The Plant Cell, Vol. 11, 769–779, May 1999, www.plantcell.org © 1999 American Society of Plant Physiologists

RESEARCH ARTICLE

The *irregular xylem3* Locus of Arabidopsis Encodes a Cellulose Synthase Required for Secondary Cell Wall Synthesis

Neil G. Taylor,[a] Wolf-Rüdiger Scheible,[b] Sean Cutler,[b] Chris R. Somerville,[b] and Simon R. Turner[a,1]

[a] University of Manchester, School of Biological Science, Manchester M13 9PT, United Kingdom
[b] Department of Plant Biology, Carnegie Institution of Washington, and Department of Biological Sciences, Stanford University, Stanford, California 94305

The *irregular xylem3* (*irx3*) mutant of Arabidopsis has a severe deficiency in secondary cell wall cellulose deposition that leads to collapsed xylem cells. The *irx3* mutation has been mapped to the top arm of chromosome V near the marker nga106. Expressed sequence tag clone 75G11, which exhibits sequence similarity to cellulose synthase, was found to be tightly linked to *irx3*, and genomic clones containing the gene corresponding to clone 75G11 complemented the *irx3* mutation. Thus, the *IRX3* gene encodes a cellulose synthase component that is specifically required for the synthesis of cellulose in the secondary cell wall. The *irx3* mutant allele contains a stop codon that truncates the gene product by 168 amino acids, suggesting that this allele is null. Furthermore, in contrast to *radial swelling1* (*rsw1)* plants, *irx3* plants show no increase in the accumulation of β-1,4–linked glucose in the noncrystalline cell wall fraction. *IRX3* and *RSW1* fall into a distinct subgroup (Csa) of Arabidopsis genes showing homology to bacterial cellulose synthases.

INTRODUCTION

For many plant cells, the cell wall is synthesized in two distinct stages. During the initial phase of cellular growth, a primary cell wall is laid down and continuously expanded by processes that include relaxation of interchain linkages and addition of new polymers and matrix materials. Cellulose usually comprises ~20 to 30% of the dry weight of the primary wall (Fry, 1988). After the cessation of expansion and division, a secondary cell wall is synthesized within the bounds of the primary wall. This wall is usually much thicker than the primary wall and comprises the bulk of terrestrial biomass. Cellulose accounts for ~40 to 90% of the secondary cell wall, depending upon the cell type. In some heavily thickened cells, such as xylem cells, the secondary wall also may contain a high proportion of lignin that contributes to the mechanical strength.

The mechanisms involved in the synthesis of secondary cell walls are not understood in detail (Emons and Mulder, 1998). It generally is accepted that for both primary and secondary walls, the cellulose component is synthesized by enzyme complexes situated at the plasma membrane. Many freeze–fracture studies have identified plasma membrane particles known as rosettes that appear to be associated with the ends of microfibrils (Brown, 1996). The spacing of these rosettes also correlates with the distribution of the microfibrils (Giddings et al., 1980). It has been suggested that each rosette consists of a hexameric complex, which results in the synthesis of 36 β-glucan chains that are thought to be present in a primary microfibril (Delmer and Amor, 1995). The differences in physical properties of primary and secondary plant cell walls are partly due to differences in the number of individual cellulose chains in the microfibril unit. In contrast to the ~36 individual chains in primary microfibrils (Delmer and Amor, 1995), the secondary cell walls of some algae contain fibrils containing up to 12,000 individual β-1,4-glucan chains (Brown et al., 1996). In addition, individual cellulose chains from the secondary wall typically contain ~14,000 β-1,4–linked glucose molecules, whereas in the primary wall approximately half of the cellulose molecules contain less than ~500 glucose moieties, and half contain ~2500 to 4500 monomers (Blaschek et al., 1982).

Although the cellulose synthase complex from higher plants has not been characterized at the molecular level, it is widely assumed that it will prove to be a multienzyme complex (Delmer and Amor, 1995). Consistent with this concept, a four-gene operon, which is responsible for cellulose synthesis, has been cloned from *Acetobacter xylinum* (Saxena et al., 1990), and five genes have been shown to be essential for cellulose synthesis in Agrobacterium (Matthysse et al., 1995). Only one of these genes shows sequence similarity between Agrobacterium and *A. xylinum*, and this gene has

[1] To whom correspondence should be addressed. E-mail simon.turner@man.ac.uk; fax 44-161-275-3938.

been identified as encoding the cellulose synthase catalytic subunit. Amino acid sequences of bacterial cellulose synthases, along with other enzymes requiring nucleotide sugars, were found to contain four regions of high conservation thought to be critical for UDP–glucose binding and catalysis (Saxena et al., 1995).

Recently, cDNA clones for two cellulose synthase homologs containing all four conserved regions were identified from a cotton cDNA library prepared from fibers at the onset of secondary cell wall synthesis (Pear et al., 1996). These genes, which are termed *CELA* genes, exhibit sequence similarity to at least 31 distinct expressed sequence tag (EST) or genomic sequences in the Arabidopsis sequence databases (Cutler and Somerville, 1997). However, it is unlikely that all of these cellulose synthase–like (*CSL*) genes actually catalyze cellulose synthesis (Cutler and Somerville, 1997; Delmer, 1998). Rather, it has been proposed that some of the *CSL* genes encode other glycan synthases, such as those responsible for the synthesis of xyloglucan, xylan, callose, and other polysaccharides. The biological function of one of the *CELA*-related genes was established recently by the characterization of a mutant of Arabidopsis deficient in crystalline cellulose deposition. The *radial swelling1* (*rsw1*) mutant exhibits temperature-sensitive radial swelling of its root tip due to a deficiency in cellulose deposition at elevated temperature (Baskin et al., 1992). The *RSW1* gene encodes a polypeptide with a high degree of sequence similarity to the cotton *CELA* genes (Arioli et al., 1998a).

Mutants of Arabidopsis carrying mutations in one of the three *irx* (for *irregular xylem*) loci are characterized by collapsed xylem in stems (Turner and Somerville, 1997). The xylem vessels are thought to collapse due to a lack of resistance to the negative pressure exerted by water transport. The deposition of cell walls in these plants is abnormal and results in the stems being weaker and less rigid. In one of these mutants, *irx3*, the increased flexibility of the stems results in an inability to support an upright growth habit. Analysis of these mutants showed a specific reduction or complete loss of cellulose deposition in the secondary cell wall (Turner and Somerville, 1997). In this study, we describe the isolation and characterization of a member of the Arabidopsis *CELA* gene family that corresponds to the *IRX3* gene. The discovery that *IRX3* is a component of the cellulose synthases involved in secondary wall synthesis provides novel insight into the function of the *CELA* gene family and creates several novel experimental opportunities for future studies of the factors that regulate secondary wall synthesis.

RESULTS

Identification of a Cellulose Synthase EST Linked to *irx3*

Because of the specific defect in secondary wall cellulose deposition in the *irx3* mutant, we tested the possibility that one of the *CSL* or *CELA* sequences present in the Arabidopsis database corresponded to the *irx3* locus. *irx3* maps to the middle of chromosome V and is close to the marker nga106 (Turner and Somerville, 1997). In a cross between the *irx3* mutant and the wild type, no recombinants were observed between *irx3* and nga106 in an analysis of 200 F_2 mutants (data not shown). Figure 1 shows that *irx3* is placed between markers nga151 and R89998. This region is represented by the seven CIC yeast artificial chromosome (YAC) clones CIC8E12, CIC9H7, CIC9F1, CIC6H3, CIC9E10, CIC11C4, and CIC6B10 (Creusot et al., 1995; Schmidt et al., 1997; http://genome-www.stanford.edu.Arabidopsis/JIC-contigsChr5_YACcontig6.GIF; Figure 1). Consequently, the *irx3* gene must be contained on one of these YACs.

Polymerase chain reaction (PCR) primer pairs were designed for each of the individual Arabidopsis *CELA* and *CSL* genes in GenBank, and each primer pair was tested to determine whether it amplified a fragment from the YAC clones spanning the region containing *irx3.* Only one of these primer pairs (75G11F and 75G11R) amplified a product corresponding to the EST clone 75G11, amplifying a 200-bp fragment (data not shown). Analysis of the individual YACs in the region demonstrated that the *75G11* gene is contained on YACs CIC9H7, CIC9F1, and CIC6H3 but not on YACs CIC8E12, CIC11C4, CIC6B10, and CIC9E10 (Figure 1). Based on the estimated relationship between physical and genetic map distance (Schmidt et al., 1997), this information localized EST clone 75G11 to an ∼150-kb region between markers nga106 and mi438 (Figure 1). Because the *irx3* mutation also maps between these two markers (results not presented), this information placed the EST *75G11* gene on a region of the chromosome that was tightly linked to *irx3.*

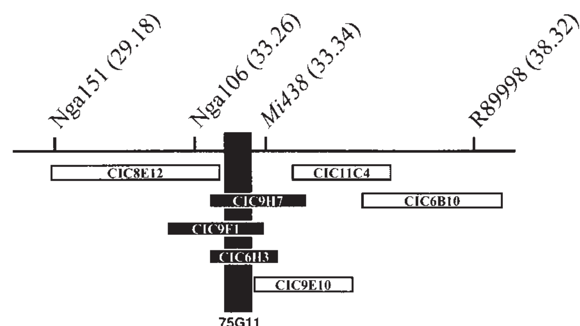

Figure 1. Localization of the *irx3* Mutation on Chromosome V.

The positions of YAC clones spanning this region are shown (from Schmidt et al., 1997). The YAC clones containing the *IRX3* gene are in filled rectangles. The filled vertical bar indicates the region of chromosome V containing the *IRX3* gene. The positions of genetic markers are taken from the map generated from recombinant inbred lines (Lister and Dean, 1993).

Isolation of Genomic Clones Corresponding to EST 75G11

To obtain the full-length sequence of the gene corresponding to EST 75G11, we used the EST clone as a hybridization probe to isolate genomic clones. A Landsberg *erecta* genomic library was screened and yielded two clones that were retained for characterization. Figure 2A shows that one of these clones (pCS1) contains a HindIII fragment of 7.5 kb that was found to encode the entire coding sequence of the gene corresponding to EST 75G11. The nucleotide sequence of this fragment and the deduced amino acid sequence of the gene product have GenBank accession number AF091713. The cDNA sequence of the gene corresponding to EST 75G11 was determined by reverse transcription–PCR (RT-PCR). To determine the sequence, we used primer pairs corresponding to the presumptive coding sequence, designed to amplify both the 3′ and 5′ halves of the gene, to amplify first-strand cDNA. The fragments were cloned before sequencing. To negate the possible effects of incorporation of incorrect nucleotides by Taq polymerase, we sequenced two independent clones isolated from individual RT-PCR reactions and found them to be identical (GenBank accession number AF088917).

Comparison of the cDNA and genomic sequences identified the presence of 11 introns and 12 exons in the genomic sequence. The cDNA sequence encodes a predicted protein of 1025 amino acids with a molecular mass of 116 kD. Figure 3 shows there is a high degree of sequence similarity between the *75G11* gene product and several other cellulose synthase gene products, notably those of the Arabidopsis *RSW1* and *Ath-A* genes (Arioli et al., 1998a) and the cotton *CELA1* gene (Pear et al., 1996). It is clear that there are significant regions of very high conservation. The only areas with no notable similarity are in a region (VR2) that has been described previously as a plant hypervariable region (Pear et al., 1996) and a region close to the N terminus (VR1; Figure 3). In common with other cellulose synthase genes that have been identified (Pear et al., 1996; Arioli et al., 1998a), the *75G11* gene product contains a cysteine-rich region at its N terminus, which has been suggested to form a LIM-like zinc finger motif that may be involved in protein–protein interactions (Delmer, 1998). As expected, the *75G11* gene product also contains the four motifs that have been identified as being conserved in cellulose synthase genes. The first three of these are centered around aspartate residues, and the fourth consists of a QxxRW motif (where x represents any amino acid), which in this case as in several other cases contains the sequence QVLRW (Figure 3).

In common with cotton *CELA* and Arabidopsis *RSW1* (Pear et al., 1996; Arioli et al., 1998a), the *75G11* gene product shares a predicted transmembrane topology consisting of two transmembrane domains at the N terminus followed by a cytoplasmic central domain containing the four conserved motifs described. Six putative transmembrane segments at the C terminus follow this domain (Figure 3).

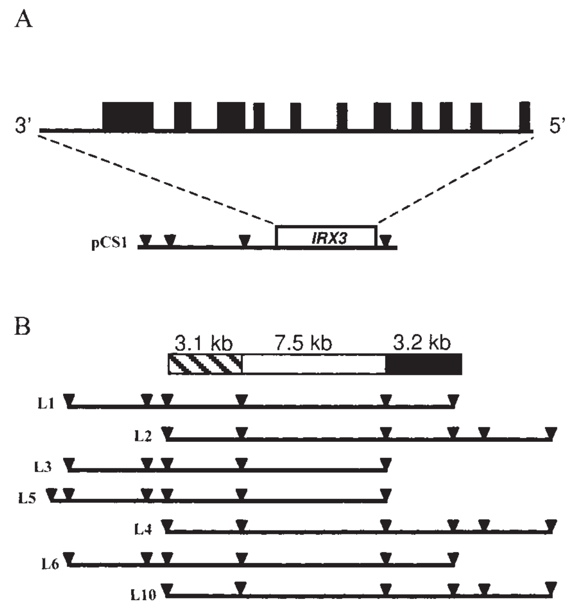

Figure 2. Map of Genomic Clones Containing the *IRX3* Gene.

Introns are represented by solid blocks, and triangles indicate the position of HindIII sites. Boxes represent the positions of the 3.1-kb (hatched), 7.5-kb (open), and 3.2-kb (filled) HindIII fragments referred to in the text. Two additional HindIII sites (not shown) occur between the 7.5- and 3.2-kb HindIII fragments.

(A) λ clone used to subclone the *IRX3* gene and intron/exon map of the *IRX3* gene.

(B) Cosmid clones used for complementation.

Isolation of a Mutant Allele of *irx3*

To test the hypothesis that the *75G11* and *IRX3* genes are identical, we determined the sequence of the *75G11* gene in the *irx3* mutant. RT-PCR was used to isolate cDNA clones of the mutant allele. The cDNA was amplified in two halves, with two independent reactions conducted to control for the possibility of nucleotide misincorporation by Taq polymerase. Both clones showed a G-to-A nucleotide substitution, which resulted in the introduction of a stop codon in place of Trp-859. The region of genomic DNA containing this mutation was amplified by PCR and two independent products sequenced to confirm the presence of this mutation. Both products contained the G-to-A nucleotide substitution. This mutation causes premature termination of translation immediately after the second of the six C-terminal putative transmembrane domains and results in a protein lacking 168 C-terminal amino acids. The identification of a mutation in the *75G11* gene in the *irx3* mutant strongly suggested that *75G11* is identical to *IRX3*.

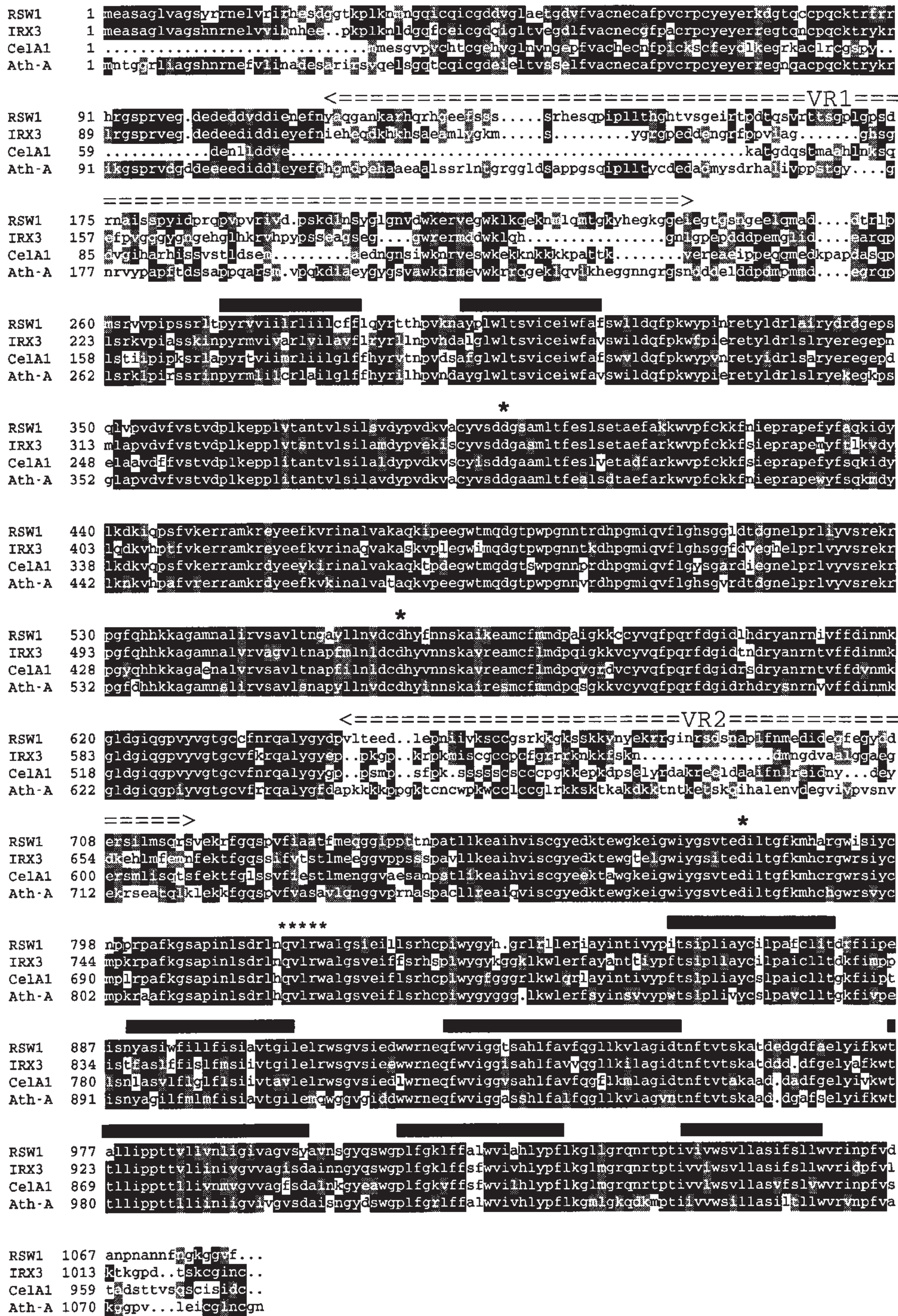

Figure 3. Alignment of the Amino Acid Sequences of Plant Cellulose Synthase Genes.

Solid boxes indicate regions in which more than half the residues are identical, and gray boxes indicate conserved residues. The positions of the three aspartic acid (d) residues and QxxRW motifs are indicated by large and small asterisks, respectively. Positions of the presumed membrane-spanning helices are indicated by solid bars. Variable regions referred to in the text also are indicated (VR1 and VR2). Dots were introduced to optimize alignment.

Complementation of *irx3* with the Wild-Type Gene

To test whether the *irx3* mutation could be complemented with the wild-type gene, we isolated several cosmid clones containing the *75G11* gene and used them to transform *irx3* plants. All of the cosmids contained a 7.5-kb HindIII fragment identified as carrying the coding region of the gene in its entirety (Figure 2B). In addition, the clone contained 90 bp of sequence at the 5′ end and 2603 bp at the 3′ end of the gene.

Figures 4 and 5 show that cosmids L1, L4, and L10 (as well as L2, L6, and L8; data not shown) complemented the *irx3* mutation. Each of these contained the 7.5-kb HindIII fragment, an adjacent 3.2-kb HindIII fragment at the 5′ end, and a 3.1-kb HindIII fragment at the 3′ end of the *IRX3* gene (Figure 2B). The 3.1-kb fragment carries no part of the *IRX3* coding region, and the nucleotide sequence of this fragment had no significant sequence similarity to any known genes as determined by BLASTX searches (Altschul et al., 1990) against the SwissProt database. It can be seen from the transverse stem sections stained with toluidine blue that in *irx3* plants, there is considerable collapse of the xylem vessels, whereas wild-type plants have clear, open xylem vessels (Figure 4). In plants transformed with cosmids L1 and L10, this collapse is not evident, and these plants have xylem elements that are visually indistinguishable from those of the wild type. Cosmids L3 and L5, which did not carry the 3.2-kb fragment, failed to complement the mutation (Figure 4). In all plants transformed with L3, the xylem vessels exhibit the collapsed phenotype evident in the mutant, whereas in some of the plants transformed with cosmid L5, there was partial complementation of the mutant phenotype (Figure 4). This suggests that the requirement for the 3.2-kb 5′ HindIII fragment is not absolute. The presence of this fragment is presumably necessary to direct correct expression of the gene. Because the 7.5-kb fragment carries only 90 nucleotides upstream of the coding sequence of the gene, the 3.2-kb fragment presumably contains the promoter required for normal correct expression of the gene. These promoter sequences are presumably found in the first 1.5 kb of this fragment, because the 5′ end of this fragment appears to encode part of a gene that exhibits weak similarity (BLASTX score 68, smallest sum probability 2×10^{-33}) to an APETALA2 domain–containing protein.

Measurements of the cellulose content of the primary transgenic plants (Figure 5) confirmed the results from qualitative

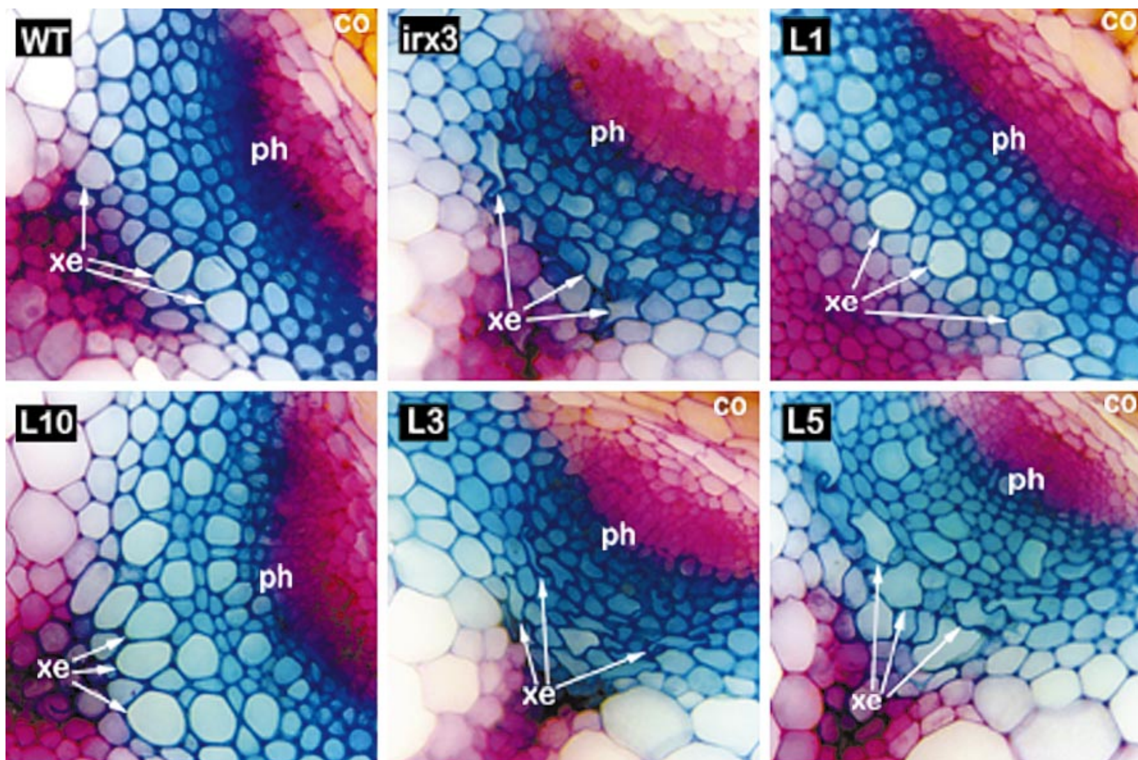

Figure 4. Toluidine Blue–Stained Sections of Arabidopsis Vascular Bundles from Wild-Type, *irx3*, and *irx3* Plants Transformed with Cosmids L1, L10, L3, and L5.

co, cortex; ph, phloem; WT, wild type; xe, xylem elements.

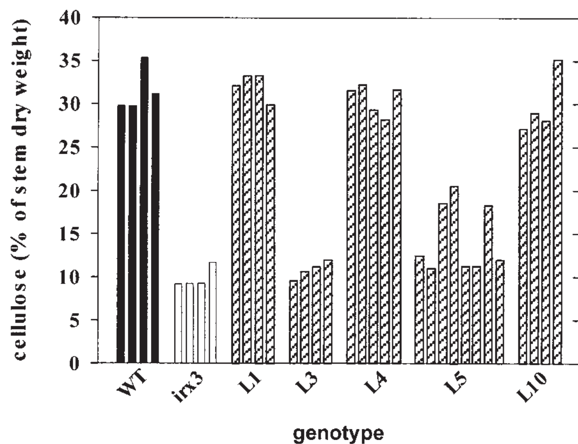

Figure 5. Cellulose Measurements Showing Complementation of the *irx3* Cellulose-Deficient Phenotype by Using Cosmid Clones.

Cellulose content of stem sections from individual wild-type (WT) and *irx3* plants together with individual *irx3* plants transformed with cosmids (L1, L3, L4, L5, and L10) containing the *IRX3* gene was determined. Details of the cosmids are provided in Figure 2.

analyses of xylem sections. Plants transformed with the cosmids L1, L4, and L10 contained cellulose levels that were indistinguishable from the wild type, whereas cosmid L3 had no effect on cellulose content. Thus, only cosmids that contained the 3.2-kb HindIII fragment effectively complemented the *irx3* mutation. Cosmids lacking this fragment (L3 and L5) did not complement or only partially complemented the mutation.

Expression Patterns of the *IRX3* Gene

RNA was isolated from leaves and from four discrete stem sections—the tip, upper middle part, lower middle part, and base of the stem—of mature wild-type and *irx3* plants. Figure 6 shows the results of probing this RNA with EST 75G11. In the wild type, there was an increase in the amount of *IRX3* mRNA as the stem matured (i.e., toward the base of the stem). There was no detectable transcript in leaves. These expression patterns correspond with secondary cell wall development. In comparison with the wild type, *IRX3* transcript levels were severely decreased in the *irx3* mutant, to ~10% of wild-type levels in the most mature stem tissue (Figure 6). An identical blot probed with a gene encoding for caffeic acid *O*-methyltransferase (COMT), which is a component of the lignin biosynthesis pathway, showed that the *irx3* mutation had little effect on the expression of a typical gene in the lignin biosynthetic pathway (Figure 6). Minor differences in COMT transcript levels are thought to be due to the difficulty in accurately staging the sections obtained from the different plants, because *irx3* plants have been shown to grow slightly more slowly than the wild type (Turner and Somerville, 1997). Two possibilities exist regarding the residual signal seen in *irx3* plants. It has been shown previously that the introduction of a premature stop codon into a transcript (as is the case with *irx3*) can lead to its degradation (Abler and Green, 1996). Thus, it would not be surprising if the message levels in *irx3* plants are reduced. It is not inconceivable that due to the close relationship between *CelA*-like genes, there is some cross-hybridization with another member of the family, but the fact that the message level is decreased 90% in *irx3* shows that the large majority of the signal seen is derived from the correct message.

IRX3 Is Part of a Large Family of Plant Cellulose Synthase Homologs

Analysis of current genomic sequence data indicates that Arabidopsis contains nine anonymous open reading frames with significant similarity to IRX3. Three other homologs previously have been described (Arioli et al., 1998a). Thus, 13 Arabidopsis genes with significant similarity to *IRX3* are present in public databases. Because only ~30% of the Arabidopsis genome sequence is available, the size of this gene family is likely to be much larger. Proteins that share a common ancestor often share similar biochemical functions; understanding the evolutionary history of this gene family may help in future predictions of gene function.

To infer the evolutionary history of this gene family, we constructed a multiple alignment of plant and bacterial sequences similar to known cellulose synthases. The alignment data were bootstrap resampled and used to generate a maximum parsimony tree using the PROTPARS algorithm (Felsenstein, 1993). The phylogenetic tree generated was rooted using a cellulose synthase homolog identified in the deeply branching prokaryote *Aquifex aeolicus* (Deckert et al., 1998). Figure 7 shows the consensus tree generated by this analysis.

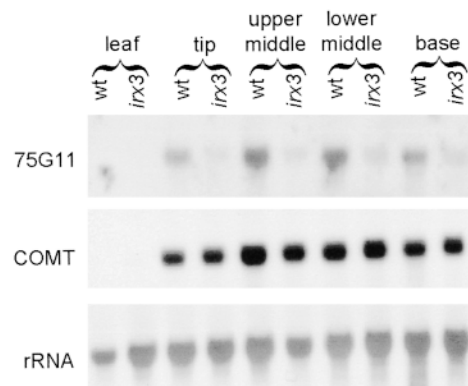

Figure 6. RNA Gel Blots Showing Expression of the *IRX3* Gene.

Blots containing RNA from developing stems and leaves from wild-type (wt) and *irx3* plants were probed with 75G11, COMT, and rRNA.

The phylogenetic tree reveals three deep branches, which divide the plant genes into three subfamilies. These branches are supported by high bootstrap values and are unlikely to be spurious. Based on these data, we suggest that the higher plant family of sequences similar to IRX3 can be broken into three subfamilies. To conform with Arabidopsis genetic nomenclature, we suggest these families be called CSA, CSB, and CSC (Figure 7). We intend for the CS prefix to indicate "cellulose synthase homolog."

The *CSA* gene family includes *RSW1*, *IRX3*, *CELA1*, and *CELA2.* These genes are likely to be cellulose synthases based on either mutational analysis or expression data. Thus, the known plant cellulose synthases form a distinct subfamily within the gene family as a whole and is not distributed throughout the family. The functions of the other branches remain to be determined. However, we believe they could function in the synthesis of one of many plant β-linked polysaccharides (Cutler and Somerville, 1997).

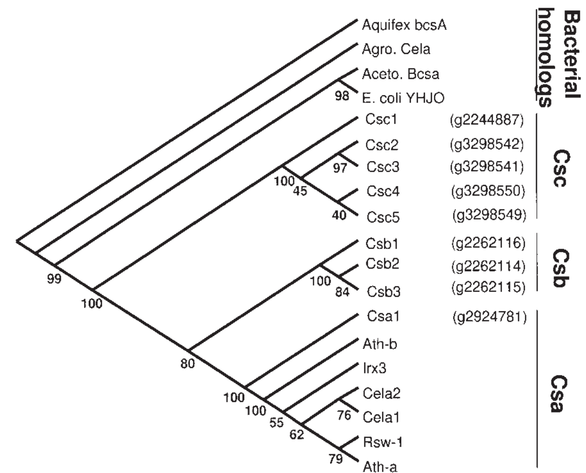

Figure 7. Phylogenetic Tree of Bacterial and Plant Cellulose Synthases and Homologs.

Alignment data are from bootstrap values sampled 100 times and used to construct the consensus tree shown. Numbers are bootstrap values and indicate the number of trees in which the proteins to the right of a bootstrap value clustered together. Shown to the right of the Csa, Csb, and Csc protein names are the GenBank accession numbers for the genes encoding each protein. Aceto, *A. xylinum*; Agro., *A. tumefaciens*; Aquifex, *A. aeolicus.*

DISCUSSION

Stems of the *irx3* mutant contain ~20 to 30% of the amount of cellulose in mature stem tissue of the wild type (Turner and Somerville, 1997). This results in an alteration of the physical properties of the stem and also leads to collapse of the xylem vessels due to an inability to withstand the negative pressure generated by water transport (Turner and Somerville, 1997).

Because of the specific defect in cellulose deposition in the mutant, we hypothesized that the *irx3* mutation may cause a defect in a subunit of cellulose synthase. To test this hypothesis, we first identified all of the EST and genomic sequences with sequence similarity to the Arabidopsis *CSL* genes and the *CELA* genes from cotton that were present in public databases. We then tested whether each of these sequences was present on the seven YAC clones that span the region of the genome where the *irx3* mutation had been genetically mapped. One EST (75G11) was found to be present on three of the relevant YACs and was therefore deemed a candidate clone for the *IRX3* gene. The observation that the *75G11* gene carries a nonsense mutation in the *irx3* background and complementation of the *irx3* mutation with cosmids carrying 75G11 confirmed the coidentity of *75G11* and *IRX3.*

IRX3 likely encodes a cellulose synthase catalytic subunit similar to other plant and bacterial cellulose synthase genes (Pear et al., 1996; Arioli et al., 1998a). It contains all of the conserved motifs that have been proposed to be essential for cellulose synthase activity (Pear et al., 1996; Arioli et al., 1998a). Moreover, the expression pattern of the *IRX3* gene in Arabidopsis is consistent with the expectation for a gene involved in the synthesis of cellulose to be deposited in heavily thickened secondary cell walls. The increased level of accumulation of *IRX3* mRNA in more mature stem tissue is consistent with the observation that the cellulose content increases toward the base of the stem. This expression pattern of the *IRX3* gene also correlates well with the *irx3*-conferred phenotype, which exhibits a large difference in cellulose content in mature stems compared with the wild type, but little difference in leaves (Turner and Somerville, 1997).

Further evidence that *IRX3* is not involved in cellulose synthesis in primary walls derives from observations that *IRX3* does not exhibit any of the radial swelling or other phenotypes characteristic of the *rsw1* mutant, despite the very severe nature of the *irx3* mutation, which suggests that it is probably a null mutation. In addition, whereas *rsw1* mutant plants exhibit a decrease in crystalline cellulose, there is an increase in noncrystalline β-1,4–linked glucose (Arioli et al., 1988a). *irx3* plants apparently show no increase in this noncrystalline β-1,4–linked glucose because, despite the very large decrease in crystalline cellulose observed in *irx3*, no increase has been observed in the proportion of glucose in the noncrystalline (soluble in 2 M sulphuric acid) cell wall fraction (Turner and Somerville, 1997). Until the definitive confirmation that recombinant proteins produced from these genes actually have cellulose synthase activity, it is still possible that these genes may encode, for example, a protein that primes rather than extends the cellulose chain. The work presented here, however, adds to the growing body of evidence (Pear et al., 1996; Arioli et al., 1998a) that these genes

do in fact encode for the catalytic subunit of the higher plant cellulose synthase complex.

The relatively large number of *CSL* sequences from Arabidopsis that are present in public databases have raised questions regarding the function of these sequences (Cutler and Somerville, 1997). The results presented here indicate that the function of at least some of the genes may be accounted for by cell type–specific gene expression. Similarly, in the *rsw1* mutant, epidermal cells are misshapen (Arioli et al., 1998a), and it is possible that only this cell type is affected. It has been suggested that of the ~40 cell types present in plants, almost all can be identified by unique features of their cell walls (Carpita and Vergara, 1998). In light of this, it may not be surprising that different cell types may use individual sets of genes for their cell wall synthesis.

The inferred phylogenetic relationship between the cellulose synthase genes aligned in Figure 3 and some genes that have been suggested to be more weakly related (Arioli et al., 1998b) is shown in Figure 7. It is clear that *IRX3* belongs to a small subfamily of cellulose synthase genes, including *RSW1* and cotton *CELA1*, but shows distant relationships to a large number of other cellulose synthase–related genes. This supports the idea that only the CSA subfamily of genes is involved in cellulose synthesis, whereas the function of other cellulose synthase–related genes remains unknown (Arioli et al., 1998a, 1998b). It can be seen that *IRX3* is closely related to *Ath-B*, an Arabidopsis cDNA of unknown function isolated by screening a cDNA library with a portion of the *RSW1* transcript (Arioli et al., 1998a), and to a gene, which we have provisionally named *CSA1*, that is evident in the currently available Arabidopsis genomic DNA sequence. *IRX3* also appears to be more closely related to the *CELA1* and *CELA2* genes from cotton (Pear et al., 1996) than it does to the Arabidopsis *RSW1* gene (Arioli et al., 1998a), based upon the results of PILEUP analysis (data not shown). Thus, it seems possible that *IRX3*, *CELA*, *CSA1*, and *Ath-B* are all involved in secondary wall synthesis, whereas *RSW1* and *Ath-A* define the class of enzymes involved in primary wall synthesis.

Comparison of these sequences may make it possible to identify features that determine which type of cell wall is produced by a particular cellulose synthase. Do cellulose synthases involved in secondary cell wall synthesis contain some sequences that allow them to form rosette structures, which cluster to produce larger cellulose microfibrils? It is clear that there are two regions of variability between plant cellulose synthase genes. One of these lies close to the N-terminal region that is predicted to be cytoplasmic and also contains a putative cysteine-rich LIM-like protein binding domain (Delmer, 1998). We speculate that this region of the protein is involved in interactions with other proteins that may make up the enzyme complex found in the membrane and possibly with other regulatory proteins as well. It should be noted, however, that there is another region of variability that has been called a hypervariable region (Pear et al., 1996). This region lies between the second and third conserved motifs and as such could be involved in the catalytic process itself. It is clear that there is still much to be learned about the synthesis of cellulose, with many important questions to be answered concerning the number of genes actually encoding cellulose synthases and their possible differences in laying down cellulose. The catalytic mode of action of cellulose synthase is also an area in which advances need to be made to further our understanding. The cloning of *IRX3*, a gene involved in the synthesis of cellulose in secondary cell walls, will allow us to investigate some of these matters. For instance, it will be instructive to test whether *RSW1* or any of the other *CELA*-like genes will functionally complement the *irx3* mutation.

The mutation in the *irx3* mutant leads to the loss of the last 168 amino acids of the mature protein. This portion contains four membrane-spanning domains and several other features conserved in *RSW1* and *CELA1*. It is very unlikely that such a gene product would retain catalytic function; therefore, the *irx3* mutation appears to be a null mutation. In support of this conclusion, electron microscopy of sections of stems from *irx3* plants shows little if any cellulose in the secondary cell wall of xylem cells (Turner and Somerville, 1997). Nevertheless, under laboratory conditions, *irx3* plants can grow and produce relatively normal plants in the absence of a normal secondary cell wall. Thus, it should be possible to recover any mutation that inactivates the cellulose synthase specifically required for secondary wall synthesis. However, if the same genes are used for components of both the primary and secondary walls, it may not be possible to identify nonconditional mutations in these genes. In this respect, the characterization of the *irx1* and *irx2* mutations (Turner and Somerville, 1997) may provide additional insights into the process of cellulose synthesis and deposition.

The identification of the *IRX3* gene was facilitated greatly by analysis of publicly available sequence data. In the near future, this sequencing initiative is likely to be an area of plant research that will revolutionize the way in which gene functions are assigned. The only other report involving the cloning of a cellulose synthase gene from Arabidopsis involved a long chromosome walk to the gene (Arioli et al., 1998a). The increasing number of ESTs that are easily mapped using PCR-based methodology and the completion of Arabidopsis genome sequencing should soon supersede the need for such chromosome walks and will greatly accelerate the identification of genes responsible for mutations.

METHODS

Library Screening

Standard molecular techniques were performed as described by Sambrook et al. (1989). An *Arabidopsis thaliana* Landsberg *erecta* library constructed in λ FIX (Voytas et al., 1990) was screened with a

1.4-kb SalI-XbaI fragment from expressed sequence tag (EST) clone 75G11, labeled nonradioactively with the Gene Images random prime labeling module (Amersham, Little Chalfont, Buckinghamshire, UK), probed, and developed with the Gene Images CDP-Star detection module (Amersham), according to the manufacturer's instructions, before visualization of signal on BioMax MR1 film (Eastman Kodak, Rochester, NY). Two rounds of screening were conducted to identify hybridizing clones.

Cosmids carrying *IRREGULAR XYLEM3* (*IRX3*) were isolated from a Landsberg *erecta* library constructed in pBIC20 (Meyer et al., 1994). Filters carrying 120,000 library clones were hybridized with a random primed digoxigenin-11-2′-deoxyuridine-5′-phosphate–labeled 200-bp polymerase chain reaction (PCR) fragment, amplified by using primers 75G11F and 75G11R (see Results), and developed, and the positive clones were detected colorimetrically as described by the kit manufacturer (Boehringer Mannheim, Germany). Two rounds of screening were conducted to identify cosmid clones harboring *75G11* genomic DNA.

RNA Gel Blot Analysis

Total RNA was isolated from 6-week-old plants by using an RNeasy Plant Mini Kit (Qiagen GmbH, Hilden, Germany). After transfer of 5 μg of electrophoresed RNA to Hybond N+ membranes (Amersham), they were probed with 75G11 (1.4-kb SalI-XbaI fragment), caffeic acid *O*-methyltransferase (COMT; Arabidopsis Biological Resource Center, Columbus, OH; stock center clone 115N5, EcoRI-HindIII 1.5-kb fragment), or rRNA (O'Donnell et al., 1998; 300-bp EcoRI fragment) probes labeled as given above, and developed according to the manufacturer's instructions, before visualization as above.

PCR and Reverse Transcription–PCR

PCR was performed using Taq polymerase (Immunogen International, Sunderland, UK) according to the manufacturer's recommendations in a PTC100 thermal cycler (MJ Research Inc., Watertown, MA). Yeast artificial chromosome (YAC) template DNA was isolated using an IGi Yeast Y1-3 kit (Immunogen International). Oligonucleotide primers were synthesized either by Gibco BRL Life Technologies UK Ltd. (Paisley, UK) or MWG Biotech UK Ltd. (Milton Keynes, UK). Primer sequences for the PCR of 75G11 from YAC clones are as follows: 75G11F, 5′-AAGGTGATAAGGAGCATTTGA-3′; and 75G11R, 5′-TCCCCACTCAGTCTTGTCTT-3′. The PCR conditions were as follows: 94°C for 60 sec followed by 10 cycles of 94°C for 45 sec, 65°C for 60 sec (reducing by 0.5°C per cycle), and 72°C for 60 sec, followed by 25 cycles at 94°C for 45 sec, 55°C for 60 sec, and 72°C for 60 sec, followed by 5 min at 72°C.

For reverse transcription–PCR (RT-PCR), first-strand cDNA was synthesized using 500 ng of mature stem total RNA in a reaction with a Ready To Go RT-PCR Bead (Pharmacia Biotechnology, Uppsala, Sweden) with 500 ng of poly(dT) primer at 42°C for 60 min. Gene-specific primers IRX3F (5′-CCTATGGAAGCTAGCGCCGGTCTT-3′) and IRX312 (5′-GTG-TTTCTGTTGGCGTAACGA-3′) were added for the 5′ end of the cDNA, and IRX3R (5′-GCTTCAGCAGTTGATGCC-ACACTT-3′) and IRX315 (5′-CGTTGAAAGTTGATTATCTCC-3′) were added for the 3′ end. PCR conditions were as follows: 95°C for 5 min followed by 30 cycles at 94°C for 60 sec, 55°C for 60 sec, and 72°C for 2 min. RT-PCR products were gel purified before cloning into the vector pGEM-T Easy (Promega) for sequencing.

For PCR amplification from plant genomic DNA to ensure presence of the A-to-G nucleotide substitution, DNA was prepared from leaf tissue by using a Phytopure plant DNA extraction kit (Scotlab, Lanarkshire, UK). Primers IRX33 (5′-TGCCTGCAACAACGCCAA-CAA-3′) and IRX317 (5′-TTGGGCACTTGGATCGGTTGA-3′) were used to amplify this fragment under the following conditions: 94°C for 60 sec followed by 30 cycles at 94°C for 60 sec, 55°C for 60 sec, and 72°C for 60 sec. Again, the products were gel purified and cloned into pGEM-T Easy for sequencing.

DNA Sequencing

Templates were generated by restriction fragment cloning or exonuclease III–generated deletions and primed with oligonucleotides annealing to either universal priming sites or gene-specific regions. Sequencing primers were synthesized and HPLC or high-purity salt-free purified by MWG Biotech or Applied Biosystems. Plasmid templates were prepared using a Qiagen QIAprep spin miniprep kit and sequenced automatically using ABI PRISM Big Dye Terminators (Applied Biosystems, Foster City, CA). DNA sequence was analyzed using the Genetics Computer Group suite of programs (program manual for the Wisconsin Package, version 8; Genetics Computer Group, Madison, WI) and programs available for use on the Internet.

Complementation of *irx*3

irx3 mutant plants were transformed by *Agrobacterium tumefaciens* (GV3101) carrying the appropriate Landsberg *erecta* binary cosmids according to Bent and Clough (1998). Primary transformants (T_1) were selected by plating sterilized T_1 seeds on Murashige and Skoog 0.8% agar plates containing 50 μg/mL kanamycin sulfate. After 3 weeks, the kanamycin-resistant plants were transplanted into pots containing a commercial soil/peat/perlite mixture. Stems from mature T_1 plants together with stems from same-aged Landsberg *erecta* wild-type and *irx3* mutant plants were sectioned and stained with toluidine blue, and the cellulose content was then measured as described previously (Turner and Somerville, 1997).

Phylogenetic Analysis

Trees were built using PROTPARS, a maximum parsimony algorithm included in the PHYLIP version 3.5 software package (Felsenstein, 1993). Robustness of tree topology was estimated using 100 bootstrapped data sets (Felsenstein, 1985). These are generated by randomly sampling input alignment data until a new data set equivalent in size to the original is generated. Topologies observed in a large percentage of trees are believed to be robust (i.e., supported by multiple characters in the alignment data).

Sequences used for alignments were identified by BLAST searches of GenBank. Several ESTs with significant similarity to *IRX3* were excluded from our alignments. ESTs typically represent a small fraction of coding sequence; consequently, we felt they did not possess enough useful (or reliable) sequence information to warrant inclusion in our data set.

Alignments were made using CLUSTALW (Thompson et al., 1994). Initially, CLUSTALW failed to align the bacterial domain B residues (as defined by Saxena et al., 1995) with the plant domain B residues. This presumably was due to the large insertion present within the

plant domain B block. This problem was rectified by manually aligning the bacterial and plant domain B sequences by inserting gaps into the bacterial sequences. This alignment was refined with a second CLUSTALW alignment. Trees made with the initial and refined alignment data sets were largely in agreement; both identified three deep branches separating the *CSA*, *CSB*, and *CSC* gene families (data not shown). Not all residues of the alignment were used to build the tree shown in Figure 7; only sequence blocks conserved among the majority of sequences were used. These blocks include domains A and B and other conserved regions visible in our alignment. For the IRX3 sequence, the following sequence blocks were used: 320 to 359, 376 to 390, 497 to 512, 518 to 574, 581 to 606, 715 to 744, 750 to 781, 784 to 868, 883 to 907, and 924 to 980.

ACKNOWLEDGMENTS

We are grateful to Erwin Grill for cosmid libraries and the Arabidopsis Biological Resource Center at Ohio State University (Columbus, OH) for providing EST clones. This work was supported in part by grants from the U.S. Department of Energy (No. DOE-FG02-97ER20133) and the Biology and Biotechnology Research Council (No. 34/P03060). W.-R.S. is the recipient of a fellowship (Sche 548 1/1) from the Deutsche Forschungsgemeinschaft.

Received October 14, 1998; accepted February 3, 1999.

REFERENCES

Abler, M.L., and Green, P.J. (1996). Control of mRNA stability in higher plants. Plant Mol. Biol. **32,** 63–78.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment sequence tool. J. Mol. Biol. **215,** 403–410.

Arioli, T., Peng, L., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Hofte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J., and Williamson, R.E. (1998a). Molecular analysis of cellulose biosynthesis in *Arabidopsis.* Science **279,** 717–720.

Arioli, T., Burn, J.E., Betzner, A.S., and Williamson, R.E. (1998b). Response: How many cellulose synthase–like gene products actually make cellulose? Trends Plant Sci. **3,** 165–166.

Baskin, T.I., Betzner, A.S., Hoggart, R., Cork, A., and Williamson, R.E. (1992). Root morphology mutants in *Arabidopsis thaliana.* Aust. J. Plant Physiol. **19,** 427–437.

Bent, A.F., and Clough, S.J. (1998). *Agrobacterium* germ-line transformation: Transformation of *Arabidopsis* without tissue culture. In Plant Molecular Biology Manual, 2nd ed, S.B. Gelvin and R.A. Schilperoot, eds (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 1–14.

Blaschek, W., Koehler, H., Semler, U., and Franz, G. (1982). Molecular weight distribution of cellulose in primary cell walls. Investigations with regenerating protoplasts, suspension cultured cells and mesophyll of tobacco. Planta **154,** 550–555.

Brown, R.M. (1996). The biosynthesis of cellulose. J. Macromol. Sci.-Chem. **A33,** 1345–1373.

Brown, R.M., Saxena, I.M., and Kudlicka, K. (1996). Cellulose biosynthesis in higher plants. Trends Plant Sci. **1,** 149–156.

Carpita, N., and Vergara, C. (1998). A recipe for cellulose. Science **279,** 672–673.

Creusot, F., Fouilloux, E., Dron, M., Lafleuriel, J., Picard, G., Billault, A., Le Paslier, D., Cohen, D., Chaboute, M.E., Durr, A., Fleck, J., Gigot, C., Camilleri, C., Bellini, C., Caboche, M., and Bouchez, D. (1995). The CIC library: A large insect YAC library for genome mapping in *Arabidopsis thaliana.* Plant J. **8,** 763–770.

Cutler, S., and Somerville, C.R. (1997). Cellulose synthesis: Cloning in silico. Curr. Biol. **7,** R108–R111.

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olson, G.J., and Swanson, R.V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus.* Nature **392,** 353–358.

Delmer, D.P. (1998). A hot mutant for cellulose synthesis. Trends Plant Sci. **3,** 164–165.

Delmer, D.P., and Amor, Y. (1995). Cellulose biosynthesis. Plant Cell **7,** 987–1000.

Emons, A.M.C., and Mulder, B.M. (1998). The making of the architecture of the plant cell wall: How cells exploit geometry. Proc. Natl. Acad. Sci. USA **95,** 7215–7219.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution **39,** 783–791.

Felsenstein, J. (1993). PHYLIP (phylogeny inference package) version 3.5. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fry, S.C. (1988). The Growing Plant Cell Wall: Chemical and Metabolic Analysis. (Harlow, UK: Longman).

Giddings, T.H., Brower, D.L., and Staehelin, A. (1980). Plasma membrane of *Micrastrias denticulata* associated with the formation of cellulose microfibrils in primary and secondary cell walls. J. Cell Biol. **84,** 327–339.

Lister, C., and Dean, C. (1993). Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana.* Plant J. **4,** 745–750.

Matthysse, A.G., White, S., and Lightfoot, R. (1995). Genes required for cellulose synthesis in *Agrobacterium tumefaciens.* J. Bacteriol. **177,** 1069–1075.

Meyer, K., Leube, M.P., and Grill, E. (1994). A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana.* Science **264,** 1452–1455.

O'Donnell, P.J., Truesdale, M.R., Calvert, C.M., Dorans, A., Roberts, M.R., and Bowles, D.J. (1998). A novel tomato gene that rapidly responds to wound- and pathogen-related signals. Plant J. **14,** 137–142.

Pear, J.P., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P., and Stalker, D.M. (1996). Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. USA **93,** 12637–12642.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Saxena, I.M., Lin, F.C., and Brown, R.M. (1990). Cloning and sequencing of the catalytic subunit of cellulose synthase from *Acetobacter xylinum*. Plant Mol. Biol. **15,** 673–683.

Saxena, I.M., Brown, R.M., Fevre, M., Geremia, R.A., and Henrissat, B. (1995). Multidomain architecture of β-glycosyl transferases: Implications for mechanism of action. J. Bacteriol. **177,** 1419–1424.

Schmidt, R., Love, K., West, J., Lenehan, Z., and Dean, C. (1997). Description of 31 YAC contigs spanning the majority of *Arabidopsis thaliana* chromosome 5. Plant J. **11,** 563–572.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22,** 4673–4680.

Turner, S.R., and Somerville, C.R. (1997). Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell **9,** 689–701.

Voytas, D.F., Konieczny, A., Cummings, M.P., and Ausubel, F.M. (1990). The structure, distribution and evolution of the TA1 retrotransposable element family of *Arabidopsis thaliana.* Genetics **126,** 713–721.