only exceptions were the PGSIP-A and GolS clades, where the sequences available from the fully sequenced plant genomes do not provide sufficient resolution to clearly discern the subclade structure of the clade.

All of the subclades in the cell wall biosynthesis-related clades (GAUT and GATL) correspond to different monophyletic clusters with significant bootstrap support values greater than or equal to 70%. Each of the GAUT and GATL subclades contain sequences from each of the seven plant genomes, indicating that the proteins in each subclade are orthologous to each other. This result is not surprising, considering that all plants need to synthesize walls and hence their polysaccharide constituents. The subclade structures of the clades not related to cell wall synthesis (PGSIP and GolS) are more complex. Not all subclades in the PGSIP and GolS groups contained representative proteins from all seven plant genomes examined, and some subclades lacked either monocot or dicot representatives. These results suggest that monocots and dicots diverged in their needs for, and uses of, enzymes involved in starch and raffinose biosynthesis, the annotated functions of the PGSIP and GolS families, respectively.

6.3 GT8 clades related to plant cell wall polysaccharide synthesis

Representatives of each of two clades (GAUT and GATL) in family GT8 were initially implicated in plant cell wall synthesis based on altered wall phenotypes exhibited by plants carrying mutations in these genes (Bouton *et al.* 2002; Lao *et al.* 2003; Shao *et al.* 2004). Subsequently, *AtGAUT1* was demonstrated to functionally encode a homogalacturonan galacturonosyltransferase (Sterling *et al.* 2006) involved in pectic polysaccharide synthesis. Additional mutational and biochemical studies on members of the GAUT and GATL clades further supported the hypothesis that members of these two clades are involved in plant cell wall polysaccharide biosynthesis (Caffall *et al.* 2009; Y. Kong, G. Zhou, Y. Yin, Y. Zhu, S. Pattathil, M.G. Hahn, submitted for publication). Bioinformatic analyses of the entire GT8 family also identified amino acid sequence motifs that are unique to the GAUT and GATL clades among the family GT8 clades (Sterling *et al.* 2006; Yin *et al.* 2010), further substantiating the functional separation of the GAUT and GATL from other clades within the GT8 family.

6.3.1 Galacturonosyltransferase (GAUT) clade

The GAUT clade consists of genes encoding proven and putative glycosyltransferases involved in plant cell wall polysaccharide biosynthesis. The GAUT proteins were first identified as being involved in cell wall polysaccharide synthesis in Arabidopsis through studies of a plant carrying a

mutation in the *QUASIMODO1/GAUT8* gene (Bouton *et al.* 2002). Biochemical proof of a role of this family of proteins in cell wall synthesis was independently obtained by enzyme purification, cloning and functional expression studies that led to the identification of Arabidopsis GAUT1 as a pectin biosynthetic homogalacturonan galacturonosyltransferase (Sterling *et al.* 2006). The proteins encoded by almost all *GAUT* genes are type II membrane proteins carrying a single transmembrane domain.

6.3.1.1 Phylogeny of GAUT clade

Initial phylogenetic studies of the GAUT family carried out in *Arabidopsis thaliana* identified three clades of GAUT proteins: clade A (AtGAUTs 1–7), clade B (AtGAUTs 8–11) and clade C (AtGAUTs 12–15) (Sterling *et al.* 2006). Subsequently, a search of the sequenced genomes of rice (*Oryza sativa*) and poplar (*Populus trichocarpa)* for GAUT homologs identified 21 poplar and 22 rice GAUTs and led to a refinement of the GAUT phylogeny into 7 subclades (Caffall *et al.* 2009). Inclusion of additional GAUTs from more recently sequenced plant genomes (*e.g.* grape, soybean, sorghum and *Brachypodium*) has substantiated the existence of seven distinct GAUT subclades, each with strong statistical support (bootstrap values >90%) (Fig. 6.2). For the purposes of this review, we have designated the seven GAUT subclades (a–g) without implying specific evolutionary relatedness between the subclades, since the evolutionary relationships between the GAUT subclades is unclear because of the relatively weak statistical support for these nodes. However, there is very strong statistical support for the relatedness of genes within a subclade (see Fig. 6.2). In addition, the structures of the *GAUT* genes are more conserved within subclades than across the subclades. For example, among the seven GAUT subclades, the GAUT-b and -c subclades have one or two introns, many fewer than the other GAUT subclades (Fig. 6.2).

All of the major GAUT subclades contain both monocot and dicot proteins (Fig. 6.2). This is, at first glance, somewhat surprising given the functional association of the GAUTs with pectic polysaccharide biosynthesis (see below) and the fact that pectin constitutes only about 10% of the walls of monocots (Mohnen *et al.* 2008). This broad distribution of GAUT proteins in all GAUT clades must then reflect a key role for these glycosyltransferases in all higher plants.

The relative distributions of monocot genes vs. dicot genes in each GAUT subclade are not uniform (Fig. 6.2). Thus, the GAUT-g clade has almost twice as many monocot representatives as dicot representatives. Furthermore, not all GAUTs have orthologues in all plants. For example, no monocot genes orthologous to *AtGAUT12* could be identified, which is consistent with functional studies (see below). Likewise there are no obvious orthologues to *AtGAUT2* in any plant included in the current analysis, suggesting that this Arabidopsis gene may be a pseudogene. Whether or not the individual GAUT subclades identified through phylogenetic analyses reflect the function(s) of their members with respect to synthesis of specific plant cell

wall polysaccharides remains unclear and awaits more detailed functional studies on representative *GAUT* genes from each subclade.

6.3.1.2 Function of GAUT proteins

All of the *GAUT* genes, with the possible exception of *GAUT2* which may be a pseudogene, are expressed in all tissues examined to date, consistent with a function for GAUTs in wall biosynthesis. Some *GAUT*s, however, show higher levels of expression in some tissues. For example, *GAUT8* and *GAUT12* are highly expressed in vascular tissues in Arabidopsis stem (Orfila *et al.* 2005; Peña *et al.* 2007; Persson *et al.* 2007; Caffall *et al.* 2009).

Functional studies of GAUT proteins have relied largely on molecular genetic and biochemical studies. The latter involved partial purification of GAUT1 from solubilized membrane fractions of Arabidopsis suspension-cultured cells. Proteomic analysis of the trypsin-digested solubilized protein fraction with the greatest amount of GalAT activity revealed the presence of only two putative glycosyltransferases; AtGAUT1 and AtGAUT7. Heterologous expression of AtGAUT1 in human embryonic kidney cells yielded α1,4-GalAT activity in the presence of homogalacturonan oligosaccharide acceptors (Sterling *et al.* 2006). Furthermore, anti-GAUT1 polyclonal antibodies immunoabsorbed GalAT activity from the partially purified Arabidopsis GalAT enzyme preparations. These data provide explicit biochemical evidence that AtGAUT1 is a pectin homogalacturonan (HG) α1,4-galacturonosyltransferase (HG:GalAT) (Sterling *et al.* 2006) which transfers GalA from UDP-GalA on to the non-reducing end of homogalacturonan acceptors. This is the only GAUT for which enzyme activity has been confirmed *in vitro*. Interestingly, although no GalAT activity was detected in heterologously expressed AtGAUT7, AtGAUT7 co-purifies and co-immunoprecipates with AtGAUT1 and with HG:GalAT activity (M. A. Atmodjo, Y. Sakuragi, X. Zhu, A. Burrell, S.S. Mohanty, J.A. Atwood III, R. Orlando, H.V. Scheller and D. Mohnen, submitted for publication), suggesting that AtGAUT7 and AtGAUT1 are components of a GalAT protein complex involved in homogalacturonan biosynthesis.

Analyses of Arabidopsis plants carrying mutations in individual *GAUT* genes have provided evidence for the involvement of GAUT proteins in the synthesis of several polysaccharides present in higher plant cell walls. For example, *gaut8* (*qua1-1*) mutants show defects in cell adhesion and have reduced wall GalA and Xyl content, and a *qua1-1* mutant stem microsomal membrane protein preparation shows reduced GalAT and xylan synthase activity compared to WT (Orfila *et al.* 2005; Brown *et al.* 2007), suggesting that that GAUT8 is involved in pectin and/or xylan synthesis. Biochemical and immunohistochemical analyses of *gaut12* (*irx8*) mutants, which have a collapsed xylem phenotype (Brown *et al.* 2005), demonstrated that these mutants have somewhat reduced levels of GalA in their walls (Persson *et al.* 2007) and reduced stem secondary wall xylan (Persson *et al.* 2007; Peña *et al.* 2007). Chemical studies of *gaut12* walls revealed the absence of a Xyl-

and GalA-containing pentasaccharide at the ends of xylan chains (Peña *et al.* 2007). The results of these biochemical analyses of *gaut12* mutants have lead to the proposal that GAUT12 may be an α1,4-GalAT that participates in the synthesis of the so-called xylan primer/cap (Peña *et al.* 2007) or of a specific subfraction of pectic homogalacturonan (Persson *et al.* 2007). Interestingly, monocots lack any apparent *GAUT12* homolog, suggesting that GAUT12 may have a specialized function in the synthesis of glucuronoxylan in dicots. Lastly, a function for GAUT11 in Arabidopsis seed mucilage and wall production has been proposed recently, based on associated changes in seed mucilage production and expansion in Arabidopsis *gaut11* mutants, suggesting a role for this protein in rhamnogalacturonan synthesis (Caffall *et al.* 2009). For GAUT8, GAUT11, and GAUT12, however, no data exist, as yet, that would permit conclusions about the biochemical functions of these proteins.

An extensive analysis of the glycosyl residue compositions of cell walls isolated from 26 homozygous T-DNA mutants of 13 of the 15 Arabidopsis genes has provided additional information about GAUT protein function (Caffall *et al.* 2009). These analyses demonstrated that *gaut6, 8, 9, 10, 11, 12, 13* and *14* mutants have significant and reproducible changes in the levels of galacturonic acid, xylose, rhamnose, galactose and/or arabinose compared to walls of wild type Arabidopsis grown under the same growth conditions (Caffall *et al.* 2009). These results strongly implicate these GAUTs as being involved in the synthesis of at least six different biosynthetic linkages in pectins and/or xylans, based on the different patterns of changes in wall composition observed in the distinct *gaut* mutant walls. An initial hypothesis for GAUT function, based on the demonstrated homogalacturonan galacturonosyltransferase activity for GAUT1, is that the GAUTs function as GalATs involved in pectin synthesis. The phenotypes of several *gaut* mutants are compatible with this hypothesis. The walls of *gaut6, 9, 10* and *11* mutants have significant reductions in GalA, suggesting a role in pectin, and possibly homogalacturonan synthesis, for the proteins encoded by these genes. Indeed, more detailed studies of GAUT6, including preliminary heterologous expression data, provide compelling evidence that GAUT6 is a putative pectin biosynthetic HG:GalAT (Caffall 2008; Caffall *et al.* 2009). Cell walls of the *gaut 13* and *14* mutants have increased GalA and Gal content and reduced Xyl and Rha content compared to wildtype plants. Such changes could reflect a role for these genes in rhamnogalacturonan I (RG-I) synthesis, although such a function would require detailed enzymological studies of the expressed proteins. The inability to recover homozygous *gaut1* or *gaut4* mutants, and the high expression level of these genes in Arabidopsis, suggests a critical function for these proteins in plants. GAUT4 has 83% amino acid similarity to GAUT1, and a reasonable hypothesis is that GAUT4 is also a GalAT involved in pectin, and possibly, homogalacturonan synthesis. However, proof of this requires critical studies of the enzyme activity of the expressed protein. In summary, the available results are consistent with a

role for the entire GAUT family as GalATs involved in pectin, and possibly xylan, synthesis.

6.3.2 Galacturonosyl transferase-like (GATL) clade

A family of genes closely related to the GAUT family was identified on the basis of protein alignments (Lao *et al.* 2003; Sterling *et al.* 2006), and named galacturonosyl transferase-like (GATL) genes based on the conservation of specific amino acid motifs between the GATL and GAUT proteins (Sterling *et al.* 2006). *GATL* genes and their corresponding proteins are distinguishable from GAUTs based on three characteristics: (1) *GATL* genes and proteins are smaller than GAUTs; (2) GATL proteins lack an identifiable trans-membrane domain that is found in almost all GAUTs; and (3) *GATL* genes for the most part lack introns (Fig. 6.3), unlike the *GAUT* genes, which typically have extensive intron/exon structures (Fig. 6.2). Nonetheless, the conservation of

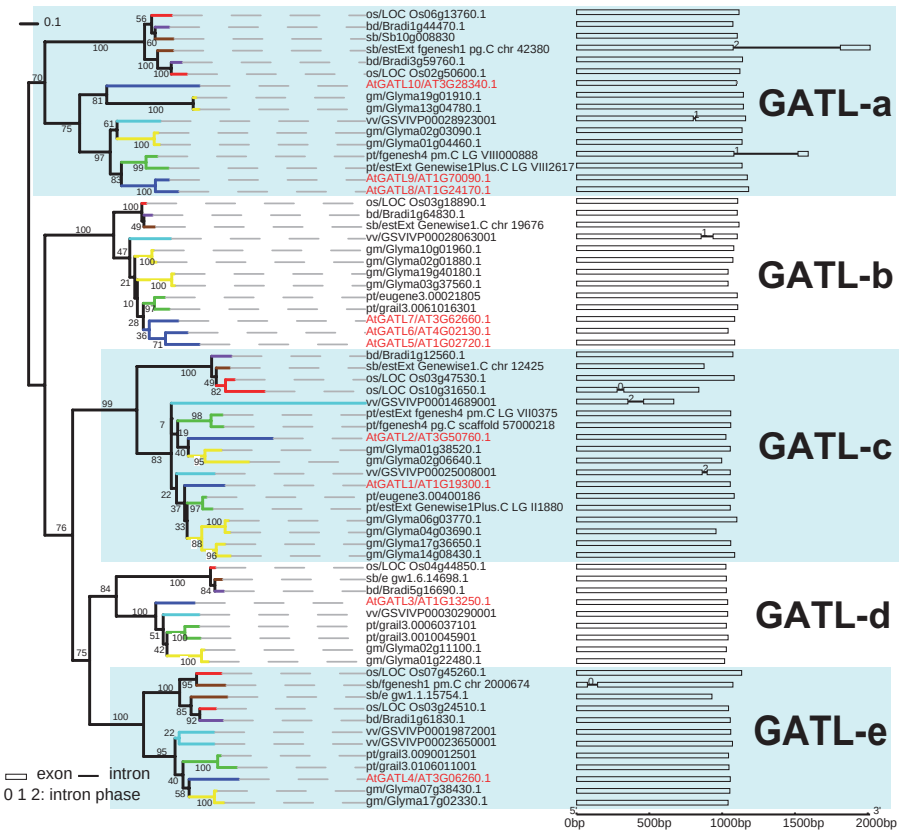

Figure 6.3 The phylogeny of the GATL clade and gene structures of the *GATL* genes. See legend of Fig. 6.2 for further details.

specific amino acid motifs in the C-terminal domains of both GAUTs and GATLs suggest possible functional commonalities, particularly with respect to their roles in cell wall biosynthesis (Sterling *et al.* 2006; Yin *et al.* 2010).

6.3.2.1 Phylogeny of GATL clade

Phylogenetic analysis of full-length GATL protein sequences based on genes from the fully sequenced plant genomes, using PhyML, a maximum likelihood algorithm (Guindon & Gascuel 2003), yields a tightly clustered tree with short branch lengths (Fig. 6.3). The latter suggests that the GATLs have diverged less than have the GAUTs, whose tree overall has significantly longer branch lengths (Fig. 6.2). The GATLs cluster into five distinct clades (a–e) with reasonable statistical support for the nodes defining these clades. Further support for this GATL tree topology is provided by the fact that each GATL clade subdivides into a monocot clade and a dicot clade, a frequent observation in gene family phylogeny comparisons between monocots and dicots. The latter also implies that the GATLs have evolved separately in the two major higher plant lineages. Thus, it may prove problematic to infer GATL function across the monocot/dicot divide.

Among the seven fully sequenced plant genomes examined in this overview of family GT8, Arabidopsis and soybean have the largest diversity of GATLs. *AtGATL10* and *Glyma19g01910* have no easily recognizable counterparts in the other five plants. The subdivision of the GATLs into five subclades tempts one to infer five distinct functions for GATL genes in higher plants. However, very few experimental studies have been done to identify GATL function (see below), so firm conclusions on this point are premature. There is some pairing of GATL genes in the tree: for example, *AtGATL5* with *AtGATL6* and *AtGATL7*, *AtGATL8* with *AtGATL9*, and the pairing of all *GATL*s from poplar (arising from the recent whole genome duplication event in the evolution of the poplar genome; see Tuskan *et al.* 2006). While such gene/protein pairings suggest functional redundancies within each pair, this must be proved experimentally. The little experimental evidence that exists to date, for *AtGATL5* and *AtGATL6* (see below) and *PdGATL1.1* and *PdGATL1.2* (Kong *et al.* 2009; Lee *et al.* 2009), suggests that these pairs of genes are not fully redundant functionally and that some functional specialization of these gene/protein pairs has occurred since duplication and divergence.

6.3.2.2 Function of GATL proteins

Currently, little is known about the function(s) of most of the GATL proteins. What functional evidence exists is largely derived from protein localization, gene expression and mutational studies.

Analysis of the domain structure of the GATLs using sequence analysis tools (as predicted by TMHMM v2.0; Krogh *et al.* 2001) reveals that about half of the GATL proteins from the fully sequenced plant genomes are

predicted to contain a single transmembrane domain (Table 6.1). However, close examination of the GATL sequences reveals that these predicted transmembrane domains are located at or only a few amino acids away from the N-terminus of the protein and are thus, most likely, signal peptides. Those GATL proteins not predicted to contain a transmembrane domain are predicted to have a signal peptide (Table 6.1). Thus, all GATL proteins appear to be directed to the endomembrane system. No other consensus membrane-anchoring sequences (e.g. for the attachment of lipid anchors) were identified in the GATL amino acid sequences. Thus, most GATLs are predicted to be soluble proteins.

The predicted domain structure of the GATL proteins raises the possibility that these proteins are not involved in cell wall synthesis at all, but rather play roles as glycosylating agents in other cellular processes (e.g. glycosylation of hormones and other metabolites), perhaps in subcellular compartments other than the Golgi where most matrix cell wall polysaccharides are thought to be synthesized (Mohnen *et al.* 2008). Thus, one key element in the discovery of GATL function is the establishment of the intracellular location of GATL proteins.

The only experimental studies on GATL localization published thus far have been carried out on GATLs from Arabidopsis and poplar. AtGATL1 (PARVUS) has been reported to be localized to the endoplasmic reticulum (ER), on the basis of heterologous expression studies carried out in carrot protoplasts (Lee *et al.* 2007). AtGATL1, PdGATL1.1 and PdGATL1.2 were localized to both ER and Golgi on the basis of transient expression studies in *Nicotiana benthemiana* leaves (Kong *et al.* 2009; Y. Kong, G. Zhou, Y. Yin, Y. Xu, S. Pattathil and M.G. Hahn, submitted for publication). Other as yet unpublished studies in Arabidopsis have demonstrated that several other AtGATL proteins (AtGATL2, AtGATL5, AtGATL6, and AtGATL9) are localized in the Golgi, and in some cases, in the ER (Y. Kong, G. Zhou, Y. Yin, Y. Xu, S. Pattathil and M.G. Hahn, submitted for publication). Thus, the available evidence indicates that GATL proteins are in the correct subcellular compartment to participate in cell wall or glycoprotein synthesis. Given the apparent absence of transmembrane or other membrane-anchoring domains in these proteins, another mechanism must be responsible for retaining the GATL proteins in the endomembrane system. Otherwise, the GATL proteins would be secreted via the default secretory pathway that targets proteins entering the endomembrane system to the apoplast, unless they are otherwise directed by signal sequences (Rojo & Denecke 2008). It is tempting to speculate that GATLs are retained in the endomembrane system because they participate with bona fide membrane-anchored proteins in biosynthetic complexes that are responsible for cell wall polysaccharide biosynthesis. However, most such enzymes that have been studied show Golgi localization patterns (Mohnen *et al.* 2008), and the ER localization of at least some GATLs is difficult to reconcile with such a model. Thus, the mechanism by which GATLs are retained within the endomembrane system remains to be experimentally determined.

Studies in Arabidopsis have demonstrated that most *GATL* genes are expressed throughout the plant, though with higher levels of expression for individual *GATL* genes in particular tissues or cell types (Kong *et al.* 2009; Y. Kong, G. Zhou, Y. Yin, Y. Xu, S. Pattathil and M.G. Hahn, submitted for publication). Transcriptomic analyses in poplar and Arabidopsis have documented co-expression of at least some of the *GATL*s with other genes known to be involved in cell wall synthesis. For example, *GATL1* has been found to be co-regulated with other cell wall genes in xylem (Ko *et al.* 2006; Leplé *et al.* 2007; Minic *et al.* 2009), as have *GATL6* and *GATL9* (Kubo *et al.* 2005). These results suggest an involvement of GATL1 in the synthesis of polysaccharides important for secondary wall formation, such as xylan (Brown *et al.* 2007; Lee *et al.* 2007; Kong *et al.* 2009). *GATL5* has been shown to be co-expressed with other genes involved in synthesis of seed mucilage (Y. Kong, G. Zhou, M. Atmodjo, A.M.A. Abdeen, T. Western, D. Mohnen and M.G. Hahn, in preparation). Other studies have placed *GATL* genes in context with other cell wall biosynthetic genes whose expression is altered by responses to nutrient depletion (*GATL8*, *GATL9*; Vlieghe *et al.* 2003) or environmental stress (*GATL4*; Tsabary *et al.* 2003) and developmental programmes (*GATL7*; Mandaokar *et al.* 2003). Whether or not such correlations observed in transcriptomic analyses allow conclusions about functional associations of GATLs with other cell wall synthetic or modifying enzymes awaits experimental verification. In any case, these transcriptomic studies imply a role for GATLs in polysaccharide biosynthesis.

Mutational studies have provided some experimental support for the hypothesis that GATLs are involved in the synthesis of several cell wall polysaccharides, including xylan (GATL1) and pectins (GATL5 and GATL6). Mutations in *GATL1* (*parvus*, *glz1*) were first isolated as plants with reduced stature (Lao *et al.* 2003; Shao *et al.* 2004). Early glycosyl composition analyses of leaves from the *parvus* mutant showed decreased xylose content and modest increases in rhamnosyl and galactosyl residues, particularly in young leaves (Lao *et al.* 2003). These results, indicating an involvement of GATL1 in xylan and possibly pectin synthesis, have been substantiated, at least in part, by recent studies of xylan synthesis in *parvus* mutants (Brown *et al.* 2007; Lee *et al.* 2007; Kong *et al.* 2009; Lee *et al.* 2009). These studies focused on secondary wall formation in stem and root xylem tissues and examined changes in xylan structure in the mutant plants. The thickness of the secondary walls in xylem tissues are dramatically reduced in *parvus*. Furthermore, a tetrasaccharide present at the non-reducing terminus of xylan chains in wildtype plants (Peña *et al.* 2007) is absent in the mutant and there is a significant decrease in xylan content of mutant walls (Brown *et al.* 2007; Lee *et al.* 2007). However, the phenotype of *parvus/glz1* mutants is complex, with many character traits (e.g. stature, organ size, pollen production and viability) affected, in addition to the effects on secondary wall formation (Lao *et al.* 2003; Shao *et al.* 2004). Not all of these mutant characters can easily be explained solely on the basis of defects in xylan synthesis and additional studies will be necessary to exclude a role for

GATL1 in pectin biosynthesis as had been suggested in earlier studies (Lao *et al.* 2003).

A systematic analysis of insertional mutations in other *GATL* genes has resulted in the association of at least some GATLs with the synthesis of pectic polysaccharides in Arabidopsis. Mutations in *AtGATL5* result in a significant decrease in the synthesis of mucilage in Arabidopsis seeds. The effect of the *gatl5* mutation is enhanced by mutation of a second, closely related gene (Fig. 6.3), *AtGATL6*, resulting in the almost complete absence of seed mucilage. The single *gatl6* mutant on its own shows little if any effect on mucilage biosynthesis. Arabidopsis seed mucilage is largely composed of unbranched rhamnogalacturonan I (Deng *et al.* 2009). Further detailed characterization of the *gatl5* mutant provides strong evidence that GATL5 participates in the synthesis of pectic polysaccharides (Y. Kong, G. Zhou, M. Atmodjo, A.M.A. Abdeen, T. Western, D. Mohnen and M.G. Hahn, in preparation). In addition to its expression in developing seeds, *GATL5* is expressed throughout the plant, particularly in the vasculature of root and stem. These results suggest that GATL5 participates in pectin synthesis in other cellular contexts besides mucilage synthesis.

An insertional mutation in *AtGATL9* leads to defects in cell adhesion (Y. Kong, G. Zhou and M.G. Hahn, unpublished results), a phenotype that has been observed in plants carrying mutations in genes thought to be associated with pectic polysaccharide biosynthesis (Bouton *et al.* 2002) (see previous section on *GAUT* genes). A 20% reduction in GalA content was detected in cell walls of *gatl9* mutants, and over-expression studies also suggest that GATL9 participates in pectin biosynthesis. However, further detailed studies are necessary to demonstrate definitively that GATL9 is a pectin biosynthetic enzyme.

Arabidopsis plants carrying single mutations in other *GATL* genes do not display easily identifiable morphological or cell wall abnormalities (Y. Kong and M.G. Hahn, unpublished results), suggesting that some redundancies exist with regard to GATL function at least in this plant.

6.4 GT8 clades not related to cell wall synthesis

The proteins in the remaining GT8 clades have not been tied experimentally to cell wall synthesis. Three clades (PGSIP clades) include proteins that were originally identified (Qi *et al.* 2005; Chatterjee *et al.* 2005) on the basis of their sequence similarity to fungal and mammalian proteins, called glycogenins, which are self-glycosylating proteins that serve as initiators of glycogen biosynthesis in those organisms (Roach & Skurat 1997). The other clade (GolS clade) contains proteins that have been associated with the synthesis of raffinose family oligosaccharides that appear to have roles in plant stress responses. These non-cell-wall-related GT8 clades include proven and putative α-D-glycosyltransferases that use protein or inositol acceptors and

glucose or galactose nucleotide sugars as donors. All of these clades have characteristic features and sequence motifs that distinguish them from each other and from the GT8 families related to cell wall synthesis, discussed previously.

6.4.1 Plant glycogenin-like (PGSIP) clades

All living organisms store chemical energy in some way in order to survive when their living conditions change. The energy-storing molecules are synthesized at times when nutrition is easily available, and are used when sustenance is scarce or in times of stress. The energy storage carbohydrate in most organisms other than plants is glycogen. The biosynthesis of glycogen, an α1,4-linked glucan with α1,6 branch points at approximately every ten glucosyl residues, is a three-step process that includes initiation, elongation and branching (Alonso *et al.* 1995a). The first step in the biosynthesis of glycogen, initiation, is carried out by glycogenin, a self-glucosylating protein that serves as an initiator of glycogen synthesis, and is believed to play a key regulatory role in glycogen synthesis (Roach & Skurat 1997; Lomako *et al.* 2004). The initiation step involves transfer of a glucosyl residue from UDP-Glc to a single tyrosyl residue on glycogenin forming a Glc-1-*O*-tyrosyl linkage on glycogenin (Smythe *et al.* 1988). This step is one of several glycogenin self-glucosylation reactions, which are repeated about ten times to yield glycogenin with a covalently bound oligoglucan oligosaccharide. It has been suggested that glycogenin exists functionally as a dimer in which one subunit glycosylates the other (Alonso *et al.* 1995b).

Starch, which is used as a storage polymer in plants and is synthesized in the chloroplast, is structurally similar to glycogen, in that it also consists of an α1,4-linked glucan with α1,6 branch points. Unlike glycogen, however, starch is made of two basic polymers, amylopectin and amylose. The branched amylopectin has approximately 1 branch point per 30 glucose units while amylose is less branched (Hoover 2001). Some plant starches, such as waxy maize starch, are composed only of amylopectin (Angellier-Coussy *et al.* 2009). Starch is semicrystalline, with granular sizes ranging from ~2 μm in rice up to 85 μm in other plant species. Another difference between glycogenin and amylopectin, aside from the frequency of branching, is that the activated sugar for glycogenin initiation is UDP-Glc, while ADP-Glc is used for starch synthesis (Ball *et al.* 1998). Although glycogen and amylopectin have very similar chemical characteristics, glycogen is fully soluble and non-crystalline (Calder 1991), while amylopectin is more crystalline (Ball *et al.* 1996).

Plant proteins with sequence similarity to mammalian glycogenin were first identified in Arabidopsis (Chatterjee *et al.* 2005) and rice (Qi *et al.* 2005), and in two thermo-acidophilic red algae (Barbier *et al.* 2005). These proteins were named plant glycogenin-like starch initiation proteins (PGSIP) (Chatterjee *et al.* 2005), which we also refer to in this review as plant

glycogenin-like proteins, and it has been suggested that they may be involved in the initiation of starch amylopectin biosynthesis in a manner similar to glycogenin in glycogen synthesis. Some of the genes encoding PGSIPs have been shown to be upregulated by auxin (Goda *et al.* 2004), wounding (Guan & Nothnagel 2004), and by environmental stress (Qi *et al.* 2005) or disease (Schaff *et al.* 2007).

6.4.1.1 Phylogeny of PGSIP clades

Phylogenetic analysis using PhyML, a Maximum Likelihood algorithm (Guindon & Gascuel 2003), distributed the glycogenin-like proteins from the seven higher plant genomes into three distinct phylogenetic clades whose evolutionary relationships are not completely clear due to the relatively weak statistical support for some of the nodes in this part of the family GT8 tree (see Fig. 6.1). The PGSIP-B and PGSIP-C clades likely derive from a common ancestor. However, the evolutionary relationship of these two clades with the PGSIP-A clade is not clear.

All PGSIP proteins contain the following consensus MEME motif (Bailey & Elkan 1994):

YSK[FL]RLWQLTDYD[KR][VI][VI]F[IL]DADL[LI][VI]L
[RK]NIDFLFA[CM][PG][QE]

In addition, each PGSIP clade is characterized by one or more MEME motifs that are unique to the proteins in a given clade. The functional significance, if any, of the distribution of the PGSIP proteins into three clades cannot be ascertained at present, as no plant PGSIP protein has yet been functionally characterized (see below).

The PGSIP-A clade contains both monocot and dicot PGSIP genes that are distributed among several subclades (Fig. 6.4). However, the orthologous relationships across the monocot/dicot divide in this clade are not clear for all subclades. For example, there are two monocot subclades most closely related to two dicot subclades containing AtPGSIP1 (At3g18660) and AtPGSIP2 (At1g77130), but the two monocot clades branch earlier than the two dicot clades. The three monocot genomes included in this analysis do not appear to contain genes orthologous to AtPGSIP4 or AtPGSIP5, suggesting that these gene products play distinctive roles in dicots.

The PGSIP-B and PGSIP-C clades each consist of single monophyletic groups of genes split among monocots and dicots (Fig. 6.5, Fig. 6.6). A distinguishing characteristic of the PGSIP-B clade is that the genes in this clade have, on average, twice as many exons as do the genes in either the PGSIP-A or PGSIP-C clades. The exon/intron gene structure is also highly conserved within both the PGSIP-B and PGSIP-C clades (Fig. 6.5, Fig. 6.6), but less well conserved within the PGSIP-A clade (Fig. 6.4).

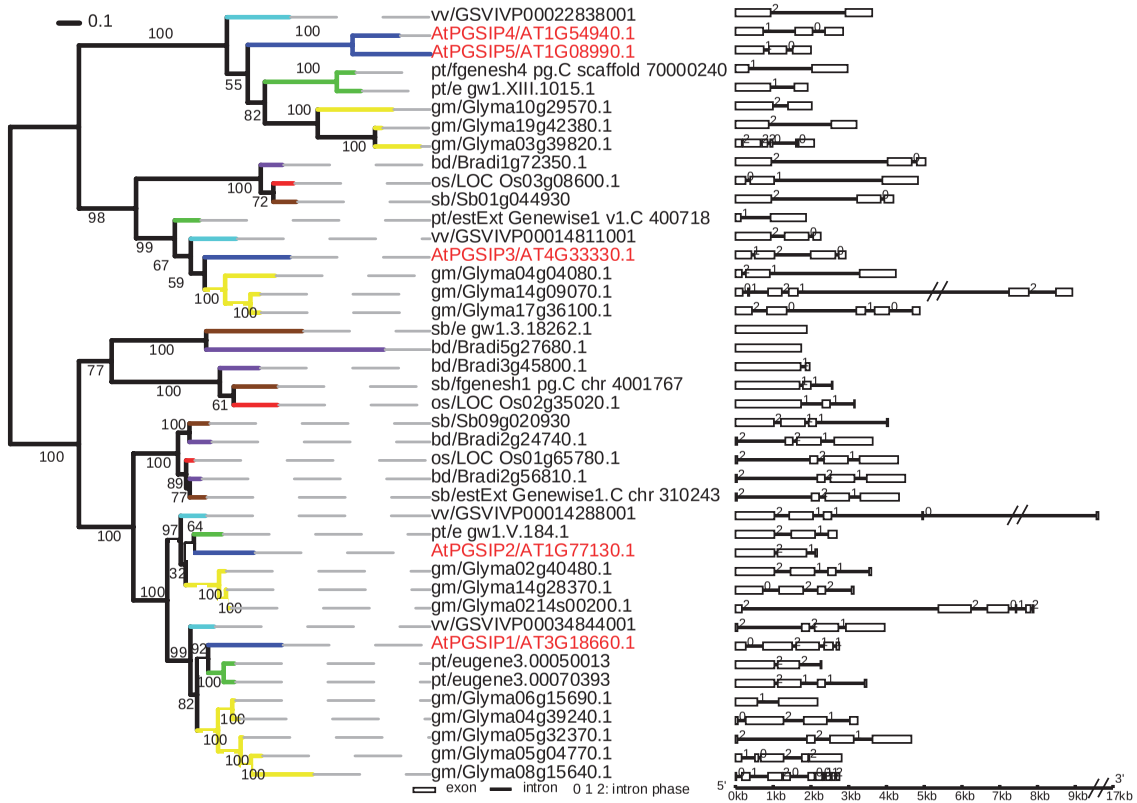

Figure 6.4 The phylogeny of the PGSIP-A clade and gene structures of *PGSIP-A* genes. See legend of Fig. 6.2 for further details.

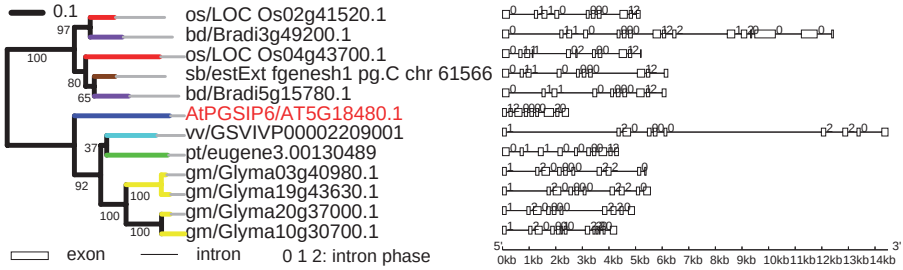

Figure 6.5 The phylogeny of the PGSIP-B clade and gene structures of the *PGSIP-B* genes. See legend of Fig. 6.2 for further details.

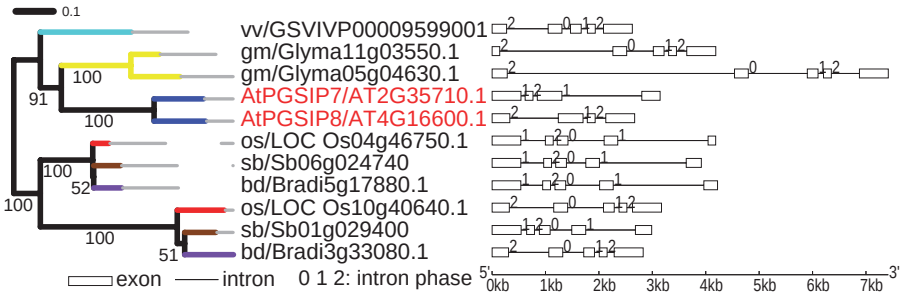

Figure 6.6 The phylogeny of the PGSIP-C clade and gene structures of the *PGSIP-C* genes. See legend of Fig. 6.2 for further details.

6.4.1.2 Function of PGSIP proteins

The function(s) of plant glycogenin-like proteins remains to be established. What little information exists originates from mutational, bioinformatic and transcriptomic analyses. There are published reports of plant self-glucosylating proteins and polypetides (Ardila & Tandecarz 1992; Singh *et al.* 1995; Langeveld *et al.* 2002), but none have been shown to be involved in the initiation of starch biosynthesis. The initial discovery and naming of the PGSIP protein clades was made based on a search of the Arabidopsis genome for genes encoding proteins with sequence similarity to mammalian and yeast glycogenin (Chatterjee *et al.* 2005). Six of these proteins (AtPGSIP1–6) have amino acid motifs consistent with a possible function of these proteins as glycosyltransferases. However, they lack the conserved Tyr residue that is the site of self-glycosylation in glycogenin. In yeast, glycogenin is glycosylated at multiple Tyr sites (Mu *et al.* 1996) and the Arabidopsis PGSIPs do contain alternative Tyr residues that could be used as the sites for self-glycosylation. More critically, only two Arabidopsis glycogenin-like proteins, AtPGSIP1 and AtPGSIP3, are predicted (see http://www.cbs.dtu.dk/services/ChloroP/; Emanuelsson *et al.* 1999) to contain a transit peptide to target the polypeptide to the chloroplast (Chatterjee *et al.*

2005 and unpublished results), the organelle in which starch biosynthesis occurs. Indeed, plants carrying a knock-out mutation in PGSIP1 are starch-deficient in their leaves (Chatterjee *et al.* 2005). These results suggest that PGSIP1 plays a role in starch biosynthesis, at least in Arabidopsis. However, no enzymatic activity has been demonstrated for this protein as yet. The absence of a clear transit peptide for chloroplast targeting in the other PGSIP proteins raises questions about the possible role(s) of these proteins in starch biosynthesis. To date, no experimental verification of the subcellular localization of any of the PGSIPs has been reported.

AtPGSIP7 and AtPGSIP8 lack conserved motifs involved in binding of the donor ligand, UDP-Glc, and also lack a conserved His residue (in the HxxGxxKPW motif at position 277) important for Mn^{2+} binding (Chatterjee *et al.* 2005). Thus, it is doubtful that the proteins in the PGSIP-C clade are involved in starch biosynthesis, and detailed studies will be required to define the function(s) of the proteins in this clade.

Transcriptomic analyses have revealed that PGSIPs are up-regulated in response to hormones (Goda *et al.* 2004; Qi *et al.* 2005), wounding (Guan & Nothnagel 2004), and environmental stress (Qi *et al.* 2005) or disease (Schaff *et al.* 2007). However, these analyses by themselves are not enlightening as to the function(s) of the up-regulated PGSIPs. Interestingly, bioinfomatic identification of Arabidopsis genes co-expressed with cellulose synthases (CESA) identified, among others, AtPGSIP1 as being co-expressed with secondary cell wall CESA 4, 7, and 8 (Persson *et al.* 2005). These results suggest a possible role for at least this PGSIP in secondary cell wall biosynthesis in Arabidopsis. These transcriptomic and bioinformatic analyses will need to be followed up with detailed molecular genetic and biochemical studies to define the functions of PGSIPs in plants.

6.4.2 Galactinol synthase (GolS) clade

Plants are constantly exposed to changes in their environment including nutrient depletion, temperature extremes, and drought. To survive such challenges plants have developed means to store energy, maintain cellular integrity and preserve cellular metabolism. One such survival mechanism is through the formation of carbohydrates that can be transported and/or stored, and that can serve as protectant molecules (Peterbauer & Richter 2001; Browse & Lange 2004). Raffinose family oligosaccharides (RFOs) are examples of such carbohydrates.

RFOs are water-soluble non-reducing carbohydrate molecules. There is a substantial body of literature on RFO accumulation and function in plants (Peterbauer & Richter 2001; Browse & Lange 2004), so only a few highlights and examples will be presented here. All plants, at some point, synthesize some RFOs, but many neither transport nor accumulate large quantities in their tissues and/or organs. For example, ryegrass has been shown to accumulate small amounts of raffinose and loliose under normal, non-stress,

conditions (Amiard *et al.* 2003). In contrast, other plants store RFOs in large concentrations, sometime 25–80% of their dry weight, in tubers (French 1954; Keller & Matile 1985) and in photosynthesizing leaves where they are localized in mesophyll cells (Senser & Kandler 1967; Bachmann *et al.* 1994). RFOs also accumulate in seeds (Korytnyk & Metzler 1962, Handley *et al.* 1983a; Handley *et al.* 1983b; Blackman *et al.* 1992; Haritatos *et al.* 2000; Peterbauer *et al.* 2001), as well as in vegetative tissues of many plants (French 1954; Zimmermann 1957; Handley *et al.* 1983a), and serve as the predominant transport carbohydrate in cucurbits (Handley *et al.* 1983a; Keller & Pharr 1996).

RFOs are believed to prolong seed survival and storage by protecting the embryo during the desiccation that occurs during seed maturation (Peterbauer *et al.* 2001). Plants can also accumulate RFOs in response to diverse abiotic stresses (Nishizawa *et al.* 2008) such as temperature extremes or drought. Under freezing conditions, for example, ice forms first in the extracellular compartments of plants that are exposed to low temperatures, causing a reduction of water potential and consequent loss of water from cells by osmosis (Browse & Lange 2004). Similar processes take place under drought conditions due to desiccation and dehydration. RFOs, under these conditions, are thought to act as osmolytes to maintain cellular integrity and function (Nishizawa *et al.* 2008).

RFOs are synthesized by a transfer of a galactosyl residue from galactinol to the glucosyl residue of sucrose, resulting in an α1,6 linkage. The first step in the synthesis of any RFO is the synthesis of galactinol that is catalysed by galactinol synthases (GolS), which are members of CAZy family GT8. The only known function for galactinol, to date, is that of a galactosyl donor in RFO biosynthesis (Saravitz *et al.* 1987; Peterbauer *et al.* 2002). Raffinose (Gal-Suc), stachyose (diGal-Suc), verbacose (triGal-Suc), etc. are formed as a result of the action of several galactosyltransferases (e.g. raffinose synthase, stachyose synthase, etc.) (Peterbauer *et al.* 2001).

6.4.2.1 Phylogeny of GolS clade

Phylogenetic analysis using PhyML, a maximum likelihood algorithm (Guindon & Gascuel 2003), distributed the galactinol synthases from the seven higher plant genomes into a single, strongly supported monophyletic clade (see Fig. 6.1). The evolutionary relationship of the GolS clade to the other non-cell-wall synthetic GT8 clades remains unclear due to the relatively weak statistical support for the nodes in this part of the family GT8 tree (see Fig. 6.1).

The subclade structure within the GolS clade is unclear, as most of the nodes between subgroups of GolS proteins have very weak statistical support (Fig. 6.7). Such weak support for subclades within the GolS clade was also reported in earlier phylogenetic analyses of GolS sequences (Volk *et al.* 2003; Zhao *et al.* 2004). Nonetheless, it is clear that monocots have a smaller number and diversity of GolS genes compared with dicots, and that

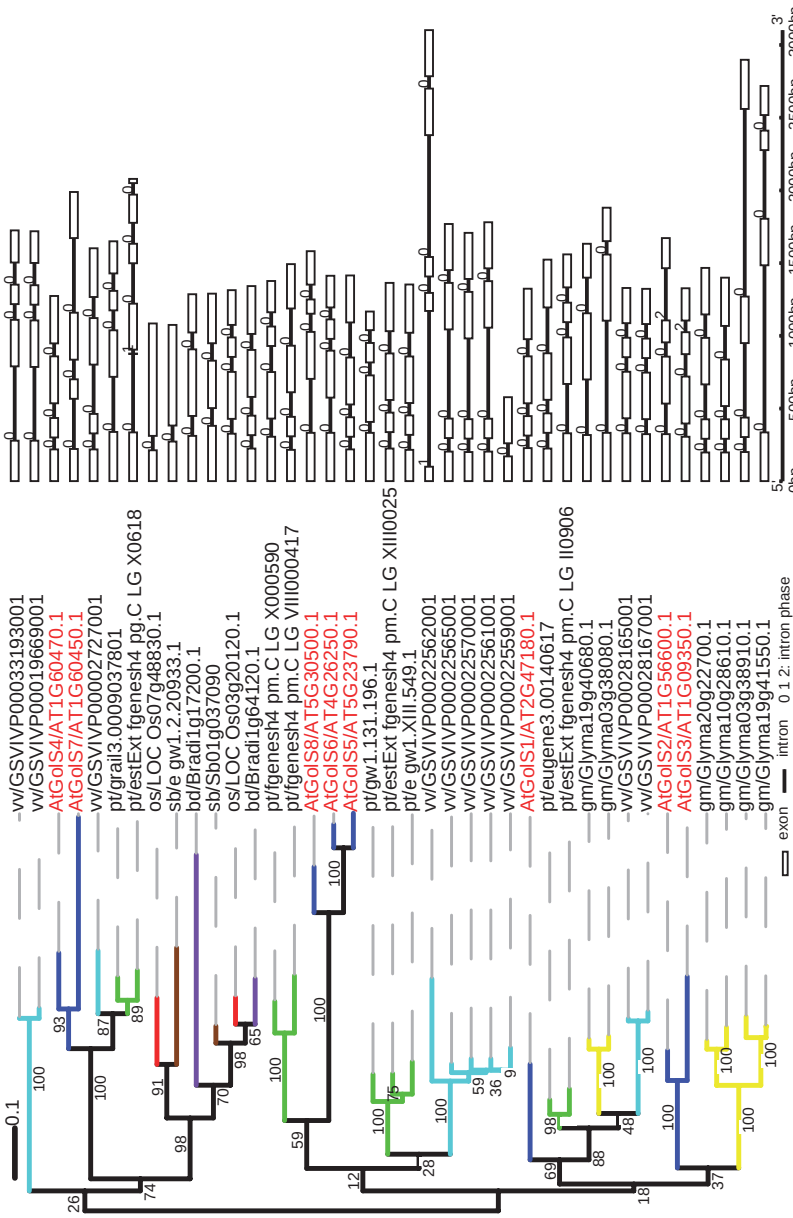

Figure 6.7 The phylogeny of the GolS clade and gene structures of the *GolS* genes. See legend of Fig. 6.2 for further details.

the monocot galactinol synthases cluster together in one subclade that is most closely related to a dicot cluster containing AtGolS4 and AtGolS7 (Fig. 6.7). Thus, it appears that dicots have developed or retained a greater diversity of uses for RFOs than have monocots.

6.4.2.2 Function of GolS proteins

Galactinol (*O*-α-D-galactopyranosyl-[1→1]-L-*myo*-inositol) is synthesized by galactinol synthase (EC 2.4.1.123) which transfers a galactosyl residue from UDP-Gal on to myo-inositol in a reversible reaction yielding galactinol and UDP. This is a regulatory step in RFO biosynthesis (Robbins & Pharr 1987; Saravitz *et al.* 1987; Hitz *et al.* 2002). Galactinol synthase was first isolated from pea seeds (Frydman & Neufeld 1963), and later from *Cucumis sativum* (Pharr *et al.* 1981), and cotyledons of kidney bean (Liu *et al.* 1995). GolS activity has been associated with RFO biosynthesis in cucumber fruit (Handley *et al.* 1983a) and maturing soybean seeds (Saravitz *et al.* 1987).

GolS enzymatic activity is readily assayable, as both donor and acceptor are known and available. Furthermore, GolS proteins do not have any transmembrane domains (Table 6.1) and are thus expected to be soluble proteins, greatly simplifying functional assays of heterologously expressed GolS proteins. Indeed, a number of GolS genes have been shown to encode functional galactinol synthases (Zhao *et al.* 2004).

There is an extensive literature on the regulation and expression of *GolS* genes in plants, particularly with respect to environmental stresses such as desiccation (including both drought and seed maturation), salinity, and low temperatures (Obendorf 1997; Peterbauer & Richter 2001). At least in dicots, it appears that different *GolS* genes have specialized functions. For example, two *GolS* genes in Arabidopsis (*AtGolS1 and AtGolS2*) (Panikulangara *et al.* 2004) and one in maize (*ZmGolS2*) (Zhao *et al.* 2004) are highly expressed in mature and dry seeds, where they most likely acting as desiccation stress response factors. Other galactinol synthase genes are expressed at low levels in seeds and are induced by diverse stresses, including temperature stresses (Taji *et al.* 2002; Fowler & Thomashow 2002; Amiard *et al.* 2003; Downie *et al.* 2003; Panikulangara *et al.* 2004; Zhao *et al.* 2004). *AtGolS1* and *AtGolS2* are also up-regulated in leaves in response to drought and salinity (Nishizawa *et al.* 2008), by a combination of high light and heat stress (Panikulangara *et al.* 2004; Nishizawa *et al.* 2006), or by treatment with peroxide (Nishizawa *et al.* 2008). *AtGolS3* (Nishizawa *et al.* 2008), and *GolS* genes in *Ajuga reptans* (Bachmann *et al.* 1994), tomato (Downie *et al.* 2003) and alfalfa (Cunningham *et al.* 2003) are induced during cold stress. GolS transcript levels decrease rapidly when plants are returned to normal growth conditions (Cunningham *et al.* 2003).

Molecular genetic studies confirmed a role for *GolS* genes and RFOs in plant responses to environmental and pathogen stress. Over-expression of *AtGOLS2* resulted in the improved drought tolerance of transgenic Arabidopsis plants (Panikulangara *et al.* 2004). These plants also showed

increased levels of galactinol and raffinose in leaves (Panikulangara *et al.* 2004). A knock-out mutation in the *AtGolS1* gene in Arabidopsis results in impaired galactinol flux and raffinose synthesis (Panikulangara *et al.* 2004).

Galactinol synthase is also shown to protect plants against pathogen infections. A cucumber galactinol synthase *CsGolS1* gene was isolated from root tissue colonized by *Pseudomonas chlororaphis* (Kim *et al.* 2008). Over-expression of *CsGolS1* in transgenic tobacco plants resulted in a constitutive resistance against pathogen infection and stimulated the accumulation of defence-related genes (Kim *et al.* 2008). The *CsGolS1* over-expressing transgenic plants also showed an increased tolerance to drought and high salt stress conditions (Kim *et al.* 2008).

6.5 Conclusions

The CAZy family GT8 genes present in higher plants are divided into two distantly related sets of proteins. One set, which includes the GAUTs and GATLs, has been strongly implicated as having roles in the synthesis of pectins and xylans, two major groups of polysaccharides present in plant cell walls. The second set, which includes PGSIPs and GolSs, appear not to be directly involved in plant cell wall synthesis. The PGSIPs have been suggested to play a role in priming starch biosynthesis, while the GolSs are key enzymes in the synthesis of the raffinose family of oligosaccharides that play important roles in environmental stress responses in plants. The large number and diversity of GT8 proteins in higher plants highlights the importance of these proteins for the biology of plants. However, much work remains to be done to define the biochemical functions of GT8 proteins, particularly those in the GAUT, GATL, and PGSIP clades.

Acknowledgements

This work was supported by the US Department of Energy (BioEnergy Science Center and grant DESG0296ER20220). The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. Research on GAUTs and GATLs is also supported in part by grants from the US National Science Foundation (MCB-0646109) and the USDA (NRI-CREES-2006-35318-17301).

References

Alonso, M.D., Lomako, J., Lomako, W.M., Whelan, W.J. (1995a) A new look at the biogenesis of glycogen. *FASEB Journal*, **9**, 1126–1137.

Alonso, M.D., Lomako, J., Lomako, W.M., Whelan, W.J. (1995b) Catalytic activities of glycogenin additional to autocatalytic self-glucosylation. *Journal of Biological Chemistry*, **270**, 15315–15319.

Amiard, V., Morvan-Bertrand, A., Billard J-P., Huault, C., Keller, F., Prud'homme M-P. (2003) Fructans, but not the sucrosyl-galactosides, raffinose and loliose, are affected by drought stress in perennial ryegrass. *Plant Physiology*, **132**, 2218–2229.

Angellier-Coussy, H., Putaux J-L., Molina-Boisseau, S., Dufresne, A., Bertoft, E., Perez, S. (2009) The molecular structure of waxy maize starch nanocrystals. *Carbohydrate Research*, **344**, 1558–1566.

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Ardila, F.J., Tandecarz, J.S. (1992) Potato tuber UDP-glucose:protein transglucosylase catalyzes its own glucosylation. *Plant Physiology*, **99**, 1342–1347.

Bachmann, M., Matile, P., Keller, F. (1994) Metabolism of the raffinose family oligosaccharides in leaves of *Ajuga reptans* L. – Cold acclimation, translocation, and sink to source transition: Discovery of chain elongation enzyme. *Plant Physiology*, **105**, 1335–1345.

Bailey, T.L., Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* (eds. R. Altman, D.L. Brutlag, P.D. Karp, R.H. Lathrop, D.B. Searls), pp 28–36. AAAI Press, Menlo Park, CA.

Ball, S., Guan H-P., James, M., *et al.* (1996) From glycogen to amylopectin: A model for the biogenesis of the plant starch granule. *Cell*, **86**, 349–352.

Ball, S.G., van de Wal, M.H.B.J., Visser, R.G.F. (1998) Progress in understanding the biosynthesis of amylose. *Trends in Plant Science*, **3**, 462–467.

Barbier, G., Oesterhelt, C., Larson, M.D., Halgren, R.G., Wilkerson, C., Garavito, R.M., Benning, C., Weber, A.P.M. (2005) Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. *Plant Physiology*, **137**, 460–474.

Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*, **340**, 783–795.

Blackman, S.A., Obendorf, R.L., Leopold, A.C. (1992) Maturation proteins and sugars in desiccation tolerance of developing soybean seeds. *Plant Physiology*, **100**, 225–230.

Bouton, S., Leboeuf, E., Mouille, G., *et al.* (2002) *QUASIMODO1* encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in Arabidopsis. *Plant Cell*, **14**, 2577–2590.

Brown, D.M., Goubet, F., Vicky, W.W., Goodacre, R., Stephens, E., Dupree, P., Turner, S.R. (2007) Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *Plant Journal*, **52**, 1154–1168.

Brown, D.M., Zeef, L.A.H., Ellis, J., Goodacre, R., Turner, S.R. (2005) Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell*, **17**, 2281–2295.

Browse, J., Lange, B.M. (2004) Counting the cost of a cold-blooded life: Metabolomics of cold acclimation. *Proceedings of the National Academy of Sciences of the U S A*, **101**, 14996–14997.

Caffall, K.H. (2008) Expression and characterization of galacturonosyltransferase-6 (GAUT6) of the galacturonosyltransferase-1 (GAUT1)-related gene family of *Arabidopsis thaliana*. PhD thesis, University of Georgia, Athens, GA.

Caffall, K.H., Pattathil, S., Phillips, S.E., Hahn, M.G., Mohnen, D. (2009) *Arabidopsis thaliana* T-DNA mutants implicate *GAUT* genes in the biosynthesis of pectin and xylan in cell walls and seed testa. *Molecular Plant*, **2**, 1000–1014.

Calder, P.C. (1991) Glycogen structure and biogenesis. *International Journal of Biochemistry*, **23**, 1335–1352.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research*, **37**, D233–D238.

Chatterjee, M., Berbezy, P., Vyas, D., Coates, S., Barsby, T. (2005) Reduced expression of a protein homologous to glycogenin leads to reduction of starch content in Arabidopsis leaves. *Plant Science*, **168**, 501–509.

Cunningham, S.M., Nadeau, P., Castonguay, Y., Laberge, S., Volenec, J.J. (2003) Raffinose and stachyose accumulation, galactinol synthase expression, and winter injury of contrasting alfalfa germplasms. *Crop Science*, **43**, 562–570.

Deng, C., O'Neill, M.A., Hahn, M.G., York, W.S. (2009) Improved procedures for the selective chemical fragmentation of rhamnogalacturonans. *Carbohydrate Research*, **344**, 1852–1857.

Dong, Q., Lawrence, C.J., Schlueter, S.D., *et al.* (2005) Comparative plant genomics resources at PlantGDB. *Plant Physiology*, **139**, 610–618.

Downie, B., Gurusinghe, S., Dahal, P., *et al.* (2003) Expression of a *GALACTINOL SYNTHASE* gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. *Plant Physiology*, **131**, 1347–1359.

Emanuelsson, O., Nielsen, H., von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*, **8**, 978–984.

Fowler, S., Thomashow, M.F. (2002) Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell*, **14**, 1675–1690.

French, D. (1954) The raffinose family of oligosaccharides. *Advances in Carbohydrate Chemistry*, **9**, 149–184.

Frydman, R.B., Neufeld, E.F. (1963) Synthesis of galactosylinositol by extracts from peas. *Biochemical and Biophysical Research Communications*, **12**, 121–125.

Gibbons, B.J., Roach, P.J., Hurley, T.D. (2002) Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. *Journal of Molecular Biology*, **319**, 463–477.

Goda, H., Sawa, S., Asami, T., Fujioka, S., Shimada, Y., Yoshida, S. (2004) Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in Arabidopsis. *Plant Physiology*, **134**, 1555–1573.

Goff, S.A., Ricke, D., Lan T-H., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). *Science*, **296**, 92–100.

Guan, Y., Nothnagel, E.A. (2004) Binding of arabinogalactan proteins by Yariv phenylglycoside triggers wound-like responses in Arabidopsis cell cultures. *Plant Physiology*, **135**, 1346–1366.

Guindon, S., Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Guo, A.Y., Zhu, Q.H., Chen, X., Luo, J.C. (2007) GSDS: a gene structure display server. *Yi Chuan*, **29**, 1023–1026.
Handley, L.W., Pharr, D.M., McFeeters, R.F. (1983a) Carbohydrate changes during maturation of cucumber fruit. Implications for sugar metabolism and transport. *Plant Physiology*, **72**, 498–502.
Handley, L.W., Pharr, D.M., McFeeters, R.F. (1983b) Relationship between galactinol synthase activity and sugar composition of leaves and seeds of several crop species. *Journal of the American Society for Horticultural Science*, **108**, 600–605.
Haritatos, E., Medville, R., Turgeon, R. (2000) Minor vein structure and sugar transport in *Arabidopsis thaliana*. *Planta*, **211**, 105–111.
Hitz, W.D., Carlson, T.J., Kerr, P.S., Sebastian, S.A. (2002) Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. *Plant Physiology*, **128**, 650–660.
Holst, O. (2007) The structures of core regions from enterobacterial lipopolysaccharides – an update. *FEMS Microbiology Letters*, **271**, 3–11.
Hoover, R. (2001) Composition, molecular structure, and physicochemical properties of tuber and root starches: a review. *Carbohydrate Polymers*, **45**, 253–267.
Hurley, T.D., Stout, S., Miner, E., Zhou, J., Roach, P.J. (2005) Requirements for catalysis in mammalian glycogenin. *Journal of Biological Chemistry*, **280**, 23892–23899.
Jaillon, O., Aury J-M., Noel, B., *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–465.
Katoh, K., Kuma, K., Toh, H., Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**, 511–518.
Keller, F., Matile, P. (1985) The role of the vacuole in storage and mobilization of stachyose in tubers of *Stachys sieboldii*. *Journal of Plant Physiology*, **119**, 369–380.
Keller, F., Pharr, D.M. (1996) Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides. In: *Photoassimilate Distribution in Plants and Crops: Source-Sink Relationships* (eds E. Zamsui, A.A. Schaffer), pp 157–183. Marcel Dekker, New York.
Kim, M.S., Cho, S.M., Kang, E.Y., *et al.* (2008) Galactinol is a signaling component of the induced systemic resistance caused by *Pseudomonas chlororaphis* O6 root colonization. *Molecular Plant Microbe Interaction*, **21**, 1643–1653.
Ko J-H., Beers, E.P., Han K-H. (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. *Molecular Genetics and Genomics*, **276**, 517–531.
Kong, Y., Zhou, G., Avci, U., *et al.* (2009) Two poplar glycosyltransferase genes, *PdGATL1.1* and *PdGATL1.2*, are functional orthologs to *PARVUS/AtGATL1* in Arabidopsis. *Molecular Plant*, **2**, 1040–1050.
Korytnyk, W., Metzler, E. (1962) Formation of raffinose and stachyose in lima beans (*Phaseolus lunatus*). *Nature*, **195**, 616–617.
Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, **305**, 567–580.
Kubo, M., Udagawa, M., Nishikubo, N., *et al.* (2005) Transcription switches for protoxylem and metaxylem vessel formation. *Genes and Development*, **19**, 1855–1860.
Langeveld, S.M.J., Vennik, M., Kottenhagen, M., *et al.* (2002) Glucosylation activity and complex formation of two classes of reversibly glycosylated polypeptides. *Plant Physiology*, **129**, 278–289.

Lao, N.T., Long, D., Kiang, S., *et al.* (2003) Mutation of a family 8 glycosyltransferase gene alters cell wall carbohydrate composition and causes a humidity-sensitive semi-sterile dwarf phenotype in Arabidopsis. *Plant Molecular Biology*, **53**, 687–701.

Lee, C., Zhong, R., Richardson, E.A., Himmelsbach, D.S., McPhail, B.T., Ye, Z.-H. (2007) The *PARVUS* gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. *Plant and Cell Physiology*, **48**, 1659–1672.

Lee, C., Teng, Q., Huang, W., Zhong, R., Ye, Z.-H. (2009) The poplar GT8E and GT8F glycosyltransferases are functional orthologs of *Arabidopsis* PARVUS involved in glucuronoxylan biosynthesis. *Plant and Cell Physiology*, **50**, 1982–1987.

Leplé, J.C., Dauwe, R., Morreel, K., *et al.* (2007) Downregulation of cinnamoyl-coenzyme A reductase in poplar: Multiple-level phenotyping reveals effects on cell wall polymer metabolism and structure. *Plant Cell*, **19**, 3669–3691.

Liu, J.J., Odegard, W., de Lumen, B.O. (1995) Galactinol synthase from kidney bean cotyledon and zucchini leaf – purification and N-terminal sequences. *Plant Physiology*, **109**, 505–511.

Lomako, J., Lomako, W.M., Whelan, W.J. (2004) Glycogenin: the primer for mammalian and yeast glycogen synthesis. *Biochimica Biophysica Acta*, **1673**, 45–55.

Mandaokar, A., Kumar, V.D., Amway, M., Browse, J. (2003) Microarray and differential display identify genes involved in jasmonate-dependent anther development. *Plant Molecular Biology*, **52**, 775–786.

Minic, Z., Jamet, E., San-Clemente, H., *et al.* (2009) Transcriptomic analysis of Arabidopsis developing stems: a close-up on cell wall genes. *BMC Plant Biology*, **9**, 6–24.

Mohnen, D., Bar-Peled, M., Somerville, C. (2008) Cell wall polysaccharide synthesis. In: *Biomass Recalcitrance – Deconstructing the Plant Cell Wall for Bioenergy* (ed. D.E. Himmel), pp 94–187. Blackwell Publishing Ltd., Oxford.

Mu, J., Cheng, C., Roach, P.J. (1996) Initiation of glycogen synthesis in yeast – requirement of multiple tyrosine residues for function of the self-glucosylating Glg proteins *in vivo*. *Journal of Biological Chemistry*, **271**, 26554–26560.

Müller, E., Hinckley, A., Rothfield, L. (1972) Studies of phospholipid-requiring bacterial enzymes. III. Purification and properties of uridine diphosphate glucose:lipopolysaccharide glucosyltransferase I. *Journal of Biological Chemistry*, **247**, 2614–2622.

Nishizawa, A., Yabuta, Y., Shigeoka, S. (2008) Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiology*, **147**, 1251–1263.

Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., Shigeoka, S. (2006) Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. *Plant Journal*, **48**, 535–547.

Obendorf, R.L. (1997) Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Science Research*, **7**, 63–74.

Orfila, C., Sørensen, S.O., Harholt, J., Geshi, N., Crombie, H., Truong H-N., Reid, J.S.G., Knox, J.P., Scheller, H.V. (2005) *QUASIMODO1* is expressed in vascular tissue of *Arabidopsis thaliana* inflorescence stems, and affects homogalacturonan and xylan biosynthesis. *Planta*, **222**, 613–622.

Panikulangara, T.J., Eggers-Schumacher, G., Wunderlich, M., Stransky, H., Schöffl, F. (2004) *Galactinol synthase1*. A novel heat shock factor target gene responsible for

heat-induced synthesis of raffinose family oligosaccharides in Arabidopsis. *Plant Physiology*, **136**, 3148–3158.
Paterson, A.H., Bowers, J.E., Bruggmann, R., *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.
Peña, M.J., Zhong, R., Zhou G-K., *et al.* (2007) Arabidopsis *irregular xylem8* and *irregular xylem9*: Implications for the complexity of glucuronoxylan biosynthesis. *Plant Cell*, **19**, 549–563.
Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G., Strynadka, N.C.J. (2001) Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nature Structural Biology*, **8**, 166–175.
Persson, S., Caffall, K.H., Freshour, G., Hilley, M.T., Bauer, S., Poindexter, P., Hahn, M.G., Mohnen, D., Somerville, C. (2007) The Arabidopsis *irregular xylem8* mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. *Plant Cell*, **19**, 237–255.
Persson, S., Wei, H.R., Milne, J., Page, G.P., Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proceedings of the National Academy of Sciences of the U S A*, **102**, 8633–8638.
Peterbauer, T., Lahuta, L.B., Blöchl, A., *et al.* (2001) Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. *Plant Physiology*, **127**, 1764–1772.
Peterbauer, T., Mucha, J., Mach, L., Richter, A. (2002) Chain elongation of raffinose in pea seeds. Isolation, charcterization, and molecular cloning of a multifunctional enzyme catalyzing the synthesis of stachyose and verbascose. *Journal of Biological Chemistry*, **277**, 194–200.
Peterbauer, T., Richter, A. (2001) Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research*, **11**, 185–197.
Pharr, D.M., Sox, H.N., Locy, R.D., Huber, S.C. (1981) Partial characterization of the galactinol forming enzyme from leaves of *Cucumis sativus* L. *Plant Science Letters*, **23**, 25–33.
Qi, Y., Kawano, N., Yamauchi, Y., Ling, J., Li, D., Tanaka, K. (2005) Identification and cloning of a submergence-induced gene OsGGT (glycogenin glucosyltransferase) from rice (*Oryza sativa* L.) by suppression subtractive hybridization. *Planta*, **221**, 437–445.
Raetz, C.R.H., Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*, **71**, 635–700.
Roach, P.J., Skurat, A.V. (1997) Self-glucosylating initiator proteins and their role in glycogen biosynthesis. *Progress in Nucleic Acid Research and Molecular Biology*, **57**, 289–316.
Robbins, N.S., Pharr, D.M. (1987) Regulation of photosynthetic carbon metabolism in cucumber by light intensity and photosynthetic period. *Plant Physiology*, **85**, 592–597.
Rojo, E., Denecke, J. (2008) What is moving in the secretory pathway of plants? *Plant Physiology*, **147**, 1493–1503.
Saravitz, D.M., Pharr, D.M., Carter, T.E., Jr. (1987) Galactinol synthase activity and soluble sugars in developing seeds of four soybean genotypes. *Plant Physiology*, **83**, 185–189.

Schaff, J.E., Nielsen, D.M., Smith, C.P., Scholl, E.H., Bird, D.M. (2007) Comprehensive transcriptome profiling in tomato reveals a role for glycosyltransferase in *Mi*-mediated nematode resistance. *Plant Physiology*, **144**, 1079–1092.

Senser, M., Kandler, O. (1967) Occurence and distribution of galactinol in leaves of higher plants. *Phytochemistry*, **6**, 1533–1540.

Shao, M., Zheng, H., Hu, Y., *et al.* (2004) The *GAOLAOZHUANGREN1* gene encodes a putative glycosyltransferase that is critical for normal development and carbohydrate metabolism. *Plant and Cell Physiology*, **45**, 1453–1460.

Singh, D.G., Lomako, J., Lomako, W.M., *et al.* (1995) β-Glucosylarginine: A new glucose-protein bond in a self-glucosylating protein from sweet corn. *FEBS Letters*, **376**, 61–64.

Smythe, C., Caudwell, F.B., Ferguson, M., Cohen, P. (1988) Isolation and structural analysis of a peptide containing the novel tyrosyl-glucose linkage in glycogenin. *EMBO Journal*, **7**, 2681–2686.

Sterling, J.D., Atmodjo, M.A., Inwood, S.E., *et al.* (2006) Functional identification of an Arabidopsis pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proceedings of the National Academy of Sciences of the U S A*, **103**, 5236–5241.

Sterling, J.D., Quigley, H.F., Orellana, A., Mohnen, D. (2001) The catalytic site of the pectin biosynthetic enzyme α-1,4-galacturonosyltransferase is located in the lumen of the Golgi. *Plant Physiology*, **127**, 360–371.

Taji, T., Ohsumi, C., Iuchi, S., *et al.* (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. *Plant Journal*, **29**, 417–426.

Tsabary, G., Shani, Z., Roiz, L., Levy, I., Riov, J., Shoseyov, O. (2003) Abnormal 'wrinkled' cell walls and retarded development of transgenic *Arabidopsis thaliana* plants expressing endo-1,4-beta-glucanase (cel1) antisense. *Plant Molecular Biology*, **51**, 213–224.

Tuskan, G.A., DiFazio, S., Jansson, S., *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. and Gray). *Science*, **313**, 1596–1604.

Vlieghe, K., Vuylsteke, M., Florquin, K., *et al.* (2003) Microarray analysis of *E2Fa-DPa*-overexpressing plants uncovers a cross-talking genetic network between DNA replication and nitrogen assimilation. *Journal of Cell Science*, **116**, 4249–4259.

Volk, G.M., Haritatos, E.E., Turgeon, R. (2003) Galactinol synthase gene expression in melon. *Journal of the American Society for Horticultural Science*, **128**, 8–15.

Yin, Y., Chen, H., Hahn, M.G., Mohnen, D., Xu, Y. (2010) Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. *Plant Physiology*, **153**, 1729–1746.

Yu, J., Hu, S.N., Wang, J., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**, 79–92.

Zhao T-Y., Thacker, R., Corum, J.W.I.I.I., *et al.* (2004) Expression of the maize *GALACTINOL SYNTHASE* gene family: (I) Expression of two different genes during seed development and germination. *Physiologia Plantarum*, **121**, 634–646.

Zimmermann, M.H. (1957) Translocation of organic substances in trees. I. The nature of the sugars in the sieve tube exudate of trees. *Plant Physiology*, **32**, 288–291.