

# Secretion Trap Tagging of Secreted and Membrane-Spanning Proteins Using Arabidopsis Gene Traps<sup>1</sup>

Andrew T. Groover<sup>2</sup>, Joseph R. Fontana, Juana M. Arroyo, Cristina Yordan, W. Richard McCombie, and Robert A. Martienssen\*

Cold Spring Harbor Laboratory, Box 100, Cold Spring Harbor, New York 11724

Secreted and membrane-spanning proteins play fundamental roles in plant development but pose challenges for genetic identification and characterization. We describe a "secretion trap" screen for gene trap insertions in genes encoding proteins routed through the secretory pathway. The gene trap transposon encodes a  $\beta$ -glucuronidase reporter enzyme that is inhibited by N-linked glycosylation specific to the secretory pathway. Treatment of seedlings with tunicamycin inhibits glycosylation, resulting in increased activity of secreted  $\beta$ -glucuronidase fusions that result from gene trap integration downstream of exons encoding signal peptides. In the 2,059 gene trap lines that we screened, 32 secretion trap expression patterns were identified in a wide variety of tissues including embryos, meristems, and the developing vasculature. Genes disrupted by the secretion traps encode putative extracellular signaling proteins, membrane transport proteins, and novel secreted proteins of unknown function missed by conventional mutagenesis and gene prediction. Secretion traps provide a unique reagent for gene expression studies and can guide the genetic combination of loss of function alleles in related genes.

The large number of receptors encoded by the Arabidopsis genome (Arabidopsis Genome Initiative, 2000) suggests that receptor-ligand signaling may be prevalent in plants, but only a modest number of peptide ligands have been identified by traditional mutagenesis. This may reflect a prevalence of lethal, redundant, or conditional mutants in this class, such that phenotypes can only be observed under certain conditions or in certain genetic backgrounds. Furthermore, small secreted proteins that lack conserved domains are difficult to detect using sequence information alone (Ride et al., 1999; Vanoosthuysen et al., 2001). For example, genes encoding putative ligands related to CLAVATA3 had previously been either overlooked or improperly annotated by automated annotation programs (Cock and McCormick, 2001). Only *clv3* had been identified genetically.

Secreted proteins can be detected experimentally via the membrane anchors or N-terminal signal peptides that route them through the secretory pathway. Genetic screens designed to reveal the presence of targeting domains were initially established in bacteria (e.g. Manoil and Beckwith, 1986) and were later modified for use in eukaryotic systems. In mouse, reporter gene insertions into genes encoding secreted proteins ("secretory traps") have been identified in

embryonic stem cells based on membrane insertion of  $\beta$ -Gal fusions with "trapped" N-terminal secretion signals (Skarnes et al., 1995). Sequencing of 5'-RACE products revealed that almost one-half encoded secreted proteins, with the remainder being expressed because of improper splicing or deletions of the vector (Townley et al., 1997). The technique has been widely successful in identifying both known and novel secreted proteins including secreted peptide ligands (Serafini et al., 1996; <http://www.genetraps.org>). Gene traps disrupt genes, and secretory trap cell lines can be used to generate knock-out mice. Genes are thus initially discovered based on marker expression in cell culture, followed by functional characterization based on expression pattern and loss of function phenotype in mice.

In plants, expression of random Arabidopsis cDNAs as invertase fusions in yeast has been used to detect signal sequences, although several non-coding sequences were also identified by this heterologous approach (Goo et al., 1999). Plant cDNAs have also been expressed as epitope-tagged transgenes in mammalian cells (Kristoffersen et al., 1996) and green fluorescent protein (GFP) fusions in Arabidopsis (Cutler et al., 2000). About 2% of the GFP fusions had distinct subcellular localization patterns, including four localized to the vacuolar membrane and two fusions localized to the endoplasmic reticulum. Although useful in determining protein localization, these fusions were not driven by the endogenous promoter and did not disrupt the endogenous gene. Thus secretion traps offer distinct advantages over random cDNA fusions.

We have established a large collection of Arabidopsis gene trap insertion lines (<http://genetraps.cshl.org>) using a modified *Ds* transposable element that carries the  $\beta$ -glucuronidase (GUS) reporter gene (Springer et

<sup>1</sup> This work was supported by the National Institutes of Health (postdoctoral fellowship no. GM19974-02 to A.G.).

<sup>2</sup> Present address: U.S. Department of Agriculture Forest Service, Institute of Forest Genetics, 1100 West Chiles Road, Davis, CA 95616.

\* Corresponding author; e-mail [martiens@cschl.org](mailto:martiens@cschl.org); fax 516-367-8369.

Article, publication date, and citation information can be found at [www.plantphysiol.org/cgi/doi/10.1104/pp.103.020099](http://www.plantphysiol.org/cgi/doi/10.1104/pp.103.020099).

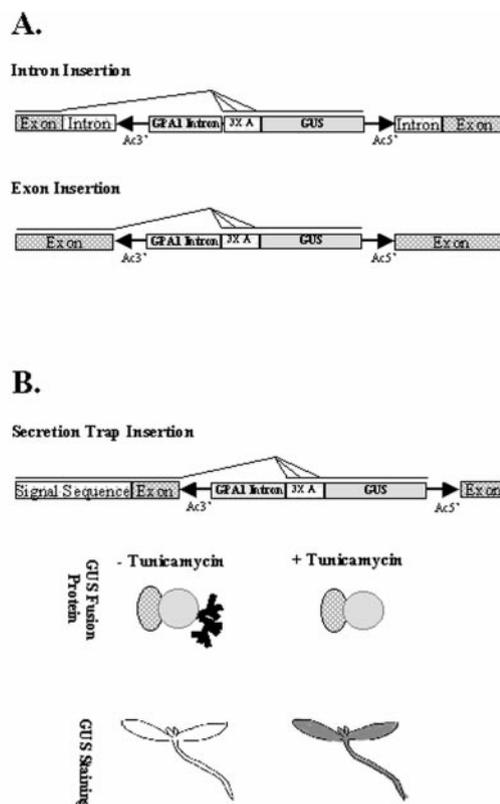
al., 1995; Sundaresan et al., 1995; Springer, 2000). The element is mobilized in a three-generation selection scheme that results in gene trap insertion lines, each carrying a unique stabilized transposon insertion. Gene trap insertion in the sense orientation leads to GUS reporter expression that mimics the normal expression pattern of the interrupted gene. Splice donor sites at the 5' end of the gene trap element and splice acceptor sites at the 5' end of the GUS gene result in translational fusions with the N-terminal portion of the interrupted gene product. The insertion sites are then amplified by PCR and sequenced to precisely map the insertion (Martienssen and Dolan, 1998). A major advantage of gene traps is that forward genetic screens can be applied to the collection based on expression pattern, loss of function phenotype, or changes in gene expression in response to experimental treatments.

Here, we describe a "secretion trap" screen for Arabidopsis gene trap insertions that disrupt genes encoding proteins routed through the secretory pathway. The screen was effective in identifying secreted and membrane-spanning proteins expressed in a wide variety of tissues. These proteins include receptors, membrane transport proteins, and novel secreted proteins of unknown function. All but one of the corresponding genes are members of gene families, and none of the insertions have a conspicuous loss-of-function phenotype. Secretion trapping is thus an effective method for characterizing secreted proteins missed by conventional mutagenesis.

## RESULTS

### Strategy for Gene Trap Tagging of Secreted Proteins

Secreted and membrane-spanning proteins contain targeting domains (transmembrane signal anchors or amino terminus signal peptides) that direct the protein to the endoplasmic reticulum (Bar-Peled et al., 1996) where they are exposed to *N*-linked glycosylation at specific sites (Asn-X-Ser/Thr; Hirschberg and Snider, 1987). The gene trap transposon used in our experiments has splice donor sites at the 5' end of the transposon and carries a plant intron and splice acceptor sites preceding the GUS reporter gene (Springer et al., 1995; Sundaresan et al., 1995). Thus, when the element inserts into exons or introns, translational fusions are generated between GUS and upstream exons from the interrupted gene (Fig. 1A). If the upstream exons encode a signal anchor or signal peptide, the resulting GUS fusion protein can be targeted to the secretory pathway. The GUS reporter protein is not normally secreted in plant cells, but contains a cryptic site for *N*-glycosylation at Asn-358 (Farrell and Beachy, 1990). GUS fusion proteins routed through the secretory pathway are enzymatically inhibited by *N*-linked glycosylation resulting in a decrease in colorimetric staining (Iturriaga et al., 1989). Enzymatic inhibition can be avoided by mu-



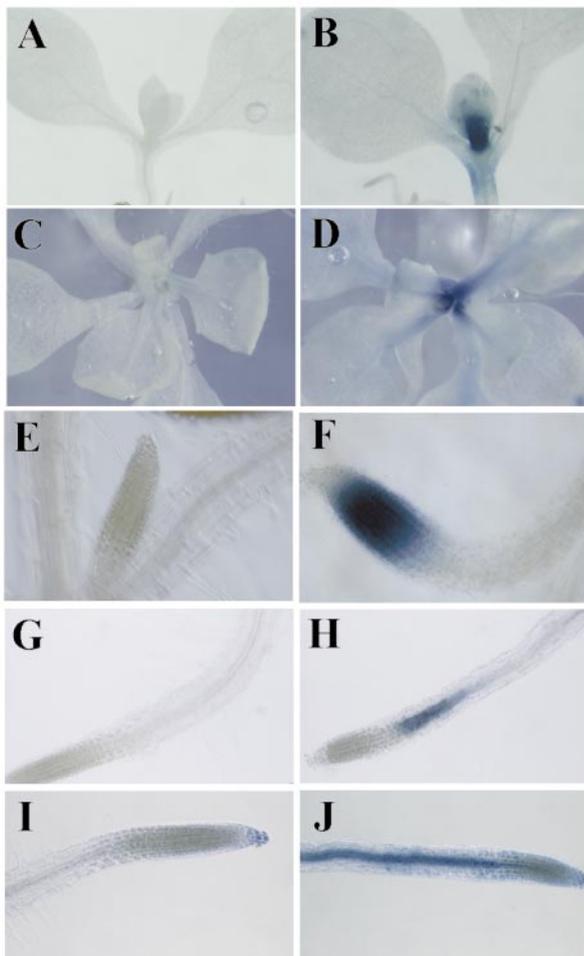
**Figure 1.** The capture of endogenous secretion signals in GUS fusions. A, Gene trap insertion into an intron of a chromosomal gene leads to use of one or more splice acceptor sequences aligned in all three reading frames ( $3 \times A$ ) with the GUS protein coding sequences (Sundaresan et al., 1995). Insertions into an exon use splice donors within the 3' Ds terminus to produce fusion GUS transcripts. B, Gene trap insertions in genes encoding secreted proteins can result in fusion GUS proteins routed through the secretory pathway. Such lines are termed secretion traps. For example, gene trap insertion downstream of a signal sequence produces a fusion protein that includes the N-terminal secretion signal. The secretion signal routes the fusion protein into the secretory pathway, where GUS is enzymatically inhibited by *N*-linked glycosylation. Glycosylation of fusion proteins is inhibited by tunicamycin and results in an increase in GUS histological staining.

tating Asn-358 to Lys (Firek et al., 1994) or else by treatment of cells with tunicamycin (Iturriaga et al., 1989), a specific inhibitor of Asn-linked glycan formation (Elbein, 1987).

We reasoned that gene trap insertions in genes encoding proteins routed through the secretory pathway could be systematically identified by comparing GUS staining of control seedlings with seedlings pretreated with tunicamycin (Fig. 1B). For example, 1-week-old transgenic seedlings expressing an  $\alpha$ -amylase-GUS fusion protein exhibited a marked increase in GUS staining after tunicamycin treatment (data not shown). Preliminary experiments using seedlings expressing the  $\alpha$ -amylase-GUS fusion protein indicated that 12-h treatment in liquid Murashige and Skoog medium containing  $20 \mu\text{M}$  tunicamycin was optimal ("Materials and Methods").

### Identification of Secretion Traps

One-week-old seedlings from 2,059 gene trap lines were transferred from Murashige and Skoog plates and grown for 12 h in liquid Murashige and Skoog medium containing either 20  $\mu\text{M}$  tunicamycin or mock solvent control before GUS staining under two stringency conditions ("Materials and Methods"). Of the 2,059 lines tested, 464 stain (23% of total), reflecting the approximate 50% gene density in Arabidopsis (Arabidopsis Genome Initiative, 2000) and the requirement for gene trap insertion in the sense orientation. Of these 464 lines, 32 secretion traps (7%) were identified that display pronounced, reproducible differences in GUS-staining intensity between tunicamycin and mock treatments (Fig. 2). GUS staining is



**Figure 2.** Examples of increased GUS histological staining in secretion trap lines after tunicamycin pretreatment. Seedlings from individual gene trap lines were pretreated with mock control solution (A, C, E, G, and I) or tunicamycin (B, D, F, H, and J) before GUS histological staining ("Materials and Methods"). Lines GT7137 (A and B) and GT7487 (C and D) have differential GUS staining in the shoot apex. Line GT5397 (E and F) stains the root tip after tunicamycin treatment. Lines GT7079 (G and H) and GT5376 (I and J) display stronger GUS staining in the elongation zone of the root after tunicamycin treatment.

unaffected by tunicamycin treatment in the majority of lines screened, indicating that this treatment does not affect GUS expression nonspecifically. Increased GUS histological staining is not the result of transcriptional induction by tunicamycin in most lines. Six of the nine lines (GT5211, GT7094, GT6249, GT5376, GT6224, and GT7059) analyzed by semi-quantitative PCR ("Materials and Methods") had indistinguishable GUS transcript levels in tunicamycin versus mock-treated samples (not shown). Two lines, GT5397 and GT6700, had elevated transcript levels in tunicamycin-treated seedlings. One line, GT6666, had decreased GUS transcript levels. Of course, elevated transcript levels do not preclude secretion.

Tunicamycin-responsive GUS staining was observed in various organs and tissues of secretion trap lines (Table I), including shoots (e.g. GT7137 and GT7487; Fig. 2, A–D) and roots (e.g. GT5397, GT7079, and GT5376; Fig. 2, E–J), indicating that most seedling tissues are susceptible to tunicamycin treatment. Some secretion trap lines display significant GUS-staining differences only within the root, especially within the elongation zone (e.g. Fig. 2, G–J). This may reflect a higher uptake of tunicamycin in the root or a high rate of secretion in the elongation zone. Secretion traps GT5397, GT7094, and GT7106 had no detectable staining in mock-treated seedlings, but residual staining was observed in the other 11 lines. Residual GUS activity in mock-treated seedlings could reflect incomplete glycosylation or splicing variants. Most secretion trap lines harbor a single gene trap insertion (Table I).

Transcriptional and translational fusion with the GUS reporter gene were demonstrated in the case of secretion trap GT5376, which lies within the second exon of At5g67600, an annotated gene predicted to encode a small protein (Table II). Reverse transcriptase (RT)-PCR was used to amplify fusion mRNAs consisting of GUS and upstream exons from the interrupted gene (Fig. 3A). The first two exons were fused to the GUS reporter via splice donor and acceptor sites from the gene trap. Western blots probed with an anti-GUS antibody ("Materials and Methods") revealed a fusion protein of the expected size (Fig. 3B). We cannot exclude the possibility that low levels of residual unfused GUS protein fragments were also present, because these were obscured by a spurious, cross-reacting protein of the same size found in plants without a gene trap insertion (WT). Nonetheless, these results demonstrate that secretion traps generate bona fide transcriptional and translational fusions of endogenous genes with the GUS reporter.

### Gene Expression and Phenotypic Characterization of Secretion Trap Lines

A wide variety of reporter gene expression patterns were detected in 1-week-old seedlings from secretion

**Table I.** Gene expression patterns of secretion trap lines

Secretion Trap Line	Root	Hypocotyl	Cotyledon	Leaf	Shoot Apex	Embryos
GT5211	Vasculature	Vasculature	Vasculature	Vasculature		Diffuse
GT5222	Root tip and lateral root	Diffuse	Diffuse	Diffuse	Stipules	None detected
GT5358	Diffuse, vasculature darker	Diffuse	Diffuse			Not tested
GT5376	Vasculature, root cap	Vasculature	Vasculature	Vasculature	Stipules	None detected
GT5397	Root tip					None detected
GT5926	Vasculature	Diffuse, with vasculature darker	Diffuse, with vasculature darker			None detected
GT6224	Vasculature	Vasculature	Vasculature	Vasculature		Not tested
GT6249 <sup>a</sup>	Root top and vasculature			Vascular	Shoot apex	Provascular
GT6434			Vasculature and petiole		Stipules	None detected
GT6548	Root tip		Diffuse	Diffuse		None detected
GT6604	Diffuse, colet		Diffuse			None detected
GT6647		Vasculature	Vasculature and hydathodes			None detected
GT6660	Diffuse					Not tested
GT6666	Root tip				Shoot apex	None detected
GT6700	Vasculature	Vasculature	Vasculature	Vasculature		None detected
GT6943	Diffuse	Diffuse	Diffuse	Diffuse, vasculature darker		None detected
GT7004	Diffuse, vasculature darker		Diffuse, with vasculature darker	Diffuse, vasculature darker		None detected
GT7010	Diffuse		Diffuse	Diffuse		None detected
GT7036			Diffuse	Diffuse		Not tested
GT7059	Root tip		Diffuse			None detected
GT7079	Diffuse, elongation zone	Vasculature	Vasculature			None detected
GT7092 <sup>b</sup>	Root tip, lateral root primordia				Stipules	None detected
GT7094	Diffuse	Diffuse	Diffuse			None detected
GT7106	Vasculature	Vasculature			Vasculature	None detected
GT7134 <sup>b</sup>	Root tip	Vasculature	Vasculature			None detected
GT7137	Diffuse	Diffuse	Diffuse	Diffuse	Shoot apex	None detected
GT7208 <sup>c</sup>	Root tip		Diffuse	Diffuse		None detected
GT7368			Diffuse	Diffuse		None detected
GT7487	Ring around lateral root, root initials	Diffuse	Diffuse		Shoot apex	Root tip, cotyledon, hypocotyl
GT7488 <sup>b</sup>	Root tip, vasculature	Diffuse	Diffuse	Diffuse		Not tested
GT7519	Root tip	Vasculature	Vasculature	Vasculature		None detected
GT7544	Diffuse			Vasculature		None detected

<sup>a</sup> Two gene trap insertions in original line. Single insertion in derived line used for analysis. <sup>b</sup> No. of gene trap insertions not determined. <sup>c</sup> Two gene trap insertions.

trap lines, reflecting the role of secreted proteins in many different tissues. For example, vascular patterning and differentiation are thought to involve secreted proteins (e.g. Groover and Jones, 1999), although few such proteins have been genetically characterized. Secretion trap lines GT5211, GT5376, GT5926, GT6224, GT6249, GT6647, GT6700, GT7079, GT7134, GT7519, and GT7106 all express GUS during vascular development or within specific vascular cell types, including provascular cells (Table I; Fig. 4). Some genes are expressed in patterns that include more than one tissue or cell type, as illustrated by GUS expression in GT6249 both within developing vasculature and the shoot apical meristem (Table I;

Fig. 4C). Other patterns do not conform to anatomically defined structures. For example, GT7487 is expressed in a ring surrounding lateral root primordia (not shown). Microarray profiling would not reveal such complex or detailed expression patterns.

Patterning of the provasculature occurs during early embryogenesis (Carland et al., 1999), but the differentiation of functional cell types does not occur until after germination (Dharmawardhana et al., 1992). Genes expressed in provascular cells during embryogenesis are thus good candidates for regulators of vascular patterning. Most secretion traps show residual GUS activity in the absence of tunicamycin, allowing us to assay GUS expression during

Table II. Secretion trap lines with insertions in annotated genes

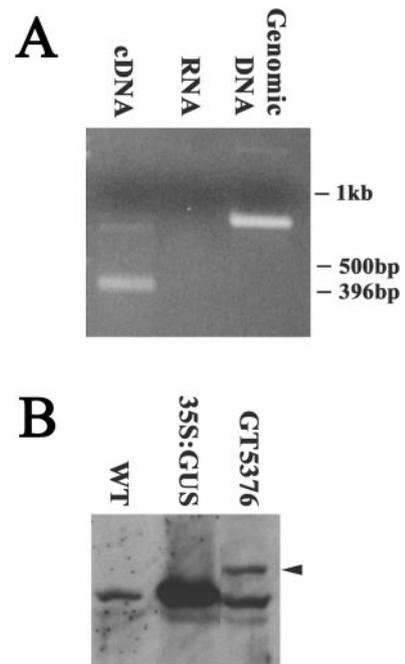
Secretion Trap Line	Insertion Site	Insertion Orientation	TAIL Sequence Accession No.	Gene Predicted at Insertion Site	Secreted	Protein Homologies	Domains	Number Similar Genes <sup>a</sup>
GT5376	g 3128137 dbj AB013390.1 AB013390 Chromosome 5, TAC clone:K919 Position 39,103	-	AF500883	Gene = K919.17 Evidence = not experimental Protein id = BAB08468.1	Type II membrane protein	Pro-rich proteins	Undefined Cys-rich C terminus	2
GT6114	g 3046848 dbj AB012240.1 AB012240 Chromosome 5, TAC clone:K18C1 Position 49,821	+	AF500884	Gene = K18C1.9 Product = disease resistance, protein-like Protein id = BAB11394.1	No significant prediction	Disease resistance proteins	1) Toll-interleukin 1-resistance 2) ATPases	>50
GT6224	g 4376087 emb Z99707.1 ATAP21 Chromosome 4, ESSA I AP2 contig fragment no. 1 Position 37,400	+	AF500885	Gene = C7A10.110 Product = receptor kinase-like protein Protein id = CAB16774.1	Type Ia membrane protein	Receptor kinases	1) Multiple LRRs 2) Protein kinase	>50
GT6434	g 12323158 gb AC027034.10 AC027034 Chromosome 1, BAC F7A10 Position 54,388	+	AF500886	Gene = F7A10.4 Product = unknown protein Protein id = AAG51563.1	Type IIIb membrane protein	No significant homologies	No domains predicted	0
GT6660	g 6598405 gb AC004136.2 AC004136 Chromosome 2, section 10 of 255 Position 60,647	+	AF500887 AF500888	Gene = A2g02610 Product = hypothetical protein Protein id = AAC18926.1	Putative transmembrane	Large number plant hypotheticals	No domains predicted	>50
GT6666	g 6598453 gb AC005309.2 AC005309 Chromosome 2, section 254 of 255 Position 43,209	+	AF500889 AF500890	Gene = A2g47800 Product = glutathione-conjugate transporter AIMRP4 Protein id = AAC63634.1	Type IIIa membrane protein	ABC transporters	Two ATPase domains	44
GT6700	g 9542859 emb AL137189.2 ATF718 Chromosome 5, BAC F718 Position 72,530	+	AF500891 AF500892	Gene = F718_180 Product = similarity to Myb-transcriptional activator Protein id = CAB69848.1	No significant prediction	Myb transcription factors	Myb	17
GT6943	g 4006885 emb Z99708.1 ATAP22 Chromosome 4, ESSA I AP2 contig fragment no. 2 Position 42,494	-	AF500893	Gene = C7A10.600 Product = aminopeptidase-like protein Protein id = CAB16823.1	No significant prediction	Proteases	Metallopeptidase family M24	1
GT7036	g 2252848 gb AF013294.1 TM018A10 Chromosome 4, BAC TM018A10 Position 52,742	+	AF500894	Gene = A_TM018A10.5 Product = unknown protein Protein id = AAB62851.1	Putative transmembrane domain	Cell surface proteins	No domains predicted	7
GT7059	g 12408717 gb AC009325.8 ATAC009325 Chromosome 3, BAC F4P13 Position 66,328	+	AF500895	Gene = F4P13.19 Product = unknown protein Protein id = AAF01562.1	Type II membrane protein	<i>Caenorhabditis elegans</i> and human copines	1) von Willebrand factor type A 2) Ring finger	5
GT7092 <sup>b</sup>	g 3859590 gb AF104919.1 T15B16 Chromosome 4, BAC T15B16 Position 23,629	+	AF500896 AF500897	Gene = T15B16.12 Product = contains similarity to wild oat DNA-binding protein Protein id = AAC72869.1	No significant prediction	Transcription factors	WRKY domain	11
GT7094	g 6598658 gb AC007069.5 AC007069 Chromosome 2, section 5 of 255 Position 32,976	-	AF500898 AF500899	Gene = A2g01850 Product = xyloglucan-specific glucanase Protein id = AAD21783.1	Cleavable N-term signal sequence	Endo-xyloglucan transferases	Glycosyl hydrolases family 16	32

(Table continues on following page.)

**Table II.** (Continued from previous page.)

Secretion Trap Line	Insertion Site	Insertion Orientation	TAIL Sequence Accession No.	Gene Predicted at Insertion Site	Secreted	Protein Homologies	Domains	Number Similar Genes <sup>a</sup>
GT7106	gj 3193311 gb AF069299.1 F6N15 Chromosome 4, BAC F6N15 Position 87,761	+	AF500900	Product = subtilisin-type serine endopeptidase XSP1 Protein id = AAF25830.1	Cleavable N-term signal sequence	Subtilisin-type serine peptidases	Subtilase Peptidase	>50
GT7134 <sup>b</sup>	gj 2098816 gb AF000657.1 ATAF000657 Chromosome 1, BAC F19G10 Position 14,182	-	AF500901 AF500902	Gene = F10G19.3 Product = hypothetical protein Protein id = AAB72157.1	Type II membrane protein	Armadillo repeat-containing proteins	1) Multiple Armadillo repeats 2) Ubox	44
GT7137	gj 12408716 gb AC009327.8 ATAC009327 Chromosome 3, BAC T12J13 Position 44,498	-	AF500903	Gene = T12J13.8 Product = $\beta$ -glucosidase Protein id = AAF03468.1	Cleavable N-term signal sequence	Glycosidases	Glycosyl hydrolase family 1	45
GT7487	gj 12408739 gb AC013483.7 ATAC013483 Chromosome 3, BAC F17A17 Position 115,174	+	AF500904 AF500905	Gene = F17A17.37 Product = unknown protein Protein id = AAF21213.1 db_xref = GI:6648215	Cleavable N-term signal sequence	Undefined plant proteins	No domains predicted	9

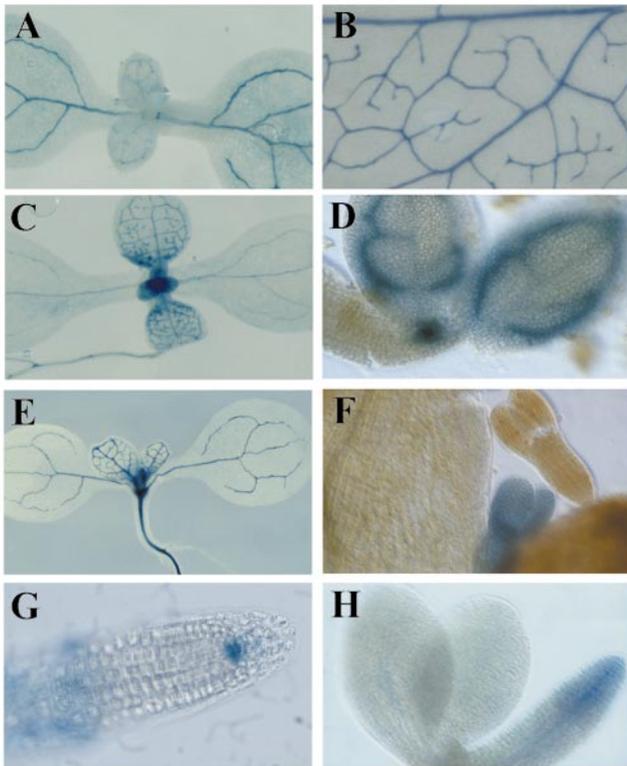
<sup>a</sup> No. of proteins encoded by the Arabidopsis genome showing BLAST *P* scores < e-20. <sup>b</sup> No. of insertions not determined.



**Figure 3.** Molecular analysis of secretion trap line GT5376. A, cDNA (lane 1) or genomic DNA (lane 3) templates from GT5376 seedlings were amplified using exon and GUS primers (“Materials and Methods”). The smaller size of the PCR product from cDNA template compared with genomic DNA reflects splicing of the fusion mRNA. Mock reaction of RNA without RT yields no product (lane 2). B, Protein gel blot incubated with anti-GUS antibodies. Fusion GUS protein from GT5376 seedlings (lane 3) migrates slower (arrow) than native GUS in extracts from a 35S::GUS transgenic line (lane 2). An additional similar-sized protein cross-reacts in wild-type seedlings (lane 1) and in GT5376 (lane 3).

embryogenesis when tunicamycin cannot be readily applied. Embryos at late torpedo to walking-stick stage were partially dissected from siliques and stained for GUS expression. Three lines, GT5211, GT6249, and GT7487, have detectable GUS expression in embryos (Table I). GT6249 is expressed in the shoot apical meristem and provascular tissues in both embryos and seedlings (Fig. 4, C and D). GT5211 is diffusely expressed during embryogenesis (Fig. 4F) but restricted to vascular tissues in seedlings (Fig. 4E). GUS expression in GT7487 is found in the root initials during seedling growth (Fig. 4G) and embryogenesis (Fig. 4H).

Secretion trap lines were scored for visible phenotypes as seedlings and as mature plants grown under both long and short day conditions. Secretion trap lines were also assayed for embryo lethal phenotypes. Three of the 33 lines assayed showed putative mutant phenotypes: GT7079 grew as a stunted, darker green plant, whereas GT7208 and GT6249 segregated embryo lethality. Subsequent cosegregation analysis indicated that these phenotypes were due to mutations unlinked to the gene trap (data not shown). Thus, gene trap mutagenesis of the genes defined by the 32 secretion trap lines did not reveal readily detectable developmental phenotypes.



**Figure 4.** Examples of secretion trap gene expression patterns. The genes disrupted in GT5376 (A) and GT6224 (B) are expressed in vascular tissues, as revealed by GUS histological staining. GT6249 is expressed in vascular tissues and the apical meristem in 1-week-old seedlings (C) and in provascular tissues and the apical meristem in embryos (D). GT5211 is expressed in vascular tissues in 1-week-old seedlings (E) and during embryogenesis (F). GT7487 is expressed in the root meristem in 1-week-old seedlings (G) and in embryos (H).

### Genes Identified by Secretion Trap Lines

Chromosomal DNA flanking the insertion sites in secretion trap lines was amplified by Thermal Asymmetric Interlaced PCR and sequenced ("Materials and Methods"). The chromosomal location of individual insertions was determined by BLAST analysis, and was confirmed by either supporting TAIL sequences from each end of the transposon or by PCR using a gene-specific primer in combination with a transposon primer (data not shown). Current annotation of the Arabidopsis genome indicates that 15 of the 21 insertion sites sequenced are within predicted genes. Of the 15 genes, 11 encode proteins with predicted signal or transmembrane domains ("Materials and Methods") or with similarity to known secreted proteins (Table II). Only one of the insertions (GT6434) interrupts a single-copy gene, whereas the other 14 interrupt genes belong to gene families of between two and >50 members (Table II). The tagged genes encode diverse classes of proteins and are described in more detail below.

### Transmembrane Transport

GT6666 disrupts a member of the ABC transporter super family. The protein, AtMRP4, is one of 15 members of the multidrug resistance-associated protein subfamily in Arabidopsis (Sanchez-Fernandez et al., 2001). Although AtMRP4 has not been functionally described, it is expected to transport glutathione S-conjugated compounds, as has been demonstrated for AtMRP1-3 (Sanchez-Fernandez et al., 2001). We considered the possibility that AtMRP4 was transcriptionally up-regulated to clear tunicamycin from challenged cells, but semiquantitative RT-PCR ("Materials and Methods") revealed that transcripts were actually reduced in treated seedlings, indicating that routing through the secretory pathway was the basis for identification in the secretion trap screen, rather than transcriptional up-regulation.

### Putative Extracellular Matrix Proteins and Cell Wall-Related Proteins

A subset of secretion trap lines defines genes whose products have primary sequence features consistent with extracellular matrix proteins or proteins involved in the assembly or modification of the cell wall. GT7094 and GT7137 disrupt genes encoding xyloglucan glucanase and  $\beta$ -glucosidase enzymes, respectively, which could modify glycoprotein or cell wall carbohydrate linkages. Both proteins contain a predicted N-terminal secretion signal and are broadly expressed within 1-week-old seedlings (Table I). Although these proteins have not been functionally characterized, their expected functions would be consistent with modification of cell wall or glycoprotein linkages ( $\beta$ -glucosidase) or loosening of cell wall through hemicellulose modification during expansive growth (xyloglucan glucanase). Both proteins are members of large gene families (Table II).

Putative extracellular matrix proteins include those disrupted by GT5376, GT7036, and GT7059. GT5376 is expressed in the vasculature (Fig. 4A), and disrupts a gene encoding a small 82-amino acid Pro-rich (22% Pro) type II membrane protein. The protein contains a conserved C-terminal Cys-rich domain of unknown function also found in proteins of similar size and amino acid composition from loblolly pine (*Pinus taeda*, a gymnosperm; accession no. AAF75822.1; length = 86 amino acids) and resurrection grass (*Sporobolus stapfianus*, a monocot; accession no. CAA71756.1; length = 86 amino acids). GT7036 disrupts a gene encoding a hypothetical protein that is Pro and Ser rich (18.6% and 17.4%, respectively). Pro-rich proteins have been implicated in diverse aspects of cell wall structure and function in plants (Cassab, 1998). GT7059 disrupts an uncharacterized gene encoding an unknown type II membrane protein with von Willebrand Factor and Ring Finger domains. The von Willebrand Factor domain mediates adhesion via metal ion-dependent adhesion sites

in diverse animal extracellular matrix proteins including cartilage matrix proteins, collagens, and von Willebrand factor, mutant forms of which are involved in the etiology of bleeding disorders (Tuckwell, 1999). The domain is also found in the membrane-spanning integrins, where it plays a central role in ligand binding. The protein identified by GT7059 is thus an excellent candidate for mediating extracellular matrix trans-membrane anchorage and/or signaling.

### Putative Receptor-Ligand Signaling Proteins

GT6224 disrupts a gene encoding a predicted type Ia membrane-spanning receptor containing Leu-rich repeats and a protein kinase domain and is expressed in vascular tissues throughout seedlings (Fig. 4B). This orphan receptor is one of the predicted 82 kinase domain-containing Leu-rich repeats in Arabidopsis, most of which await characterization (Arabidopsis Genome Initiative, 2000). Recently, a vascular expressed Leu-rich repeat receptor kinase was identified through enhancer trapping that regulates pro-vascular development (Clay and Nelson, 2002).

Ser proteases play diverse roles including the processing of both ligands and receptors to active forms in animals. The GT7106 insertion disrupts a gene encoding a Ser protease belonging to a large protein family. The predicted gene model in GenBank (NP\_191934.1) does not include the complete coding sequence, and an alternative gene model (Zhao et al., 2000) predicts an upstream secretion signal. GUS expression in this line is specific to differentiating tracheary elements. Interestingly, a secreted Ser protease has been implicated in the regulation of programmed cell death in this cell type (Groover and Jones, 1999).

### Unknown Genes

Three of the secretion trap lines disrupt genes for which no functional assignment could be made based on similarity searches. GT7487 disrupts a gene encoding a predicted 39-kD unknown protein that lacks any obvious functional domains, except for a strongly predicted secretion signal. The tagged gene is expressed in the root meristem quiescent center in 1-week-old seedlings, in the root tip during embryogenesis (Fig. 4, G and H), and in the shoot apical meristem in 1-week-old seedlings (Fig. 2, C and D). The protein is highly similar (e-123 to 2e-032) to nine other proteins in Arabidopsis, and to proteins from rice (*Oryza sativa*; accession no. BAB21293), castor bean (*Ricinus communis*; accession no. T10174), alfalfa (*Medicago sativa*; accession no. T09642), and chickpea (*Cicer arietinum*; accession no. CAA06490). There is no functional information currently available for any of these proteins.

Secretion trap GT6660 disrupts a 72-kD hypothetical protein with a putative transmembrane domain. Although there are over 100 similar family members in Arabidopsis, none have a known function. GT6434 disrupts a gene encoding a predicted type III membrane protein that could not be assigned a putative function based on domains, amino acid composition, or homologies. In contrast to the other secretion traps, GT6434 identifies a potential single-copy gene. The gene is expressed within the vasculature and petiole of the cotyledon in 1-week-old seedlings (Table I).

The gene disrupted by GT7134 encodes a type III membrane protein containing a U box motif and four putative Armadillo repeats. The U box motif is a modified ring finger found within a subset of proteins involved in ubiquitination (Aravind and Koonin, 2000), whereas the Armadillo repeat consists of 40 amino acids that mediate interaction of Armadillo/ $\beta$ -catenin proteins with their ligands.

Two secretion trap lines identified genes that are not predicted to encode secreted or membrane-spanning proteins (GT6700 and GT6943). GT6943 disrupts a gene that encodes an amino peptidase that does not contain a secretion signal. It is possible that the N terminus and signal sequence have not been included in the gene model. However, GT6700 disrupts an myb-class transcription factor expressed in the vasculature which is transcriptionally induced by tunicamycin treatment (data not shown). This may be the basis for its identification in the secretion trap screen, although secreted transcription factors are known in animals (e.g. Maizel et al., 1999).

### Insertions outside Annotated Genes

In mouse stem cells, secretion trapping has been very successful in identifying novel cell-cell signaling proteins (<http://socrates.berkeley.edu/~skarnes/resource.html>). Even so, 60% of mouse secretion trap cell lines did not detect secreted proteins and were only selected because of improper splicing or vector deletion (Townley et al., 1997). Two of our Arabidopsis secretion trap lines have a single insertion outside annotated genes (GT5211 and GT6249). We considered the possibility that transcription could initiate within the gene trap element (Cocherel et al., 1996) and come under the control of a tunicamycin-induced enhancer, but tests indicated that these insertions are not transcriptionally induced by tunicamycin (Table III). Furthermore, GUS expression patterns were specific to individual cell types, suggesting that they do not reflect rogue transcription. For example, GT6249 is expressed in developing vascular tissues (Fig. 4C), the shoot apical meristem, and the root columella initials in 1-week-old seedlings (Table I). Expression is also found in provascular tissues and the shoot apical meristem in developing embryos (Fig. 4D).

**Table III.** Secretion trap lines with insertions outside of predicted protein coding regions

Secretion Trap Line	Insertion Site	Insertion Orientation	TAIL Sequence Accession	Transcript Induced <sup>a</sup>
GT5211	gi 4159700 dbj AB022211.1 AB022211 Chromosome 5, TAC clone:K1L20 Position 34,371	–	AF500872 AF500873	No
GT5397	gi 4218109 emb AL035353.1 ATF16A16 Chromosome 4, BAC clone F16A16 Position 21,512	+	AF500874	Yes
GT6249	gi 10086525 gb AC079041.4 AC079041 Chromosome 1, BAC F5M6 Position 42,977	–	AF500875 AF500876	No
GT7208 <sup>b</sup>	i 4467094 emb AL035538.1 ATF20D10 Chromosome 4, BAC F20D10 Position 91,963	+	AF500877 AF500878	ND
GT7368	gi 7363407 gb AC025290.3 F9P14 Chromosome 1, BAC F9P14 Position 4,341	–	AF500879 AF500880	ND
GT7488 <sup>c</sup>	gi 2828278 emb AL021687.1 ATT18B16 Chromosome 4, BAC T18B16 Position 31,210	–	AF500881 AF500882	ND

<sup>a</sup> GUS transcript levels were compared by semi-quantitative RT-PCR for  $\pm$  tunicamycin-treated seedlings (“Materials and Methods”). <sup>b</sup> Two gene trap insertions. <sup>c</sup> No. of gene trap insertions not determined.

GUS fusion transcripts were amplified from GT6249 using 5'-RACE to determine the sequences responsible for GUS expression (“Materials and Methods”). The single amplified product was cloned and sequenced, and the 3' end of the corresponding endogenous gene was amplified with 3'-RACE using a primer (5'-cgaccgggttcgtctctgttctc-3') complementary to the 5'-RACE product (“Materials and Methods”). Collectively, the RACE cDNA and genomic sequences indicate that the gene trap insertion is in the first intron of a gene with seven exons. The 3' portion of the RACE cDNA includes exons from a downstream, annotated gene encoding a putative membrane-spanning major intrinsic protein channel (F5M6.11; accession no. AC079041.4) on chromosome 1. The annotated translation start (position 40,306 on bacterial artificial chromosome [BAC] F5M6) is nearly 3 kb downstream of the gene trap insertion (position 42, 977 on BAC F5M6), but the RACE cDNA sequence shows there are at least three unannotated exons upstream, comprising 462 bp of transcript. An alternative gene model places the translation start in the second exon (position 42,785 on BAC F5M6) defined by the RACE cDNA (rather than the third, as annotated). Although the gene trap insertion is in the 5'-untranslated region in this gene model, it is possible that alternate upstream translation start sites are used or that alternative splicing results in GUS fusions to additional upstream exons.

## DISCUSSION

We describe a new genetic method that uses Arabidopsis gene traps to identify novel proteins based on targeting to the secretory pathway as well as their

expression pattern. The secretion trap screen was effective in enriching for genes encoding secreted and membrane-spanning proteins. Of the 15 annotated proteins identified, 11 (73%) are predicted by computer algorithms to have sequence features consistent with routing through the secretory pathway (“Materials and Methods”) or to have enzymatic functions requiring endomembrane insertion or whose substrate is unique to the secretory pathway. In addition to proteins resembling known secreted proteins, we found several genes not previously identified by conventional means. Encouragingly, the frequency of bona fide secretion traps identified in our screen is substantially higher than in mouse secretory trap screens, possibly reflecting the compact structure of plant genes (Arabidopsis Gene Initiative, 2000).

The secretion trap screen potentially identifies any protein containing routing signals capable of directing GUS fusions to the secretory pathway. Tunicamycin treatment inhibits early steps in glycosylation, but the critical step at which glycosylation inhibits GUS enzymatic activity is not currently known. It is thus possible that only more extensive elaboration of carbohydrate linkages later in the secretory pathway is sufficient to inhibit GUS, in which case, a more limited subset of proteins would be identified by the screen. Tunicamycin treatment may lead to changes in GUS activity based on destabilization of fusions with normally glycosylated proteins, but this is not a significant disadvantage of the screen because this would reduce rather than increase GUS activity. A more likely source of false positives results from direct transcriptional induction of a tagged gene by

tunicamycin, but our results show that is not the case for the majority of genes identified here.

Secretion traps are a valuable resource because they report the pattern of gene expression within individual cells in complex tissues, as well as the timing of gene expression at multiple stages of development. Most important, they disrupt the genes they report. Detailed knowledge of gene expression can then guide efforts to identify subtle mutant phenotypes resulting from insertion, especially when paired with functional information. For example, the potassium channel gene *AKT1* is only expressed in the roots, and T-DNA insertion alleles are phenotypically wild type, except when roots were challenged in medium containing minimal potassium (Hirsch et al., 1998). The effect of experimental treatments and mutant backgrounds can be easily monitored using secretion traps. For example, several of the genes we have identified are expressed in the developing vasculature. Auxin regulates vascular development, and the effect of exogenous auxin on gene expression could be readily monitored in these lines, as could the effects of auxin-related mutants.

Secretion traps can also identify redundant genes. Insertion of the gene trap transposon typically leads to gene disruption (Springer et al., 1995), but none of the secretion trap lines have a conspicuous developmental phenotype and none of the genes have been found previously through conventional mutagenesis. The observation that all but one of the genes disrupted is a member of a gene family is consistent with the notion that genetic redundancy masks the phenotypic consequence of disruption. In principle, secretion trap expression patterns can be used to guide the construction of double mutants in redundant gene family members. This is because genes expressed in the same cell types are much more likely to have overlapping functions (Martienssen and Irish, 1999; Ferrandiz et al., 2000; Pelaz et al., 2000; Byrne et al., 2002). In addition, misexpression of these genes could be used to produce dominant phenotypes, a particularly powerful approach in the case of developmental signals responsible for morphogenesis. Finally, transposons can be remobilized to generate genetic mosaics, allowing the cellular autonomy of any phenotypes to be examined (Jenik and Irish, 2001). Overall, we believe secretion traps will provide a powerful resource for detecting cell-cell signaling proteins in plants.

## MATERIALS AND METHODS

### Plant Cultivation, Drug Treatment, and Histological Staining

Seed from individual *Arabidopsis* gene trap lines (ecotype Landsberg *erecta*) was surface sterilized for 10 min in 10% (w/v) bleach and 0.1% (w/v) Tween, washed twice with sterile water, and sown onto plates containing Murashige and Skoog medium (1× Murashige and Skoog salts [Invitrogen, Carlsbad, CA], 0.5 g L<sup>-1</sup> MES, and 1% [w/v] Suc, pH 5.7), solidified with 8 g L<sup>-1</sup> phytigel (Sigma-Aldrich, St. Louis). Plates were incubated under con-

tinuous illumination at an angle of 35°. After 7 d, the roots of seedlings had grown across the surface of the agar, allowing seedlings to be lifted off intact and placed into cell culture plates (no. 3524, Costar, Corning, NY) containing liquid Murashige and Skoog medium. Preliminary experiments using transgenic seedlings expressing an  $\alpha$ -amylase signal peptide-GUS fusion construct (Firek et al., 1994) indicated that overnight growth with gentle shaking in liquid medium supplemented with 20  $\mu$ M tunicamycin resulted in marked increase in GUS histological staining compared with mock solvent control (dimethyl sulfoxide) treated seedlings.

After incubation in liquid medium, seedlings were vacuum infiltrated with GUS-staining solution (100 mM sodium phosphate, pH 7, 10 mM EDTA, 0.1% [w/v] Triton X-100, 0.5 mg mL<sup>-1</sup> X-glucuronide, and 100  $\mu$ g mL<sup>-1</sup> chloramphenicol) either with or without 2 mM potassium ferricyanide and 2 mM potassium ferrocyanide. After staining 2 d at 37°, seedlings were cleared with several changes of 70% (v/v) ethanol. Preliminary experiments indicated that reporter gene expression patterns were not affected by incubation for 12 h in liquid medium (data not shown).

### Insertion Site Analysis

Chromosomal DNA flanking gene trap insertion sites was amplified using TAIL PCR and directly sequenced as described previously (<http://genetraps.cshl.org>). Flanking sequences were used in nucleotide BLAST searches at The Arabidopsis Information Resource (<http://www.Arabidopsis.org>) to determine the approximate chromosomal insertion site in the *Arabidopsis* genome. Insertion sites were confirmed either by supportive sequences from both sides of the insertion site or by PCR using a primer annealing in the putative flanking chromosomal region paired with a primer annealing in the gene trap vector.

Genome annotation at GenBank (National Center for Biotechnology Information) and Munich Information Center for Protein Sequences Arabidopsis database (<http://mips.gsf.de/proj/thal/db/index.html>) was used to putatively assign insertions to regions within an annotated gene, annotated exons not within a complete gene model, or between annotated genes. For annotated gene products, predicted subcellular localization was determined by Psort (<http://psort.nibb.ac.jp/>; Nakai and Kanehisa, 1992), and predicted transmembrane domains were identified using TopPred 2 (<http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html>; von Heijne, 1992). Putative signaling domains were identified using the SMART search tool (<http://smart.embl-heidelberg.de/>; Schultz et al., 2000).

### Molecular Analysis

Protein was isolated by macerating 1-week-old seedlings in extraction buffer (100 mM Tris, pH 7.5, 10% [w/v] Suc, 5 mM EDTA, 40 mM 2-mercaptoethanol, and 2 mM phenylmethylsulfonyl fluoride), followed by centrifugation at 12,000g at 4°C for 15 min to pellet cell debris. Protein was electrophoresed on a 10% (w/v) SDS-PAGE gel, blotted to nitrocellulose membrane, and probed with anti-GUS antibody from Molecular Probes (Eugene, OR).

RNA for RT-PCR was isolated using TRIZOL reagent (Invitrogen) or RNeasy (Qiagen USA, Valencia, CA) and treated with RQ1 DNase (Promega, Madison, WI) using the manufacturers' protocols. Synthesis of cDNA was primed with an anchored poly(T) primer, using Moloney murine leukemia virus RT (Invitrogen). RNA was prepared from equal amounts of seedlings from tunicamycin or mock-treated seedlings, and subjected to semiquantitative RT-PCR at three different cycle numbers empirically determined to include non-saturation levels of amplification (25, 32, and 36 cycles for most transcripts); using the primers TrmGUSR (5'-aaaatcgggcgaattccatctacctg-3') and TrGUSL (5'-cgattacccttacgctgaagaga-3') to quantify GUS transcripts and the primers Tin1 (5'-ttgggtgatgccctgata-3') and Tin 2 (5'-taatttcgaatccaaatc-3') to amplify a control transcript (NM\_118843). Annealing temperatures of 58°C and 50°C were used for the GUS and Tin1 PCRs, respectively, with an extension time of 1 min. 5'-RACE was performed using the SmartRace kit (BD Biosciences Clontech, Palo Alto, CA) according to the manufacturer's protocols. Nested GUS primers are 5'-gtatagccgctgatgctcactact-3' and 5'-tcacgggtgggtttctacaggac-3'. RT-PCR primer 5376\_R was 5'-aaccaactcctctgctgtct-3' (anneals to chromosome 5, TAC clone:K9I9, 39379-39360).

## ACKNOWLEDGMENTS

We thank Rulan Shen and Joe Simorowski (Cold Spring Harbor Laboratory, NY) for assistance in creating, screening, and analyzing gene trap lines, and Bruce May (Cold Spring Harbor Laboratory, NY) analysis of insertion sites. We thank Tim Mulligan of the Cold Spring Harbor Uplands Farm facility for plant care. We thank Alina Rabinovich (US Department of Agriculture Forest Service) for assistance with RT-PCR, and the Harada lab for control RT-PCR primers.

Received January 7, 2003; returned for revision January 24, 2003; accepted February 10, 2003.

## LITERATURE CITED

- Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815
- Aravind L, Koonin E** (2000) The U box is a modified RING finger: a common domain in ubiquitination. *Curr Biol* **10**: R132–R134
- Bar-Peled M, Bassham D, Raikhel N** (1996) Transport of proteins in eukaryotic cells: more questions ahead. *Plant Mol Biol* **32**: 223–249
- Byrne M, Simorowski J, Martienssen R** (2002) *ASYMETRIC LEAVES1* reveals *knox* gene redundancy in Arabidopsis. *Development* **129**: 1957–1965
- Carland F, Berg B, FitzGerald J, Jinamornphongs S, Nelson T, Keth B** (1999) Genetic regulation of vascular tissue patterning in Arabidopsis. *Plant Cell* **11**: 2121–2137
- Cassab G** (1998) Plant cell wall proteins. *Annu Rev Plant Physiol Plant Mol Biol* **49**: 281–309
- Clay N, Nelson T** (2002) VH1, a provascular cell-specific receptor kinase that influences leaf cell patterns in Arabidopsis. *Plant Cell* **14**: 2707–2722
- Cocherel S, Perez P, Degroote F, Genestier S, Picard G** (1996) A promoter identified in the 3' end of the *Ac* transposon can be activated by cis-acting elements in transgenic Arabidopsis lines. *Plant Mol Biol* **30**: 539–551
- Cock J, McCormick S** (2001) A large family of genes that share homology with *CLAVATA3*. *Plant Physiol* **126**: 939–942
- Cutler S, Ehrhardt D, Griffiths J, Somerville C** (2000) Random GFP::cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at high frequency. *Proc Natl Acad Sci USA* **97**: 3718–3723
- Dharmawardhana D, Ellis B, Carlson J** (1992) Characterization of vascular lignification in *Arabidopsis thaliana*. *Can J Bot* **70**: 2238–2244
- Elbein A** (1987) Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains. *Annu Rev Biochem* **56**: 497–534
- Farrell L, Beachy R** (1990) Manipulation of  $\beta$ -glucuronidase for use as a reporter in vacuolar targeting studies. *Plant Mol Biol* **15**: 821–825
- Ferrandiz C, Gu Q, Martienssen R, Yanofsky M** (2000) Redundant regulation of meristem identity and plant architecture by *FRUITFULL*, *APETALA1* and *CAULIFLOWER*. *Development* **127**: 725–734
- Firek S, Whitelam G, Draper J** (1994) Endoplasmic reticulum targeting of active modified  $\beta$ -glucuronidase (GUS) in transgenic tobacco plants. *Transgenic Res* **3**: 326–331
- Goo J, Park A, Park W, Park O** (1999) Selection of Arabidopsis genes encoding secreted and plasma membrane proteins. *Plant Mol Biol* **41**: 415–423
- Groover A, Jones A** (1999) Tracheary element differentiation uses a novel mechanism coordinating programmed cell death and secondary cell wall synthesis. *Plant Physiol* **119**: 375–384
- Hirsch R, Lewis B, Spalding E, Sussman M** (1998) A role for the AKT1 potassium channel in plant nutrition. *Science* **280**: 918–921
- Hirschberg C, Snider M** (1987) Topography of glycosylation in the rough endoplasmic reticulum and Golgi apparatus. *Annu Rev Biochem* **56**: 63–87
- Iturriaga G, Jefferson R, Bevan M** (1989) Endoplasmic reticulum targeting and glycosylation of hybrid proteins in transgenic tobacco. *Plant Cell* **1**: 381–390
- Jenik P, Irish V** (2001) The Arabidopsis floral homeotic gene *APETALA3* differentially regulates intercellular signaling required for petal and stamen development. *Development* **128**: 13–23
- Kristoffersen P, Teichmann T, Stracke R, Palme K** (1996) Signal sequence trap to clone cDNAs encoding secreted or membrane-associated plant proteins. *Anal Biochem* **243**: 127–132
- Maizel A, Bensaude O, Prochiantz A, Joliot A** (1999) A short region of its homeodomain is necessary for engrailed nuclear export and secretion. *Development* **126**: 3183–3190
- Manoil C, Beckwith J** (1986) A genetic approach to analyzing membrane topology. *Science* **233**: 1403–1408
- Martienssen R, Dolan L** (1998) Patterns in vegetative development. *In* M Anderson, J Roberts, eds, *Arabidopsis*. Annual Plant Reviews, Vol 1. Sheffield Academy Press, Sheffield, UK, pp 262–297
- Martienssen R, Irish V** (1999) Copying out our ABCs: the role of gene redundancy in interpreting genetic hierarchies. *Trends Genet* **15**: 435–437
- Nakai K, Kanehisa M** (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**: 897–911
- Pelaz S, Ditta G, Baumann E, Wisman E, Yanofsky M** (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**: 200–203
- Ride J, Davies E, Franklin F, Marshall D** (1999) Analysis of Arabidopsis genome sequence reveals a large new gene family in plants. *Plant Mol Biol* **39**: 927–932
- Sanchez-Fernandez R, Davies T, Coleman J, Rea P** (2001) The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. *J Biol Chem* **276**: 30231–30244
- Schultz J, Copley R, Doerks T, Ponting C, Bork P** (2000) SMART: a Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* **28**: 231–234
- Serafini T, Colamarino S, Leonardo E, Wang H, Beddington R, Skarnes W, Tessier-Lavigne M** (1996) Netrin-1 is required for the commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**: 1001–1014
- Skarnes W, Moss J, Hurtle S, Beddington R** (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc Natl Acad Sci USA* **92**: 6592–6596
- Springer P, McCombie W, Sundaresan V, Martienssen R** (1995) Gene trap transposon tagging of proliferin, an MCM2–3–5 like gene in Arabidopsis. *Science* **268**: 877–880
- Springer P** (2000) Gene traps: tools for plant development and genomics. *Plant Cell* **12**: 1007–1020
- Sundaresan V, Springer P, Volpe T, Haward S, Jones J, Dean C, Ma H, Martienssen R** (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev* **9**: 1797–1810
- Townley D, Avery B, Rosen B, Skarnes W** (1997) Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Res* **7**: 293–298
- Tuckwell D** (1999) Evolution of von Willebrand factor A (VWA) domains. *Biochem Soc Trans* **27**: 835–840
- Vanoosthuysen V, Miede C, Dumas C, Cock M** (2001) Two large *Arabidopsis thaliana* gene families are homologous to the *Brassica* gene superfamily that encodes pollen coat proteins and the male component of the self-incompatibility response. *Plant Mol Biol* **16**: 17–34
- von Heijne G** (1992) Membrane protein structure prediction, hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225**: 487–494
- Zhao C, Johnson B, Kositsup B, Beers E** (2000) Exploiting secondary growth in Arabidopsis: construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. *Plant Physiol* **123**: 1185–1196