The plant glycosyltransferase clone collection for functional genomics

Jeemeng Lao[1], Ai Oikawa[1], Jennifer R. Bromley[1], Peter McInerney[1,2], Anongpat Suttangkakul[1], Andreia M. Smith-Moritz[1], Hector Plahar[1], Tsan-Yu Chiu[1], Susana M. González Fernández-Niño[1], Berit Ebert[1], Fan Yang[1], Katy M. Christiansen[1], Sara F. Hansen[1], Solomon Stonebloom[1], Paul D. Adams[1,3], Pamela C. Ronald[1,4], Nathan J. Hillson[1], Masood Z. Hadi[1,2], Miguel E. Vega-Sánchez[1], Dominique Loqué[1], Henrik V. Scheller[1,5] and Joshua L. Heazlewood[1]*

[1]*Joint BioEnergy Institute and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720 USA,*
[2]*Sandia National Laboratory, Livermore, CA 94551, USA,*
[3]*Department of Bioengineering, University of California, Berkeley, CA 94720, USA,*
[4]*Department of Plant Pathology and the Genome Center, University of California, Davis, CA 95616, USA, and*
[5]*Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA*

SUMMARY

The glycosyltransferases (GTs) are an important and functionally diverse family of enzymes involved in glycan and glycoside biosynthesis. Plants have evolved large families of GTs which undertake the array of glycosylation reactions that occur during plant development and growth. Based on the Carbohydrate-Active enZymes (CAZy) database, the genome of the reference plant *Arabidopsis thaliana* codes for over 450 GTs, while the rice genome (*Oryza sativa*) contains over 600 members. Collectively, GTs from these reference plants can be classified into over 40 distinct GT families. Although these enzymes are involved in many important plant specific processes such as cell-wall and secondary metabolite biosynthesis, few have been functionally characterized. We have sought to develop a plant GTs clone resource that will enable functional genomic approaches to be undertaken by the plant research community. In total, 403 (88%) of CAZy defined Arabidopsis GTs have been cloned, while 96 (15%) of the GTs coded by rice have been cloned. The collection resulted in the update of a number of Arabidopsis GT gene models. The clones represent full-length coding sequences without termination codons and are Gateway® compatible. To demonstrate the utility of this JBEI GT Collection, a set of efficient particle bombardment plasmids (pBullet) was also constructed with markers for the endomembrane. The utility of the pBullet collection was demonstrated by localizing all members of the Arabidopsis GT14 family to the Golgi apparatus or the endoplasmic reticulum (ER). Updates to these resources are available at the JBEI GT Collection website http://www.addgene.org/.

Keywords: glycosyltransferase, Arabidopsis, rice, particle bombardment, GT14, cell wall, endomembrane, subcellular localization, biolistics.

INTRODUCTION

The glycosyltransferases (GTs; EC 2.4.x.y) are enzymes that catalyze the transfer of activated carbohydrate moieties from donor molecules to an acceptor molecule such as saccharides, nucleic acids, lipids, proteins, and an assortment of organic compounds. Glycosylation is a pivotal reaction for all organisms including eukaryotes, archaea, bacteria, and viruses. The Carbohydrate-Active enZymes (CAZy) database has defined over 90 distinct GT families from known biological systems (Lombard *et al.*, 2014). Higher plants have evolved large families of GTs that produce an array of glycoconjugates. Collectively these compounds are involved in a range of structural and functional roles throughout the plant cell. More recently, several studies have indicated that plants likely encode GT families not currently classified by CAZy (Zhou *et al.*, 2009; Hansen *et al.*, 2012; Nikolovski *et al.*, 2012).

Major roles for GTs in plants include the biosynthesis of polysaccharides in the plant cell wall. The most abundant polysaccharide in the cell wall is cellulose, a linear polymer synthesized from β-1,4-linked glucose molecules by

cellulose synthase GT2 family members (CESA) at the plasma membrane (Somerville, 2006). Matrix polysaccharides are branched molecules, largely synthesized in the Golgi apparatus by a myriad of GT family members and secreted to the cell wall to form a flexible cross-linking matrix (Scheller *et al.*, 2007; Scheller and Ulvskov, 2010). Protein glycosylation occurs in the endoplasmic reticulum (ER) and Golgi apparatus and is one of the most common post-translational modifications in eukaryotic systems (Stanley, 2011). The process involves the construction of branched polysaccharide chains by a range of GT family members and associated partners in a well regulated sequential process through the plant secretory system (Oikawa *et al.*, 2013). These polysaccharide chains are comprised of a variety of glycosidic linkages and are attached to the protein backbone via either *O-* or *N*-linked glycosidic bonds (Kang *et al.*, 2008). Glycolipids are a major constituent of biological membranes and are generally classified as compounds comprised of lipids covalently bound to a monosaccharide or polysaccharide chain. In plants, glycolipid biosynthesis has been poorly characterized with few GTs functionally associated with these biosynthetic pathways. The involvement of the GT4 family members (DGDG synthases) in plastid galactolipid biosynthesis represent the best characterized example (Kelly and Dörmann, 2004). Plants synthesize a multitude of specialized or secondary metabolites, many of which exist in various glycoforms which can affect their physical and chemical properties (Vaistij *et al.*, 2009). Glycosylation reactions involving small molecule acceptors are generally undertaken by GT1 family members, which comprise the largest clade in the family of plant GTs (Caputi *et al.*, 2012). However, other GT families are also involved in glycosylations of small molecule acceptors, especially retaining glycosylation reactions as found, for example in sucrose phosphate synthase (GT4), galactinol synthase (GT8), and trehalose synthase (GT20).

The subcellular localization of proteins can assist in defining a functional context within the eukaryotic cell. Plant organelle proteomics and fluorescent tagged proteins (FP) have been extensively applied to determine subcellular localizations, especially in the reference plant *Arabidopsis thaliana.* Only around 10% of the GTs encoded by Arabidopsis have been localized using fluorescently tagged proteins and about 25% by mass spectrometry (Tanz *et al.*, 2013). There are a host of plasmid collections and marker sets commonly used for localization studies in plants (Nelson *et al.*, 2007; Geldner *et al.*, 2009). However, these vectors are generally not very effective for transient co-localizations experiments. These large binary vectors have genes necessary for *Agrobacterium*-based transformations, which serve no real purpose for transient expression using particle bombardment. It has been shown that the vector backbone can have negative effects on the transgene, thus having unnecessary genes may be detrimental

when performing particle bombardment (Fu *et al.*, 2000). In addition, these plasmids often come with the low copy RK2 origin, oriV (Kues and Stahl, 1989), resulting in poor DNA yields from plasmid preparations and reduced transient transformation efficiencies. When performing co-localization studies, two plasmids must be used since the marker and gene of interest are located on separate plasmids. This is problematic, as co-expression of two genes on separate plasmids occurs at a much lower rate (Tang *et al.*, 1999) and to achieve higher rates requires plasmid ratio optimization (Chen *et al.*, 1998).

Large-scale functional genomic approaches provide a mechanism to determine information relating to protein function e.g. protein–protein interactions (Arabidopsis Interactome Mapping Consortium, 2011), heterologous complementation (Ton and Rao, 2004), functional screens (Eswaran *et al.*, 2010), *in vitro* expression for enzyme assays (Greving *et al.*, 2012), overexpression in plants (Lindbo, 2007) and subcellular localization (Nelson *et al.*, 2007). In the reference plants of Arabidopsis and *Oryza sativa* (rice), there are currently around 450 and over 600 GTs in each system respectively (Lombard *et al.*, 2014). Most of these GTs do not have a clear functional designation and few GTs have been successfully characterized. Unfortunately, unlike model organisms such as yeast and mouse, complete cDNA and open reading frame (ORF) collections are not readily available in plants (Gelperin *et al.*, 2005; MGC (Mammalian Gene Collection) Project Team 2009). Some international projects have sought to create large-scale ORF collections in Arabidopsis and rice including the SSP/RIKEN Consortium (Yamada *et al.*, 2003), Arabidopsis Interactome Mapping Consortium (Arabidopsis Interactome Mapping Consortium, 2011) and The Rice Full-Length cDNA Consortium (The Rice Full-Length cDNA Consortium, 2003). These clones are available at various stock centers and encompass over 250 Arabidopsis GT clones (60% of the encoded GTs) and over 600 rice GT clones (95% of the encoded GTs). However, these clones are derived from a variety of collections and often found in different vector backbones; most contain termination codons, are partial sequences, are not fully validated (sequenced) or contain untranslated regions making them unsuitable for many applications without further manipulation.

In an effort to address these limitations and to accelerate research in the area of plant GTs, especially as they pertain to cell-wall biosynthesis, we have cloned and verified over 500 GTs from Arabidopsis and rice. These plant GT coding sequences (CDS) have been cloned in-frame into pDONR™ vectors using Gateway® technology to readily enable downstream applications. Furthermore, the entire collection will be made available to the community to drive functional genomic approaches related to GT function. To highlight the application of this resource, known as the JBEI GT Collection, we have designed and constructed

particle bombardment plasmids (pBullet) with co-localization markers for the plant endomembrane. We have tested this collection of bombardment plasmids to highlight their effectiveness by localizing the 11 members of the GT14 family from Arabidopsis.
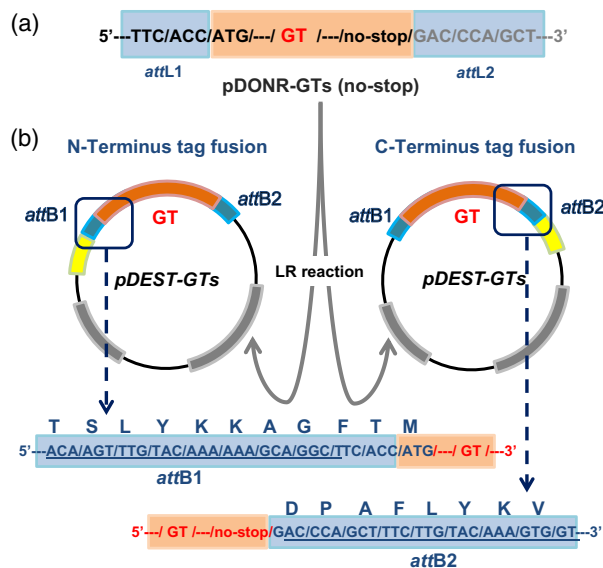
RESULTS AND DISCUSSION

Construction of the Arabidopsis and rice GT collections

Annotated GT genes for both Arabidopsis and rice were obtained from the CAZy database (Lombard *et al.*, 2014). These data were integrated with the most recent coding DNA Sequence (CDS) models for Arabidopsis (Lamesch *et al.*, 2012) and rice (Kawahara *et al.*, 2013; Sakai *et al.*, 2013), the latter models were also integrated with data in the Rice GT Database (Cao *et al.*, 2008). Mixed organ and developmental cDNA libraries from Arabidopsis and rice were employed with two rounds of PCR (Figure S1). PCR products from the second round were gel purified and cloned into pDONR™ vectors. Both forward and reverse gene specific primers were designed to ensure all clones were in frame with the *att*L region of the pDONR™ vectors and the reverse primers excluded the stop codon from all cDNAs (Figure 1a). Employing Gateway® recombination cloning technology, the pDONR-GT clones can be transferred to destination vectors to produce the expression

Figure 1. Overview and application of the JBEI GT Collection.
(a) A schematic plasmid map of the pDONR-GT clones in the JBEI GT Collection highlighting in-frame sequences of both the 5′ and 3′ ends of the target sequence.
(b) Workflow outlining the utilization of the Gateway® technology with the JBEI GT collections. The resultant coding sequence (*att*B *sites)* of an expression vector (pDEST-GTs) after the LR reaction of the pDONR-GT entry clone with a destination vector containing either N- or C-terminus tag is also shown. The Universal antibody recognizes the *att*B2 sequence DPAFLYKV at the C-terminus (Eudes *et al.*, 2011).

clone. The provision of in-frame coding regions at both the 5′ and 3′ ends enables this collection of clones (JBEI GT Collection) to be used in applications employing either N- or C-terminus fusions with various pDEST vectors (Table 1). An added benefit of enabling translation of the *att*B2 region of a pDEST vector (encoding the amino acid sequence DPAFLYKVV) is that the translated clone can be detected by immunoblotting with the Universal (UNI) antibody (Eudes *et al.*, 2011) after any downstream application (Figure 1b). A variety of expressed proteins has been successfully detected in various heterologous systems using this epitope (Eudes *et al.*, 2011, 2012; Chiu *et al.*, 2012).

The Arabidopsis JBEI GT collection

The total number of Arabidopsis GTs curated by CAZy with corresponding gene models in the most recent genome release from The Arabidopsis Information Resource (TAIR10) is 456 (Lamesch *et al.*, 2012). In addition, several proteins have been tentatively assigned as GT-like after their identification in plant Golgi proteomes (Nikolovski *et al.*, 2012) and through bioinformatic analyses (Hansen *et al.*, 2012). Many of these have now been incorporated into CAZy and designated not classified (GTnc). The current number of loci designated as GTnc in Arabidopsis is 98. Although we have made some efforts to clone these GTnc sequences from Arabidopsis, our primary focus has been the core set of glycosyltransferases as defined by CAZy. Out of the 456 CAZy designated GTs in Arabidopsis, 403 (88%) were successfully cloned and are available in this collection (Table S1). From the GTnc annotated genes, 25 (25.5%) are available, with the majority comprising members of the GT14-like (GT14L) gene families (Table S1). In total, 18 clones were obtained that did not correspond to the representative gene models as defined by TAIR (The Arabidopsis Information Resource); these splice forms may actually represent the dominant gene model in Arabidopsis. All 42 GT gene families are represented in the Arabidopsis JBEI GT Collection with the majority of family members cloned (Figure 2). Under-represented families include callose synthase (GT48) of which only two of 13 members (15%) are included in this collection. All members of the GT48 family are encoded by large cDNAs (approximately 6 kb), which likely contributed to difficulties in amplification. The high quality of clones available in the Arabidopsis JBEI GT Collection is highlighted by the small number of cloning errors resulting in sequence variations or single nucleotide polymorphisms (SNPs). There are 26 SNPs in 25 clones which do not result in a change to the amino acid sequence. In total, 10 clones contain variations that result in amino acid changes (non-synonymous SNPs), with only one of these clones containing multiple changes, namely AT3G46720.1 (Table S1). An analysis of nsSNPs in sequenced Arabidopsis accessions at the 1001 Proteomes portal (Joshi *et al.*, 2012) confirmed that all these amino

Table 1 Gateway® destination vectors for functional genomics that are compatible with the pDONR-GT clones available from the JBEI GT collections

Vector category	Application	Method and/or epitope tags	Reference
Binary vectors (e.g. p7GW2)	Functional analysis	Overexpression in Arabidopsis/tobacco; N- and C-terminus fusions to GFP	Karimi *et al.* (2002)
Binary vectors (pMDC)	Functional analysis	Overexpression in Arabidopsis/tobacco; N- and C-terminus fusions to GFP	Curtis and Grossniklaus (2003)
Transient expression vectors (pSAT)	Functional analysis	Transient expression into plant cells; N- and C-terminus fusions to GFP, YFP, CFP, and DsRed	Tzfira *et al.* (2005)
Binary vectors (pEarley)	Functional analysis	Overexpression in Arabidopsis/tobacco; N- and C-terminus fusions to YFP, CFP, GFP, FLAG, HA, cMyc, AcV5, and TAP tags	Earley *et al.* (2006)
Binary vectors (pGWB)	Functional analysis	Overexpression in Arabidopsis/tobacco; GFP, LUC, YFP, CFP, 6xHis, FLAG, 3xHA, 4xMyc, 10xMyc, GST, T7-epitope, and TAP tags	Nakagawa *et al.* (2007)
Binary virus vectors (pTMV)	Functional analysis	Protein production with large amount in tobacco; N- and C-terminus fusions to YFP, CFP, FLAG, HA, cMyc, His, and Strep-tags	Kagale *et al.* (2012)
Binary vectors (mcBiFC)	Protein–protein interactions	Multicolor bimolecular fluorescence complementation (mcBiFC) using tobacco expression system	Gehl *et al.* (2009)
Binary vectors (FLuCl)	Protein–protein interactions	Luciferase complementation imaging (LuCl) using tobacco expression system	Gehl *et al.* (2011)
Yeast vectors (mbSUS)	Protein–protein interactions	Mating-based split-ubiquitin system (mbSUS) in yeast expression system	Grefen *et al.* (2009)

acid changes were likely due to cloning errors and not inaccuracies in the reference genome (Col-0). Ultimately, our intention is to create a complete collection of Arabidopsis GTs and intend to correct errors and acquire remaining clones in the future.

Updated Arabidopsis gene models in the collection

The Arabidopsis genome has been intensely curated and on its 10th revision (Lamesch *et al.*, 2012). We identified eight CAZy designated GTs with varied intron–exon junctions after their sequences were compared with the current CDS at TAIR (Figure 3). These updated GT models are the result of corrected intron–exon splicing. An examination of transcript coverage can explain many of these incorrectly or varied splice boundaries. Many of these intron–exon sites are annotated computationally and have poor transcript coverage. As a result, we have identified three new gene models for the loci AT4G15270, AT1G77810 and AT2G03370 and updated models for five loci (AT2G25540, AT5G37180, AT2G46480, AT3G07620 and AT5G37000). For example, the cellulose synthase 10 gene AT2G25540.1 (GT2), while from an extensively studied gene family (Somerville, 2006), has poor transcript coverage at the 5′ end of the cDNA. There is no coverage of the incorrectly annotated fourth exon and it appears to have a misassigned splice donor site (GT) in the following intron. The updated gene model removes 21 bp (7 amino acids) from

the end of the fourth exon and now better aligns to the sequence of its closest paralog AT4G32410.1 (CESA1).

The rice JBEI GT collection

The initial stage of the rice GT collection was the development of the Rice GT Database (Cao *et al.*, 2008). After its initial construction, the Rice GT Database contained 605 loci annotated as rice GTs. Since the resource was created, some new families have been designated by CAZy (e.g. GT90 and GT92) and an updated version of the rice genome has been made available by MSU Rice Genome Annotation Project (release 7) as well as updates to the Rice Annotation Project (Sakai *et al.*, 2013). Thus we have now integrated this updated information with the Rice GT Database to create 622 loci (Table S2). Phase one of the rice collection was recently completed, resulting in 96 gene models (15% of the total) from a variety of GT family members (Figure S2). The initial focus was on GT families that contain clades that are expanded or divergent in grasses compared to dicots, e.g. GT2, GT47 and GT61 (Cao *et al.*, 2008). Well represented family members include those from GT61 (17 clones), GT47 (15 clones), GT8 (11 clones) and GT2 (10 clones). Current corrections to gene models in the rice collection include a GT47, LOC_Os06 g23420.2 which has been updated to include an extra codon which inserts an alanine (A488) at an intron/exon splice site and the GT48, LOC_Os03 g02756.1 which was discontinued in
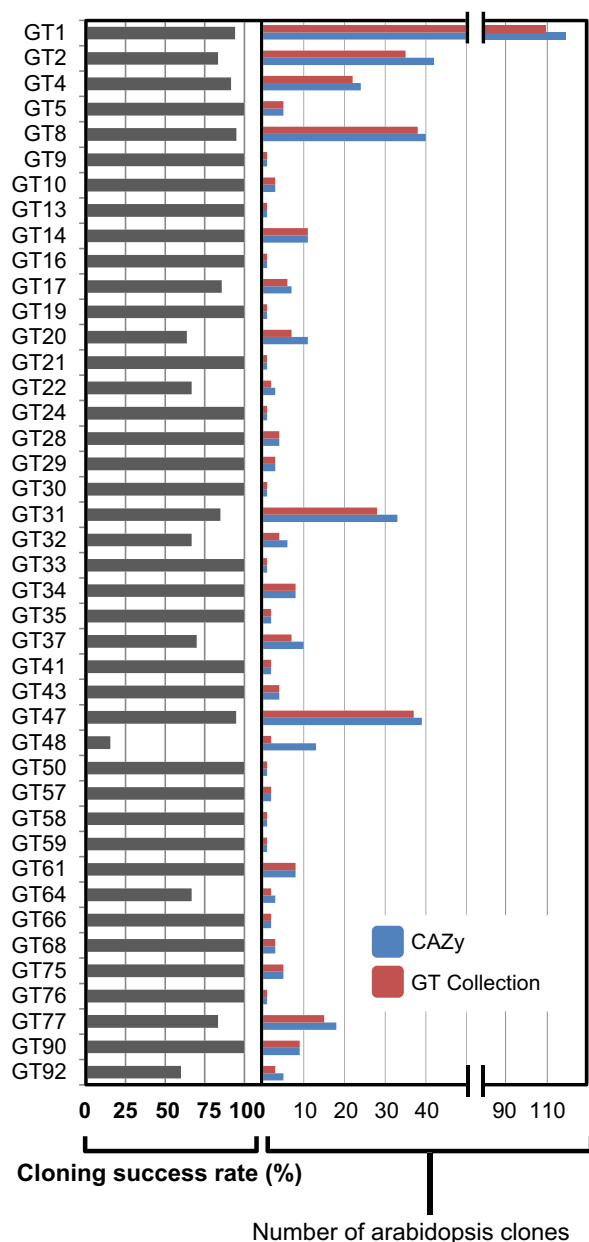
Figure 2. Cloning status of the Arabidopsis JBEI GT Collection.
The grey bars indicate the percentage of successfully cloned GTs in each family. The blue bars indicates the total numbers of genes in each GT family according to Table S1. The red bars indicate the number of clones currently available in the collection for each GT family.

MSU Rice Genome Annotation Project (release 7), but is still present in Rice Annotation Project as Os03 g0119500 and is clearly a valid rice locus. Many small GT families are also not well represented in this collection, but a more targeted cloning approach is now underway.

### Subcellular localization of Arabidopsis GTs

The subcellular localization of proteins can assist functional classifications and potential interaction partners. Of the over 550 Arabidopsis GTs (including the GTnc genes), only 54 have previously been localized using tagged fluorescent proteins according to the SUBcellular Arabidopsis database (SUBA) (Tanz *et al.*, 2013). The majority of these proteins has been localized to the Golgi apparatus, followed by the cytosol, the ER and plasma membrane (Table S1). In total, 127 Arabidopsis GTs have been identified and localized by subcellular proteomic surveys (Tanz *et al.*, 2013). Roughly half were localized to the Golgi apparatus, a fifth to the plasma membrane, and a sixth to the cytosol. However, many of the GTs from subcellular proteomic studies display contradictory subcellular locations. The utilization of multiple techniques to confirm subcellular localization of proteins is important to increase confidence for a given assignment due to the high levels of false positives (Millar *et al.*, 2009). In order to accurately determine the subcellular localization of proteins, multiple studies and a variety of methods should be examined.

The pBullet biolistic plasmids

An effective means to localize plant proteins is using particle bombardment, as it allows rapid transient expression of genes in plants (Rech *et al.*, 2008). The current collection of vectors used for biolistics are dual purpose and are also used for stable transformation of plant tissue using *Agrobacterium* (Geldner *et al.*, 2009). The pBullet collection of plasmids was constructed to address issues associated with current options for biolistic-based subcellular localizations approaches in plants. The plasmid backbone was synthesized in two sections and assembled using ligation-independent cloning. Complete synthesis enabled a number of features to be incorporated into the plasmid including a collection of unique restriction sites, a modular design, minimal junk DNA, a high-copy origin of replication from pUC (Chambers *et al.*, 1988), optimized fluorescent proteins EYFP and ECFP (Shaner *et al.*, 2005), Gateway® recombination site for fast direction cloning (Hartley *et al.*, 2000), inclusion of organelle markers and multiple selectable markers (Figure 4a). The pBullet vectors were specifically designed to overcome the yield limitations of the commonly employed binary vectors (Nelson *et al.*, 2007). The incorporation of the high-copy origin of replication into pBullet provides adequate yields of plasmid from a miniprep, usually 10–20 μg of plasmid DNA (Table S3). This enables multiple bombardments to be undertaken from a single preparation with high rates of transformation using about 0.5 μg of plasmid DNA (Figure S3).

To ensure co-transformation and thus simultaneous expression of marker and the protein of interest in the same cell after biolistics, subcellular markers (ECFP) for a range of compartments were introduced into the pBullet backbone (Table 2). The cytosol marker employs the ECFP with no signal sequence; ER localization employs the
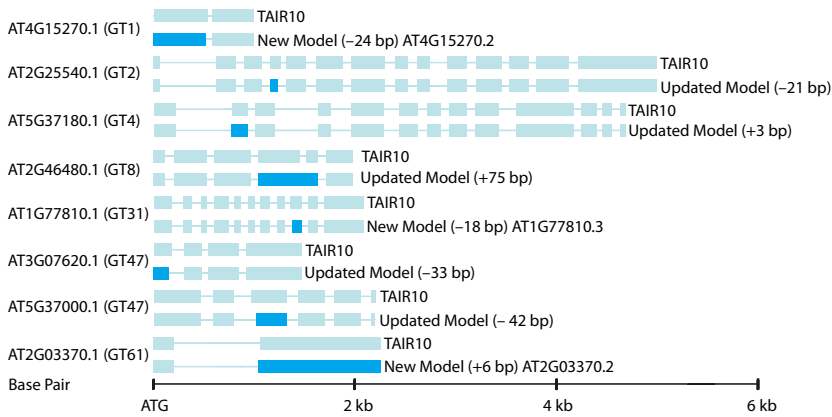
Figure 3. Diagram of Arabidopsis glycosyltransferases with modified gene models. Blocks and lines are used to indicate exons and introns for each locus. The light blue corresponds to exons according to TAIR10 while the dark blue indicates an update in the clone in the JBEI GT Collection. The change or update to the revised gene model is shown below the current model designated by TAIR10. The variation in nucleotides in the revised gene models is shown on the right.
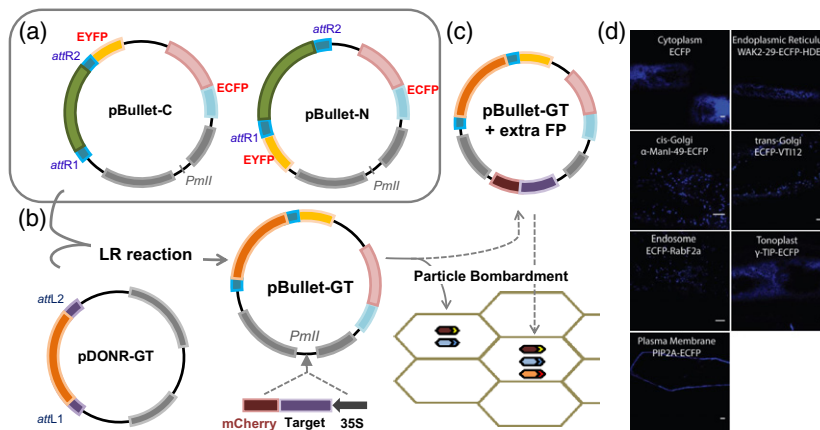
Figure 4. Overview demonstrating utilization of the pBullet plasmid collection.
(a) Schematic diagrams of the both the C- and N-terminal pBullet plasmids indicating major components: GT (orange); *attL* (purple); Ccdb and chloramphenicol (blue); ECFP(light blue); *attR* (green); EYFP (yellow); organelle marker (pink or purple); mCherry (red); various promoter and antibiotic resistance (grey).
(b) Diagram outlining the production of the pBullet expression clone after the LR reaction with a pDONR vector.
(c) Diagram highlighting the utility of the pBullet backbone demonstrating the use of the PmlI restriction site to add a second subcellular marker e.g. mCherry.
(d) Confirmation of pBullet co-localization marker with ECFP in onion epidermal cells. Scale bar = 10 μm.

Table 2 Summary of the pBullet plasmid collection indicating nomenclature and subcellular markers employed for co-localization

Compartment	Plasmid name	Protein marker (ECFP)	GenBank accession (c/n)	Addgene plasmid number
Cytosol	pBullet-cyt-c/n	ECFP	KJ081785/KJ081787	53066/53065
Endoplasmic reticulum	pBullet-er-c/n	WAK2-29-ECFP-HDEL	KJ081780/KJ081783	53068/53067
cis-Golgi	pBullet-cg-c/n	α-MANI-49-ECFP	KJ081792/KJ081781	53070/53069
trans-Golgi network	pBullet-tgn-c/n	ECFP-VTI12	KJ081789/KJ081791	53072/53071
Endosome	pBullet-end-c/n	ECFP-RABF2A	KJ081790/KJ081788	53074/53073
Tonoplast	pBullet-vac-c/n	γ-TIP-ECFP	KJ081784/KJ081786	53076/53075
Plasma membrane	pBullet-pm-c/n	PIP2A-ECFP	KJ081782/KJ081779	53078/53077

WAK2 (AT1G21270) signal peptide (29 amino acids) on the N-terminus and the HDEL retention sequence on the C-terminus of ECFP; *cis*-Golgi localization of ECFP employs an N-terminus fusion with the first 49 amino acids from α-ManI protein from soybean (Nelson *et al.*, 2007); the *trans*-Golgi localization of ECFP is achieved using VTI12 (AT1G26670, V-SNARE 12) fused to the C-terminus (Geldner *et al.*, 2009); the endosomal marker RABF2A (AT5G45130) was fused to the C-terminus of ECFP (Geldner *et al.*, 2009); tonoplast localization was achieved using γ-TIP (AT2G36830) fused to the N-terminus of ECFP (Nelson *et al.*, 2007) and PIP2A (AT3G53420) was fused to the N-terminus of ECFP to produce a plasma membrane marker (Nelson *et al.*, 2007). The localization of each subcellular

marker was assessed all successfully localized as reported in the literature (Figure 4d).

To increase the utility of the pBullet plasmid for high-confidence localizations of a gene of interest, the ability to carry out N- or C-terminus fusions with EYFP is also possible (Figure 4a). To ensure successful translation of the fusion construct in a pDONR™ vector with either the N- or C-terminus of EYFP in the pBullet vectors, *att*B PCR primers require additional bases to ensure in-frame translations (Figure 1). All pBullet plasmids contain *att*R sites for Gateway® cloning and the inclusion of an organelle marker means only one plasmid is required. Specifically engineered restriction sites enable extensive modification of the core plasmids. This could include the addition of a second marker (e.g. mCherry) at the *Pml*I site found between the two resistance markers (Figure 4b,c).

Application of pBullet with the Arabidopsis GT14 family

In order to demonstrate the utility of the pBullet vectors we examined the GT14 family of the GT superfamily from Arabidopsis (Lombard *et al.*, 2014). There is currently little functional knowledge regarding this family in plants, although members have been implicated in arabinogalactan biosynthesis (Zhou *et al.*, 2009). This was recently confirmed for three members of the GT14 family (At5 g39990, At2 g37585, and At5 g15050) where the catalytic domain was expressed and shown to transfer glucuronic acid to β-1,6-galactan and β-1,3-galactan, both of which are present in type II arabinogalactans (Knoch *et al.*, 2013; Dilokpimol and Geshi, 2014). There are 11 annotated members of the GT14 family in Arabidopsis as determined by CAZy (Coutinho *et al.*, 2003). Our previous analysis of plant GT14 and GT14-like families indicated a distinct separation between these two clades (Hansen *et al.*, 2012). An examination of the Arabidopsis GT14 and GT14-like members confirmed that there are 11 GT14 family members and that the GT14-like clades were distinct (Figure S4).

The 11 Arabidopsis GT14s were initially cloned into the pBullet-cg-c vector (*cis*-Golgi marker and C-terminus EYFP) and the pBullet-cg-n vector (*cis*-Golgi marker and N-terminus EYFP) to assess their subcellular distributions. The *cis*-Golgi marker was initially selected as three members (AT5 G39990.1, AT3G24040.1 and AT1G71070.1) were recently identified in the Arabidopsis Golgi proteome (Parsons *et al.*, 2012). Using the C-terminus fusion, eight members (AT5G39990.1, AT3G24040.1, AT4G03340.1, AT1G03520.1, AT2G37585.1, AT3G03690.1, AT3G15350.1, and AT5G150 50.1) displayed punctate structures and co-localized to the *cis*-Golgi (Figure 5a,e,g,i,k,m,o,s). Two members (AT4G27480.1 and AT1G53100.1) had a more diffuse structure, and one (AT1G71070.1) had no signal. With the N-terminus fusion, three (AT5G39990.1, AT3G24040.1, AT2G37585.1) were co-localized to the *cis*-Golgi, two (AT1G71070.1 and AT4G03340.1) were diffuse and

network-like, and six gave no signal. pBullet-er was used to test the co-localization of the GT14 members showing diffuse and non-punctate signals. The four genes were cloned into corresponding pBullet-er-c and pBullet-er-n plasmids and introduced into onion epidermal cells. AT4G27480.1 and AT1G53100.1 co-localized with the ER marker using C-terminus fused EYFP while AT1G71070.1 and AT4G03340.1 co-localized with the ER marker using N-terminus fused EYFP (Figure 5w,x,y,z).

Correlation analysis was used to further assess co-localization of markers and GT14 fusion constructs. Values above 0.5 indicate confidence in the co-localization assignment (Manders *et al.*, 1992; Bolte and Cordelieres, 2006). All Golgi-localized GT14s have a Pearson's correlation coefficient above 0.6 for either the N- or C-terminus construct, indicating co-localization with the *cis*-Golgi marker (Figure 5). The Pearson's correlation coefficient values for three of the GT14s with ER-like localization patterns (AT1G71070.1, AT1G53100.1 and AT4G03340.1) were above 0.5, but was inconclusive for AT4G27480.1 (Figure 5). This is probably due to low signal intensity from the FP construct and the noise associated with these images. However, the observed structure appears to be ER upon manual inspection (Figure 5q,w) and will be designated 'likely ER' (Table 3). It should be noted that the images outlined in Figure 5 are representative for at least three independent transformations. In summary, seven GT14 proteins localized to the Golgi, while three localized to the ER and one co-localized to both the ER and the Golgi (Table 3).

The localization results for AT1G71070.1, AT1G53100.1 and AT4G03340 conflict with several subcellular proteomic studies. In each of these cases, the proteins were only found in one other subcellular preparation. However, over 25 plasma membrane proteomes have been characterized (Tanz *et al.*, 2013), with AT1G53100.1 and AT4G03340.1 having only been identified in one preparation each (Keinath *et al.*, 2010; Li *et al.*, 2012). This indicates that they were likely contaminants in these plasma membrane preparations. The GT14 AT1G71070.1 was previously identified in the Golgi proteome (Parsons *et al.*, 2012), which is frequently contaminated with ER membranes due to their close physical proximity (Hawes, 2012). Therefore, an FP localization to the ER with proteomic evidence suggesting an association with the secretory membrane (Golgi apparatus) should more likely be attributed to the ER (Table 3).

This survey highlights the utility of the pBullet collection as it was possible to localize all 11 members of the Arabidopsis GT14 family using either the N or C-terminus FP fusions to the *cis*-Golgi and/or the ER. The entire co-localization experiment was conducted in less than 2 weeks, demonstrating the pBullet plasmid an effective tool for subcellular localizations using particle bombardment.
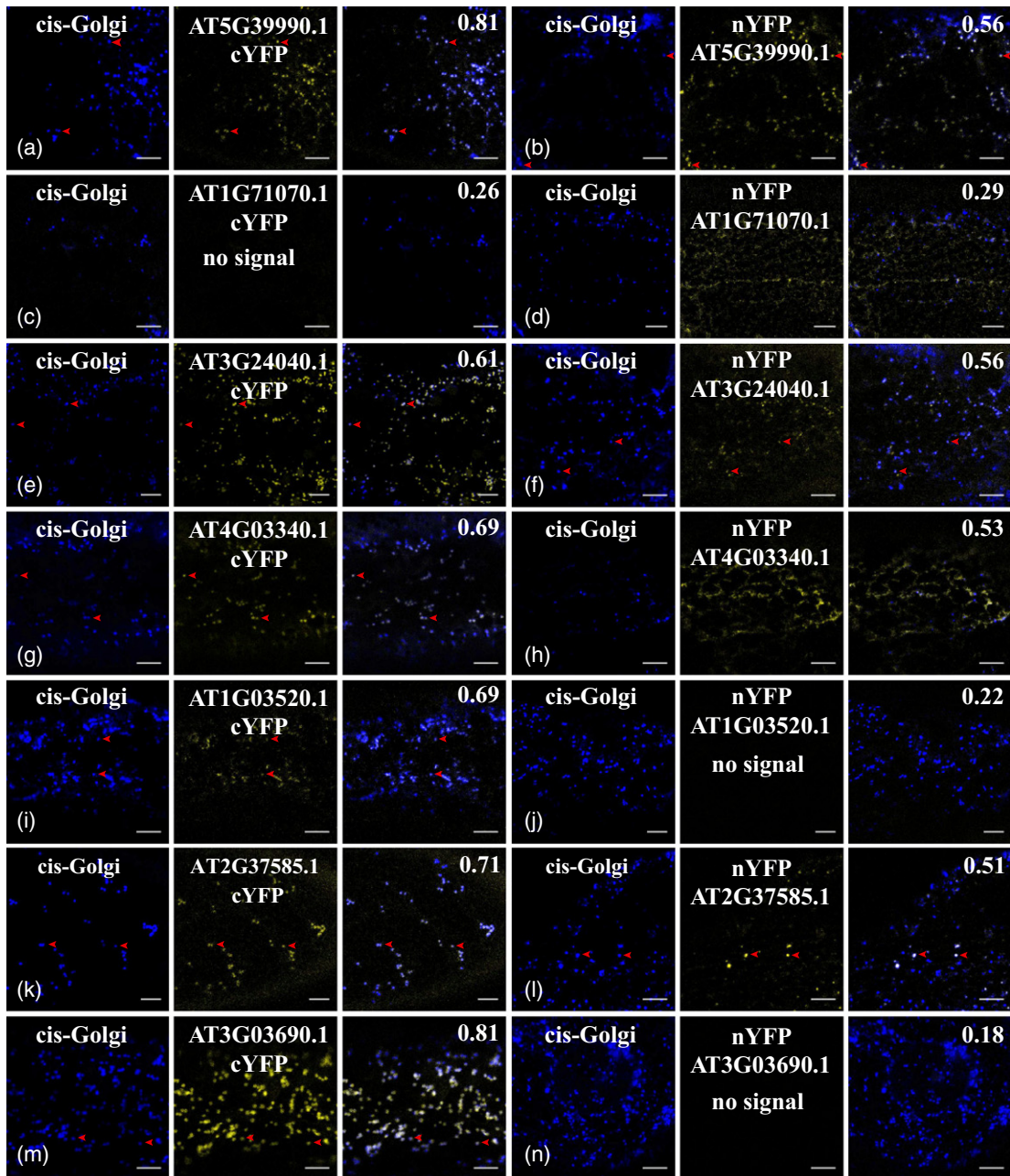
Figure 5. Subcellular localization of the Arabidopsis GT14 family.
Subcellular localization of the Arabidopsis GT14 proteins using N- and C-terminal fluorescent proteins with the pBullet vectors. For each set of three panels, the first contains the organelle marker, the second contains the protein of interest (GT14), and the third shows the overlay image with the Pearson's correlation coefficient on the top right corner. Arrows highlight examples of co-localized punctate structures. Subcellular localization of GT14 proteins using the pBullet-cg-c plasmid (cYFP) containing the *cis*-Golgi marker: (a) AT5G39990.1, (c) AT1G71070.1, (e) AT3G24040.1, (g) AT4G03340.1, (i) AT1G03520.1, (k) AT2G37585.1, (m) AT3G03690.1, (o) AT3G15350.1, (q) AT4G27480.1, (s) AT5G15050.1, (u) AT1G53100.1. Subcellular localization of the Arabidopsis GT14 proteins using the pBullet-cg-n (nYFP) containing the *cis*-Golgi marker: (b) AT5G39990.1, (d) AT1G71070.1, (f) AT3G24040.1, (h) AT4G03340.1, (j) AT1G03520.1, (l) AT2G37585.1, (n) AT3G03690.1, (p) AT3G15350.1, (r) AT4G27480.1, (t) AT5G15050.1, (v) AT1G53100.1. Clarification of non-punctate signal observed from four of the GT14s using the pBullet-er-c and pBullet-er-n, which contain the ER marker: (w) AT4G27480.1, (x) AT1G71070.1, (y) AT1G53100.1 and (z) AT4G03340.1. Results are summarized in Table 3. Scale bar = 10 μm.

Collection availability

The Arabidopsis JBEI GT collection will be made available through Arabidopsis stock centers (e.g. Arabidopsis Biological Resource Center) using the stock numbers outlined in Table S1. The rice collection is available directly upon request, but is currently still being
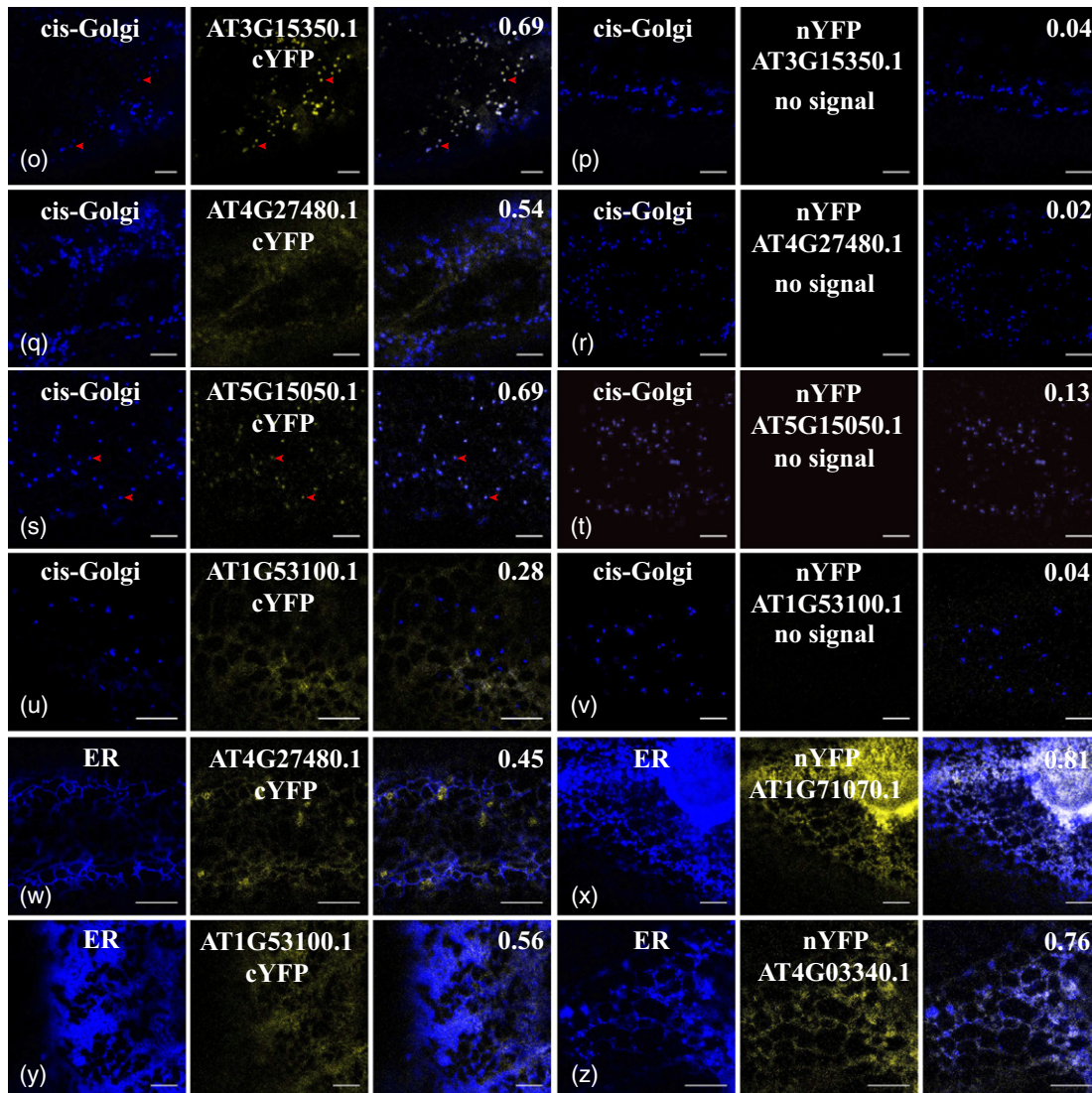
Figure 5. continued

Table 3 Summary of subcellular locations of the Arabidopsis GT14 family using the pBullet plasmid collection

AGI	C-terminus EYFP	N-terminus EYFP	SUBA (MS)	SUBA (FP)	Consensus location
AT5G39990.1	Golgi	Golgi	Golgi	Golgi	Golgi
AT1G71070.1	No signal	ER	Golgi	–	ER
AT3G24040.1	Golgi	Golgi	Golgi	–	Golgi
AT4G03340.1	Golgi	ER	PM	–	ER/Golgi
AT1G03520.1	Golgi	No signal	–	–	Golgi
AT2G37585.1	Golgi	Golgi	–	–	Golgi
AT3G03690.1	Golgi	No signal	–	–	Golgi
AT3G15350.1	Golgi	No signal	–	–	Golgi
AT4G27480.1	ER	No signal	–	–	Likely ER
AT5G15050.1	Golgi	No signal	–	–	Golgi
AT1G53100.1	ER	No signal	PM	–	ER

Subcellular interpretations from both C- and N-terminus EYFP constructs have been made based on co-localizing markers. Prior localization information was obtained from the SUBcellular Arabidopsis database (SUBA) with data for subcellular proteomics (MS) and tagged fluorescent proteins. ER, endoplasmic reticulum; PM, plasma membrane.

assembled. The pBullet collection of plasmids is available from Addgene (https://www.addgene.org/). Further clones for both rice and Arabidopsis are currently in production. The JBEI GT Collection website provides additional details and current progress with regard to the GT collections (http://gt.jbei.org/). The location of each clone in plate format, listed with its GT family number and locus code is provided. Full-length sequence reads were performed on each clone and are available along with vector maps. This information is hosted by a public instance of the Inventory for Composable Elements (ICE) a web-based application for biological parts (Ham *et al.*, 2012).

CONCLUSION

In total, 403 (88%) Arabidopsis GTs and 96 (15%) rice GT genes have been successfully cloned into pDONR™ vectors for improved utility. This collection enables a number of downstream applications when used in conjunction with destination vectors via Gateway® technology. In conjunction with the GT collection, the pBullet series of destination vectors have been developed to assist with efficient subcellular localization experiments in plants with an emphasis on the endomembrane. Collectively, these technologies can be used to undertake a variety of functional genomic applications in plants.

EXPERIMENTAL PROCEDURES

Plant material

Arabidopsis thaliana (L.) Heynh. accession Columbia (Col-0) were grown with 16-h photoperiod at 22°C with 90 μmol m^{-2} sec^{-1} illumination during the day period. *Oryza sativa* L. (cultivar Nipponbare) was grown in chambers under the following conditions: 12 h daylight, 470 μmol m^{-2} sec^{-1} illumination, 80% relative humidity, 26°C for 1 h at the beginning and end of the cycle, and 28°C for the remaining 10 h; 12 h dark, 80% relative humidity, 26°C. For each species, plant tissue from various organs and developmental stages (Arabidopsis: flowers, seedling, stem, leaf, silique or rice: leaves, tillers, stems, roots, seedling) were harvested, frozen in liquid nitrogen and RNA isolated using the plant RNeasy kit (Qiagen) for Arabidopsis material and TRIzol® reagent (Life Technologies, https://www.lifetechnologies.com/) for rice material. First strand synthesis was undertaken using M-MLV Reverse Transcriptase (Sigma-Aldrich, https://www.sigmaaldrich.com/).

Arabidopsis and rice GT gene models

The latest coding sequence (CDS) models for both Arabidopsis (TAIR10) and rice (OSA7) were obtained from The Arabidopsis Information Resource (Swarbreck *et al.*, 2008) and the MSU Rice Genome Annotation Project Database (Kawahara *et al.*, 2013), respectively. Further rice GT models were supplemented from the Rice Annotation Project (Sakai *et al.*, 2013). Annotated GTs for Arabidopsis and rice were obtained from the Carbohydrate-Active enZYmes Database (CAZy) (Lombard *et al.*, 2014), with further annotations for rice GTs acquired from the Rice GT Database (Cao *et al.*, 2008).

Polymerase chain reaction

Gene specific primers (forward: 5′-CAGGCTTCACC-gene specific region-3′ and reverse: 5′-AAAGCTGGGTC- gene specific region-3′) were designed to amplify from mixed cDNA libraries using Phusion High Fidelity DNA Polymerase (Thermo Scientific, https://www.thermofisher.com/). A second PCR was performed using Gateway® compatible primers (forward: 5′-GGGGACAAGTTTG TACAAAAAAGCAGGCTTCACCATG-3′ and reverse: 5′-GGGGAC CACTTTGTACAAGAAAGCTGGGTC-3′). PCR products were electrophoresed on a 1% agarose gel and purified using QIAquick PCR purification kit (Qiagen, http://www.qiagen.com/).

Cloning into pDONR™ vectors

Purified PCR products were cloned into pDONR™223 or pDONR-F1-Zeo (Lalonde *et al.*, 2010) using Gateway® BP Clonase® II Enzyme Mix (Life Technologies) and incubated overnight at room temperature before bacterial transformation.

Construction of the initial pBullet plasmid

pBullet-cg-c-r (with mCherry) was created using two fragments (Table S4), one from a restriction digest with *Not*I and *Mlu*I of pBullet-cg-c-r (Part1) (Figure S5) and the other through amplification of the pUC origin from pDRf1-GW (Loque *et al.*, 2007) using PCR. The two fragments were fused together using In-Fusion HD kit (Clontech, http://www.clontech.com/). The original pBullet-cg-c plasmid was produced from a fragment from pBullet-cg-c-r cut with *Age*I and *Nsi*I and DNA amplified from a synthesized clone (pBullet-cg-c-ECFP) fused using In-Fusion HD (Table S4 and Figure S5). In order to convert pBullet-cg-c into a destination vector, the cytotoxic protein CcdB and the antibiotic resistance gene chloramphenicol acetyltransferase were inserted by performing a BP reaction using BP Clonase™ II Enzyme Mix and the donor vector Gateway® pDONR™/Zeo (Life Technologies).

Construction of the pBullet plasmids collection

Each additional pBullet plasmids were created from various combinations of DNA fragments (Table S4). Restriction digest of the expression clone pBullet-cg-c-APY1 (Parsons *et al.*, 2012) was used to generate most of the pBullet plasmid collection (Table S4). The pBullet-cg-n-APY1 DNA template was created using an LR reaction of apyrase 1 (At3 g04080) outlined in (Chiu *et al.*, 2012) and pBullet-cg-n. Markers for tonoplast (γ-TIP) and plasma membrane (PIP2A) were obtained from previously published subcellular marker plasmids (Nelson *et al.*, 2007). DNA fragments for markers of the *trans*-Golgi network (VTI12) and endosome (RABF2A) were created using parts from the WAVE plasmid collection (Geldner *et al.*, 2009). The pBullet-er-c and pBullet-cg-n were created using synthesized fragments (Figure S5) outlined in Table S4. Collectively, after obtaining the necessary DNA fragments, pieces were fused together using In-Fusion (Clontech) resulting in the pBullet expression clones. The genes encoding CcdB and chloramphenicol acetyltransferase were inserted into expression clones using BP reaction and the donor vector pDONR™/Zeo (Life Technologies) which led to the creation of the final pBullet destination vectors (Table 2).

Construction of pBullet expression clones

The GT14 pBullet Gateway® expression clones were created using an LR reaction (Life Technologies) with the entry clone and the appropriate pBullet plasmid. Reactions were left to incubate

overnight prior to bacterial transformation. The insertion of clones into the pBullet plasmids was verified by sequencing using: forward primer 5′-ACAAGTTTGTACAAAAAAGCAGGCTTC-3′ and reverse primer 5′-ACCACTTTGTACAAGAAAGCTGGGTC-3′.

Bacterial transformation

All bacterial transformations were performed with chemically competent *Escherichia coli* (DH5α). After heat shock, cells were spread on Luria Broth (LB) agar plates under appropriate selection and incubated overnight at 37°C. Plasmid DNA was isolated from 4 ml overnight cultures using a miniprep kit.

Particle bombardment

Microcarriers (gold 1 μm diameter, Bio-Rad, http://www.bio-rad.com/) were prepared following manufacturer's instructions. Specifically, 15 mg ml^{-1} of microcarrier was used instead of 60 mg ml^{-1}. A total of 2.5 μl of DNA was added into 25 μl of microcarrier/glycerol solution, followed by 25 μl of 2.5 M CaCl$_2$ and 10 μl of 0.1 M spermidine. Solution was vortexed for 10 min at 3000 rpm, spun down for 5 sec; supernatant removed. The pellet was washed with 140 μl of 100% ethanol, resuspended with 20 μl of 100% ethanol, loaded onto the macrocarrier, and allowed to dry. The original rupture disk retaining cap is used with the hepta adapter macrocarrier holder when setting up the PDS-1000/He. The macrocarrier is placed onto the centre holder, leaving the other six empty. Yellow onion epidermal peels were bombarded with a vacuum of 28 inHg, target distance of 6 cm, and helium pressure of 1100 psi. After bombardment, onions were kept on moist plates, overnight in the dark.

Confocal microscopy

Approximately 16–24 h after bombardment, the epidermal cells were removed from the onions and visualized using a Zeiss LSM 710 (Carl Zeiss, http://www.zeiss.com/) as previously outlined (Parsons *et al.*, 2012). Image analysis and processing (scale bar, brightness, and contrast) were done using IMAGEJ (Version 1.6r) (Schneider *et al.*, 2012).

Pearson's coefficient analysis

The Pearson's coefficient analysis was conducted on split channel images scaled to 256 × 256 pixels using IMAGEJ and with the JA-CoP plug-in (Bolte and Cordelieres, 2006).

Phylogenetic analysis

Protein sequences were used to build the tree using Muscle (maximum likelihood) from the MEGA 5 software package (Tamura *et al.*, 2011).

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. The GT cDNA cloning workflow used to create the JBEI GT collections.

Figure S2. Cloning status of the rice JBEI GT Collection.

Figure S3. Typical transformation efficiency after particle bombardment of onion epidermal cells using approximately 0.5 μg of pBullet-er-n.

Figure S4. Phylogenetic analysis of the Arabidopsis GT14 and GT14-like families.

Figure S5. Template DNA used for the construction of the pBullet plasmid collection.

Table S1. The collection of Arabidopsis glycosyltransferases according to CAZy.

Table S2. The current set of cloned rice glycosyltransferases in the collection and the curated list of all rice encoded glycosyltransferases.

Table S3. Typical plasmid yield (μg) from a 4 mL overnight culture of pBullet plasmids using a standard plasmid preparation (e.g. Qiagen QIAprep Spin Miniprep Kit).

Table S4. DNA template used to create additional pBullet plasmids containing various organelle markers.

REFERENCES

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an *Arabidopsis* interactome map. *Science*, **333**, 601–607.

Bolte, S. and Cordelieres, F.P. (2006) A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232.

Cao, P.J., Bartley, L.E., Jung, K.H. and Ronald, P.C. (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. *Mol. Plant.* **1**, 858–877.

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S. and Martens, S. (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J.* **69**, 1030–1042.

Chambers, S.P., Prior, S.E., Barstow, D.A. and Minton, N.P. (1988) The pMTL *nic-* cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene*, **68**, 139–149.

Chen, L.L., Marmey, P., Taylor, N.J. *et al.* (1998) Expression and inheritance of multiple transgenes in rice plants. *Nat. Biotechnol.* **16**, 1060–1064.

Chiu, T.Y., Christiansen, K., Moreno, I., Lao, J., Loque, D., Orellana, A., Heazlewood, J.L., Clark, G. and Roux, S.J. (2012) AtAPY1 and AtAPY2 function as Golgi-localized nucleoside diphosphatases in *Arabidopsis thaliana*. *Plant Cell Physiol.* **53**, 1913–1925.

Coutinho, P.M., Deleury, E., Davies, G.J. and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**, 307–317.

Curtis, M.D. and Grossniklaus, U. (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469.

Dilokpimol, A. and Geshi, N. (2014) *Arabidopsis thaliana* glucuronosyltransferase in family GT14. *Plant Signal. Behav.* **9**, e28891.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629.

Eswaran, N., Parameswaran, S., Sathram, B., Anantharaman, B., Kumar, G.R. and Tangirala, S.J. (2010) Yeast functional screen to identify genetic determinants capable of conferring abiotic stress tolerance in *Jatropha curcas*. *BMC Biotechnol.* **10**, 23.

Eudes, A., Baidoo, E.E., Yang, F., Burd, H., Hadi, M.Z., Collins, F.W., Keasling, J.D. and Loque, D. (2011) Production of tranilast [N-(3′,4′-dimethoxycinnamoyl)-anthranilic acid] and its analogs in yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **89**, 989–1000.

Eudes, A., George, A., Mukerjee, P. *et al.* (2012) Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnol. J.* **10**, 609–620.

Fu, X.D., Duc, L.T., Fontana, S., Bong, B.B., Tinjuangjun, P., Sudhakar, D., Twyman, R.M., Christou, P. and Kohli, A. (2000) Linear transgene constructs lacking vector backbone sequences generate low-copy-number transgenic plants with simple integration patterns. *Transgenic Res.* **9**, 11–19.

Gehl, C., Waadt, R., Kudla, J., Mendel, R.R. and Hansch, R. (2009) New GATEWAY vectors for high throughput analyses of protein–protein interactions by bimolecular fluorescence complementation. *Mol. Plant.* **2**, 1051–1058.

Gehl, C., Kaufholdt, D., Hamisch, D., Bikker, R., Kudla, J., Mendel, R.R. and Hansch, R. (2011) Quantitative analysis of dynamic protein-protein interactions in planta by a floated-leaf luciferase complementation imaging (FLuCI) assay using binary Gateway vectors. *Plant J.* **67**, 542–553.

Geldner, N., Denervaud-Tendon, V., Hyman, D.L., Mayer, U., Stierhof, Y.D. and Chory, J. (2009) Rapid, combinatorial analysis of membrane compartments in intact plants with a multicolor marker set. *Plant J.* **59**, 169–178.

Gelperin, D.M., White, M.A., Wilkinson, M.L. et al. (2005) Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* **19**, 2816–2826.

Grefen, C., Obrdlik, P. and Harter, K. (2009) The determination of protein-protein interactions by the mating-based split-ubiquitin system (mbSUS). *Methods Mol. Biol.* **479**, 217–233.

Greving, M., Cheng, X., Reindl, W. et al. (2012) Acoustic deposition with NIMS as a high-throughput enzyme activity assay. *Anal. Bioanal. Chem.* **403**, 707–711.

Ham, T.S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N.J. and Keasling, J.D. (2012) Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Res.* **40**, e141.

Hansen, S.F., Harholt, J., Oikawa, A. and Scheller, H.V. (2012) Plant glycosyltransferases beyond CAZy: a perspective on DUF families. *Front. Plant Sci.* **3**, 59.

Hartley, J.L., Temple, G.F. and Brasch, M.A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795.

Hawes, C. (2012) The ER/Golgi interface – is there anything in-between? *Front. Plant Sci.* **3**, 73.

Joshi, H.J., Christiansen, K.M., Fitz, J. et al. (2012) 1001 Proteomes: a functional proteomics portal for the analysis of *Arabidopsis thaliana* accessions. *Bioinformatics*, **28**, 1303–1306.

Kagale, S., Uzuhashi, S., Wigness, M., Bender, T., Yang, W., Borhan, M.H. and Rozwadowski, K. (2012) TMV-Gate vectors: gateway compatible tobacco mosaic virus based expression vectors for functional analysis of proteins. *Sci. Rep.* **2**, 874.

Kang, J.S., Frank, J., Kang, C.H. et al. (2008) Salt tolerance of *Arabidopsis thaliana* requires maturation of *N*-glycosylated proteins in the Golgi apparatus. *Proc. Natl Acad. Sci. USA* **105**, 5933–5938.

Karimi, M., Inze, D. and Depicker, A. (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* **7**, 193–195.

Kawahara, Y., de la Bastide, M., Hamilton, J.P. et al. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, **6**, 4.

Keinath, N.F., Kierszniowska, S., Lorek, J., Bourdais, G., Kessler, S.A., Shimosato-Asano, H., Grossniklaus, U., Schulze, W.X., Robatzek, S. and Panstruga, R. (2010) PAMP (pathogen-associated molecular pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity. *J. Biol. Chem.* **285**, 39140–39149.

Kelly, A.A. and Dörmann, P. (2004) Green light for galactolipid trafficking. *Curr. Opin. Plant Biol.* **7**, 262–269.

Knoch, E., Dilokpimol, A., Tryfona, T. et al. (2013) A beta-glucuronosyltransferase from *Arabidopsis thaliana* involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth. *Plant J.* **76**, 1016–1029.

Kues, U. and Stahl, U. (1989) Replication of plasmids in Gram-negative bacteria. *Microbiol. Rev.* **53**, 491–516.

Lalonde, S., Sero, A., Pratelli, R. et al. (2010) A membrane protein/signaling protein interaction network for *Arabidopsis* version AMPv2. *Front. Physiol.* **1**, 24.

Lamesch, P., Berardini, T.Z., Li, D. et al. (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–1210.

Li, B., Takahashi, D., Kawamura, Y. and Uemura, M. (2012) Comparison of plasma membrane proteomic changes of *Arabidopsis* suspension-cultured cells (T87 Line) after cold and ABA treatment in association with freezing tolerance development. *Plant Cell Physiol.* **53**, 543–554.

Lindbo, J.A. (2007) TRBO: a high-efficiency tobacco mosaic virus RNA-based overexpression vector. *Plant Physiol.* **145**, 1232–1240.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–495.

Loque, D., Lalonde, S., Looger, L.L., von Wiren, N. and Frommer, W.B. (2007) A cytosolic trans-activation domain essential for ammonium uptake. *Nature*, **446**, 195–198.

Manders, E.M., Stap, J., Brakenhoff, G.J., van Driel, R. and Aten, J.A. (1992) Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J. Cell Sci.* **103**, 857–862.

MGC Project Team (2009) The completion of the Mammalian Gene Collection (MGC). *Genome Res.* **19**, 2324–2333.

Millar, A.H., Carrie, C., Pogson, B. and Whelan, J. (2009) Exploring the function-location nexus: using multiple lines of evidence in defining the subcellular location of plant proteins. *Plant Cell*, **21**, 1625–1631.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T. and Kimura, T. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**, 34–41.

Nelson, B.K., Cai, X. and Nebenfuhr, A. (2007) A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J.* **51**, 1126–1136.

Nikolovski, N., Rubtsov, D., Segura, M.P., Miles, G.P., Stevens, T.J., Dunkley, T.P., Munro, S., Lilley, K.S. and Dupree, P. (2012) Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. *Plant Physiol.* **160**, 1037–1051.

Oikawa, A., Lund, C.H., Sakuragi, Y. and Scheller, H.V. (2013) Golgi-localized enzyme complexes for plant cell wall biosynthesis. *Trends Plant Sci.* **18**, 49–58.

Parsons, H.T., Christiansen, K., Knierim, B. et al. (2012) Isolation and proteomic characterization of the *Arabidopsis* Golgi defines functional and novel targets involved in plant cell wall biosynthesis. *Plant Physiol.* **159**, 12–26.

Rech, E.L., Vianna, G.R. and Aragao, F.J. (2008) High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nat. Protoc.* **3**, 410–418.

Sakai, H., Lee, S.S., Tanaka, T. et al. (2013) Rice Annotation Project Database (RAP-DB): an Integrative and Interactive Database for Rice Genomics. *Plant Cell Physiol.* **54**, e6(1–11).

Scheller, H.V. and Ulvskov, P. (2010) Hemicelluloses. *Annu. Rev. Plant Biol.* **61**, 263–289.

Scheller, H.V., Jensen, J.K., Sorensen, S.O., Harholt, J. and Geshi, N. (2007) Biosynthesis of pectin. *Physiol. Plant.* **129**, 283–295.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods.* **9**, 671–675.

Shaner, N.C., Steinbach, P.A. and Tsien, R.Y. (2005) A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909.

Somerville, C. (2006) Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* **22**, 53–78.

Stanley, P. (2011) Golgi glycosylation. *Cold Spring Harb. Perspect. Biol.* **3**, 1–9. pii: a005199.

Swarbreck, D., Wilks, C., Lamesch, P. et al. (2008) The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* **36**, D1009–1014.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Tang, K.X., Tinjuangjun, P., Xu, Y., Sun, X.F., Gatehouse, J.A., Ronald, P.C., Qi, H.X., Lu, X.G., Christou, P. and Kohli, A. (1999) Particle-bombardment-mediated co-transformation of elite Chinese rice cultivars with genes conferring resistance to bacterial blight and sap-sucking insect pests. *Planta*, **208**, 552–563.

Tanz, S.K., Castleden, I., Hooper, C.M., Vacher, M., Small, I. and Millar, H.A. (2013) SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. *Nucleic Acids Res.* **41**, D1185–1191.

The Rice Full-Length cDNA Consortium (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science*, **301**, 376–379.

Ton, V.K. and Rao, R. (2004) Functional expression of heterologous proteins in yeast: insights into Ca^{2+} signaling and Ca^{2+}-transporting ATPases. *Am. J. Physiol. Cell Physiol.* **287**, C580–589.

Tzfira, T., Tian, G.W., Lacroix, B., Vyas, S., Li, J., Leitner-Dagan, Y., Krichevsky, A., Taylor, T., Vainstein, A. and Citovsky, V. (2005) pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. *Plant Mol. Biol.* **57**, 503–516.

Vaistij, F., Lim, E.-K., Edwards, R. and Bowles, D. (2009) Glycosylation of secondary metabolites and xenobiotics. In *Plant-derived Natural Prod-ucts* (Osbourn, A.E. and Lanzotti, V., eds). Philadelphia, PA: Springer, pp. 209–228.

Yamada, K., Lim, J., Dale, J.M. *et al.* (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science*, **302**, 842–846.

Zhou, Y., Li, S., Qian, Q. *et al.* (2009) BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (*Oryza sativa* L.). *Plant J.* **57**, 446–462.