A Comparative Analysis of the Plant Cellulose Synthase (*CesA*) Gene Family[1]

Neta Holland[2], Doron Holland[2], Tim Helentjaris, Kanwarpal S. Dhugga, Beatriz Xoconostle-Cazares[3], and Deborah P. Delmer*

Section of Plant Biology, University of California, Davis, California 95616 (N.H., D.H., B.X.-C., D.P.D.); and Pioneer-Hibred International, Johnston, Iowa 50131–1004 (T.H., K.S.D.)

CesA genes are believed to encode the catalytic subunit of cellulose synthase. Identification of nine distinct *CesA* cDNAs from maize (*Zea mays*) has allowed us to initiate comparative studies with homologs from Arabidopsis and other plant species. Mapping studies show that closely related *CesA* genes are not clustered but are found at different chromosomal locations in both Arabidopsis and maize. Furthermore, sequence comparisons among the CesA-deduced proteins show that these cluster in groups wherein orthologs are often more similar than paralogs, indicating that different subclasses evolved prior to the divergence of the monocot and dicot lineages. Studies using reverse transcriptase polymerase chain reaction with gene-specific primers for six of the nine maize genes indicate that all genes are expressed to at least some level in all of the organs examined. However, when expression patterns for a few selected genes from maize and Arabidopsis were analyzed in more detail, they were found to be expressed in unique cell types engaged in either primary or secondary wall synthesis. These studies also indicate that amino acid sequence comparisons, at least in some cases, may have value for prediction of such patterns of gene expression. Such analyses begin to provide insights useful for future genetic engineering of cellulose deposition, in that identification of close orthologs across species may prove useful for prediction of patterns of gene expression and may also aid in prediction of mutant combinations that may be necessary to generate severe phenotypes.

Evidence is accumulating to support the notion that some, if not all, of the members of the family of *CesA* genes in plants encode a glycosyltranferase that plays a key role in the process of cellulose synthesis (for recent reviews, see Brown et al., 1997; Kawagoe and Delmer, 1997, 1998; Delmer, 1999). The deduced proteins from members of this gene family are characterized by the presence of domains that share significant sequence homology with other family 2 glycosyltransferases that are characterized by having conserved motifs surrounding three conserved D residues and a QXXRW motif downstream of D_3 (Campbell et al., 1997). Recent crystallographic evidence supports a model in which the three D residues, in conjunction with a divalent cation, are involved in binding of the UDP-sugar substrate and in catalysis of glycosyltransfer (Charnock and Davies, 1999). In the deduced proteins encoded by most family 2 glycosyltransferases, the domains containing these conserved D residues are consecutive, but the predicted proteins in plants contain a plant-specific conserved and a hypervariable (HVR-2) domain that separate the domains containing these conserved residues. A conserved, extended N-terminal region containing two zinc fingers resembling LIM/Ring domains (Kawagoe and Delmer, 1997) followed by the HVR-1 region also characterizes the plant CesA proteins. Many of these glycosyltransferases, including the plant and bacterial CesA proteins, are predicted to be anchored in the plasma membrane by two transmembrane helices upstream of D_1 and six more downstream of the QXXRW motif. A model has been proposed (Delmer, 1999) in which the three D residues may be positioned close to each other by formation of a channel comprised of the eight transmembrane helices through which a potential growing glucan chain might be secreted.

Several lines of evidence implicate the plant *CesA* genes in the process of cellulose synthesis. In cotton fibers, *GhCesA-1* and *GhCesA-2* (previously referred to as CelA-1 and CelA-2, see Delmer [1999] and http://www-plb.ucdavis.edu/labs/delmer/genes.html for recent changes in nomenclature of these genes) are highly expressed at the onset of secondary wall synthesis when the rate of cellulose synthesis in vivo rises over 100-fold (Pear et al., 1996). Furthermore, a recombinant protein comprising the central catalytic domain was shown to bind UDP-Glc, the predicted substrate. Mutation of the *AtCesA-1* (*Rsw1*) gene of Arabidopsis leads to marked reduction in deposition of crystalline cellulose in young, expanding cells of seedlings (Arioli et al., 1998), whereas mutation of *AtCesA-7* (*Irx3*) leads to a phenotype of collapsed

[1] This work was supported by Pioneer-Hibred (grants to T.H. and K.S.D.), by the U.S. Department of Energy (grant no. DE–FG–03–963ER 20238 to D.P.D.), and by the National Science Foundation Plant Genomics Program (grant no. DBI 9872627).

[2] Present address: The Volcani Institute, P.O. Box 6, Bet Dagan, Israel.

[3] Present address: CINVESTAV-IPN, Department of Biotechnology and Bioengineering, Ave IPN 2508, Zacatenco, Mexico D.F. 07360

* Corresponding author; e-mail dpdelmer@ucdavis.edu; fax 530–752–5410.

xylem vessels and reduced cellulose content in the inflorescence stem (Turner and Somerville, 1997; Taylor et al., 1999). A polyclonal antibody against the central domain of a cotton CesA protein has recently been shown to react with plasma membrane rosettes, structures proposed to represent cellulose synthase complexes in plants (Kimura et al., 1999). Taken together, these results strongly support a critical role for at least some of the *CesA* genes in cellulose synthesis. By analogy with the presumed reaction catalyzed by the bacterial homologs, *AxCesA-1* and *AxCesA-2*, the most logical reaction catalyzed by the plant proteins is use of UDP-Glc as substrate to effect elongation of the growing 1,4-β-glucan chains of cellulose, although one still cannot exclude the possibility of other glycosyltransferase reactions such as synthesis of a protein- or lipid-linked intermediate.

To date in Arabidopsis, 12 distinct *CesA* genes can be identified from sequencing of expressed sequence tags (ESTs) and genomic DNA. In addition, this family appears to be part of a much larger superfamily that includes related members referred to as cellulose-synthase-like genes (Cutler and Somerville, 1997; http://cellwall.stanford.edu/cellwall/). Thus, it will be a challenge to determine the patterns of expression and precise functions of the various members of this superfamily in plants. For the *CesA* family, one likely possibility is that all members catalyze the same reaction and that many related genes evolved via gene duplication to provide distinct cell- and tissue-specific patterns of expression; in addition, the distinct HVR regions in the proteins might conceivably play a role in interaction with other unique cell-type-specific proteins involved in the process.

Alternatively, these differences in amino acid sequence may represent alternate glycosyltransferase activities. As mentioned previously, the only data available for Arabidopsis indicate that *AtCesA-1* and *AtCesA-7* are critical for cellulose synthesis but are expressed in different cell types. In cotton the fact that the two reported *GhCesA* genes are both highly expressed in the unique cell type, the fiber, at the same time in development suggests another possible role for multiple genes—to provide redundancy and possibly to allow for the very high rates of cellulose synthesis that have been selected for in these unique cells. Since cellulose is a key polymer in the walls of almost all plant cells, understanding the function of the various *CesA* genes is important for the understanding of plant development. In addition, the commercial importance of cellulose as a major component of cotton fibers, wood, and sources of forage make these genes attractive targets for genetic engineering. Understanding where the genes are expressed, which genes may be redundant, and what phenotypes are generated when each is mutated or overexpressed thus become important goals for such engineering.

The commercial importance of maize (*Zea mays*), as well as the large collection of ESTs from diverse tissue-specific cDNA libraries and a large transposable mutator (Mu)-insertion-generated mutant population that exist in the Pioneer-Hibred collections, make maize another attractive system for study of the *CesA* genes. We report here on the identification of nine distinct *CesA* cDNAs from maize. Having these clones available has allowed us to carry out a more extensive comparative analyses of chromosomal locations, sequence similarities, and patterns of expression for some of these genes in maize, Arabidopsis, and other plant species. Taken together, this analysis begins to provide a basis upon which predictions useful for future genetic engineering can be made concerning patterns of expression and expected mutant phenotypes for previously uncharacterized members of the family in both dicotyledonous and monocotyledonous plants.

RESULTS AND DISCUSSION

Identification of Maize *CesA* cDNA Clones

Pioneer-Hibred, working with Human Genome Sciences, constructed a maize EST collection and database based upon the principles outlined originally by Adams et al. (1991). This involved the production and analyses of over 130 cDNA libraries produced from a wide variety of maize organs obtained at various developmental stages or under different environmental conditions. Several hundred to a few thousand clones from each of these libraries were analyzed via a single sequencing run from the presumed 5′ end of the original mRNA to generate a database totaling over 200,000 entries. This clone and sequence database provides a rich source of information regarding the general set of expressed genes in maize.

This EST database was screened for sequences with significant similarity to the previously described *GhCesA-1* cDNA from cotton (Pear et al., 1996) by BLASTing (Altschul et al., 1990) their deduced amino acid sequences against all possible sequences deduced for each of the maize entries. This produced a preliminary list of greater than 350 entries with some similarity to *GhCesA-1*. By aligning these individual sequence runs into contigs to reduce their number to sets representing unique genes and emphasizing those that appeared to also represent full-length cloned versions of the original mRNAs, a set of nine clones was identified for further characterization (Table I). Each of these clones was then fully sequenced, producing full-length cDNA sequences for all but *ZmCesA-3* (828 amino acids of sequence, missing the 5′-most end of the gene). Using all of these sequences to probe back against the database, it was found that they each subsumed some subset of the previously mentioned *CesA*-like entries (A-1 = 34, A-2 = 18, A-3 = 21, A-4.1 = 9, A-4.2 = 25, A-5 = 28, A-6 = 33,

Table I. *Full-length or near full-length maize CesA cDNAs*

Name	Map to Chromosome	ESTs Found in These Libraries[a]
ZmCesA-1	8.02	IE, IT, K, P, R, S, S/S, T, TC
ZmCesA-2	6.05	IE, K, L, R, T, TC
ZmCesA-3	2.08	E, IT, L, T, TC
ZmCesA-4	2.06/7.01	IE, IT, K, R, L, S/S, T, TC
ZmCesA-5	1.10–1.11	C, E, IE, K, L, P, T, TC
ZmCesA-6	1.10	E, K, L, P, R, S/S, TC
ZmCesA-7	Not mapped	IE, IT, L, P, S, S/S, T, TC
ZmCesA-8	7.01	IE, K, P, S, S/S, T, TC
ZmCesA-9[b]	7.02	IE, IT, K, S/S, T

[a] Libraries: C, coleoptile; E, endosperm; IE, immature ear; IT, internal, ground tissue of stalk internodes; K, developing kernels; L, leaf; P, pericarp; R, root; S, seedling; S/S, shank/stalk; T, tassel; TC, tissue culture. [b] This has identical deduced protein sequence to *ZmCesA-4*, but 3′-UTR sequence differs significantly; thus, it was not possible to distinguish between these two genes for some ESTs.

A-7 = 34, and A-8 = 36), totaling approximately 238 entries and leaving approximately another 130+ to be accounted for by other genes not analyzed further. What might this number of unaccounted-for genes amount to? This is difficult to determine precisely but, by examination of the remaining contigs and singletons, there were approximately 82 entries included in contigs, with the remaining entries existing as "scattered" or unaligned singletons. It would appear that there may exist in the maize genome at least six and probably more genes in addition to the nine described herein, but possibly none of them are expressed as abundantly as this set.

Sequence Comparison of CesA Proteins and Mapping of the Maize *CesA* Genes

Figure 1 shows an unrooted cladogram that is based upon parsimony analysis for the deduced proteins from the known *CesA* genes from Arabidopsis, maize, cotton, rice, and poplar. The analysis shown excluded the major HVR regions, although similar results were obtained when they were included, and analyses of just the HVR regions also yielded similar cladograms (not shown). Although we recognize that not all *CesA* genes from these plants have yet been identified, some preliminary conclusions can be made. As noted before, for at least some large gene families (e.g. plant resistance genes; Michelmore and Meyers, 1998), orthologs are often more similar than paralogs in the major groupings of the cladogram. Thus, many of the clades contain members from plants of both monocot and dicot lineages, indicating that the divergences into at least some of these subclasses may have arisen relatively early in the evolution of these genes. Although the HVR regions appear to have been under less constraint and show more variability, the overall clade-specific patterns for these regions were also apparently established early in the evolution of these genes.

All but one of the *ZmCesA* genes (*ZmCesA-7*) were analyzed for their genomic origins via RFLP association and segregation. It was normal for each of these clones to detect more than one hybridizing fragment, although there was often one of much greater signal intensity than the others. Map locations for each gene are listed in Table I and are also shown on the cladogram of Figure 1. Three general patterns are seen among this group of genes, one pair mapping on chromosomes 6 and 8 (*ZmCesA-1* and *-2*), another group mapping on chromosomes 2 and 7 (*ZmCesA-3*, *-4*, *-8*, and *-9*), and another pair mapping to the long arm of chromosome 1 (*ZmCesA-5* and *-6*). These chromosomal locations were clustered in that the areas on chromosomes 6 and 8 and the regions on chromosomes 2 and 7 each represent ancient paired duplications within the maize genome (Helentjaris et al., 1988). *ZmCesA-1* and *-2* represent very similar sequences (see Fig. 1), even at the nucleotide level, and so it is conceivable that these reflect ancient duplications of the same original gene. Similarly, one pair of genes, *ZmCesA-4* and *-9*, differ almost exclusively in their 3′-untranslated regions (3′-UTRs) and so probably define a paired set of duplicated genes on chromosome 7. At the same time, there are two other genes clustered within these same regions (*ZmCesA-3* and *-8)* that show much less similarity to each other. Also, *ZmCesA-5* and *-6* both map to the same general region of chromosome 1 but are much less related by sequence and are positioned within very different regions of the cladogram. This latter type of genomic clustering is also seen with genes in the Arabidopsis genome via inspection of their map locations and cluster patterns on the cladogram of Figure 1. For example, AtCesA-1 and -13 and AtCesA-2 and -9 are

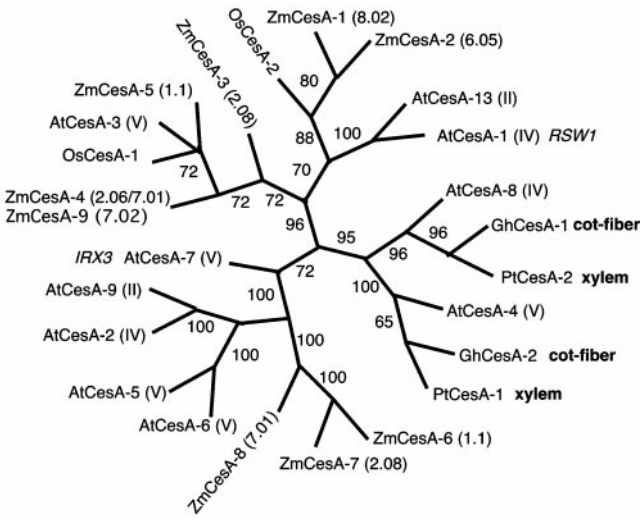

Figure 1. Unrooted cladogram (strict consensus tree) for plant CesA proteins. Numbers in parentheses after the protein names refer to the chromosomal map locations of the genes. ZmCesA-4 and -9 are listed as one entry since their deduced amino acid sequences are essentially identical.

two pairs that are very similar in sequence, but the genes map to different chromosomes, and quite often the CesA proteins derived from genes of any single chromosome are scattered into different branches on the cladogram.

Thus, two consistent conclusions are observed with these mapping results—that closely related cellulose synthase genes are often found at different chromosomal locations that reflect previously described ancient duplications within a plant's genome and that cellulose synthase genes are also found close to each other in the genomes that are much less related to each other by sequence. Presumably, genes were duplicated in specific locations by unequal recombination in the very distant past and then these genes subsequently diverged from each other, at least by sequence and possibly even by function. In a more recent event in maize, a duplication of these entire regions occurred that created the set of genes that we see today.

General Patterns of Expression for the Maize *CesA* Genes

Since each of these genes is represented in the database by multiple entries, the libraries of origin for each can represent a crude "electronic northern" to indicate something about the expression patterns of these genes. A summary of the library types from which the original ESTs were identified for each gene is included in Table I. The most striking results from the preliminary examination of expression of these genes is the diversity of their source libraries, indicating that very few if any of these genes exhibit a strong organ type of preferred expression, i.e. we see no genes that are only expressed in roots or tassels or kernels. Instead, the ESTs for each gene collectively seem to originate from a variety of tissues/organs, many of them overlapping extensively with other gene family members. It seems unlikely that all of these genes are functionally redundant, and perhaps it is more likely that their expression, if it has any specificity at all, might be more related to individual cell types or their developmental state. Clearly this type of analysis is only sufficient to determine the absence of any higher order of tissue-specific expression.

The conclusions drawn from analyzing the libraries of origin are further supported by measurements of mRNA levels in various organs or organ parts. Signals from some of the genes were weak in our attempts at northern blot analysis, and therefore we used the more sensitive technique of reverse transcriptase (RT)-PCR to obtain a semiquantitative assessment of mRNA levels for these genes in various regions of maize seedlings using amplification of a conserved region of actin as an internal reference. However, we did note that for several of the genes that gave the strongest signals on northern blots (*ZmCesA-2* and *ZmCesA-8*), two sizes of mRNAs, one around the predicted size of 3.7 to 3.8 kb and a second a few hundred base pairs smaller, were routinely observed using probes based upon the HVR-2 region (not shown). This may indicate alternative splicing of these genes, but we have not yet explored this finding further. For the RT-PCR studies, we used gene-specific probes based upon the 3′-UTR of each gene. We verified their specificity by PCR using the cDNA as template and demonstrated that each primer set resulted in amplification of a single band of predicted size only with the appropriate cDNA and not with any cDNAs from the other *ZmCesA*s. An example of this specificity is shown for *ZmCesA-8* in Figure 2A, and similar clearly specific results were obtained for the other genes, with the exception of *ZmCesA-2* and *ZmCesA-4 (ZmCesA-9*, being so similar to *ZmCesA-4,* was not analyzed*)*. Because minor reactions occurred with other members of the gene family using the primers for these genes, we do not present results for these genes. We also verified that, within a certain range, the amplification of the predicted bands was linear with respect to RNA concentration (Fig. 2B), and we chose a concentration of RNA within this range for our studies.

Results that are representative of several different experiments are shown in Figure 3. Although reproducible, these results should be considered only semiquantitative since we do not know how actin gene expression varied in these organs; also, it is obvious that we have only examined the plant parts

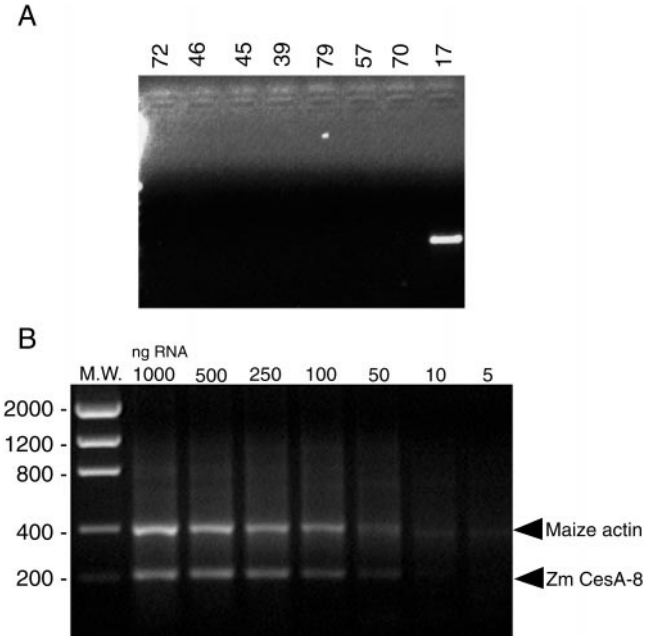

Figure 2. Controls for RT-PCR. A, Examples of specificity of primers. Example shown is for primers for *ZmCesA-8* used in PCR with each of the *ZmCesA1–8* cDNA clones (lanes 1–8). B, RT-PCR using maize RNA from etiolated maize seedlings. Reactions contained primers for both actin and ZmCesA-8.

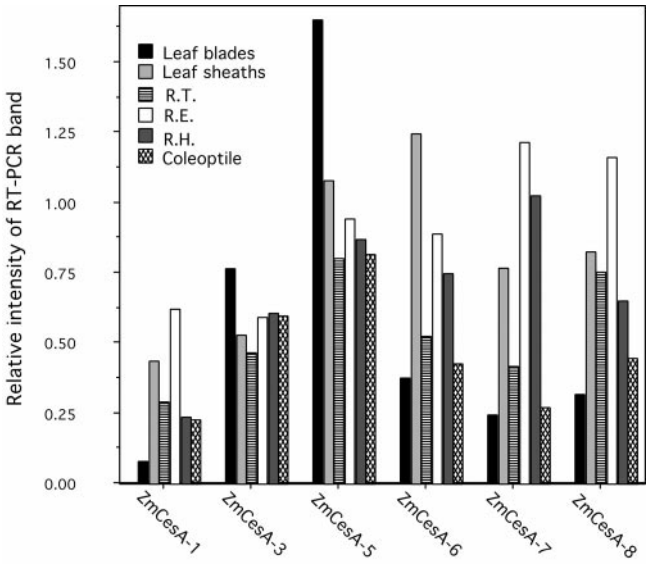

Figure 3. Expression of *ZmCesA* genes in various organs or organ parts of maize seedlings as determined by RT-PCR using gene-specific primers. Results are expressed relative to the intensity of the actin fragment that was generated within the same reaction. R.T., First 0.5 cm of root tip; R.E., root elongation zone from 0.5 cm behind tip up to region where root hair elongation begins; R.H., root hair zone.

present at a very early stage of development. Nevertheless, these result do confirm that each gene is expressed in most of the organs or parts of organs chosen—leaf blades and sheaths, coleoptile, and the root tip, root elongation zone, and the region where root hair differentiation begins. However, as might be expected, the levels of expression for each gene do differ for different regions of the seedling, although we note that the three genes *ZmCesA-6* to *-8* that cluster in the cladogram also show very similar patterns of expression. Although expression is not specific to any single plant part, one still cannot conclude that these genes may not show cell type-specific expression. For example, since vascular tissue is present in essentially all of the plant parts examined, a gene might show very specific expression in such a tissue, but this would be reflected in its expression in each of the plant parts examined here by RT-PCR.

Cell-Specific Expression Patterns for Several *CesA* Genes

It is notable that, in the comparisons shown in Figure 1, the smallest clade that contains both GhCesA-1 and GhCesA-2 also contains other members that are known to be expressed in tissues engaged in secondary wall formation. The two GhCesA proteins are known to be expressed during secondary wall cellulose synthesis in cotton fibers; GhCesA-1 is grouped with PtCesA-2, a protein deduced from a cDNA isolated from a xylem-specific cDNA library. GhCesA-2 is grouped with another xylem-specific poplar protein (PtCesA-1). Each of these subgroups also contains one Arabidopsis protein (AtCesA-8 or AtCesA-4), the expression patterns of which have not been reported. Several of the maize proteins (ZmCesA-1 and ZmCesA-2) fall within a clade that contains AtCesA-1 (Rsw1), which is known to be functional in cells engaged in primary wall synthesis (Arioli et al., 1998). We also note that the three maize proteins from genes *ZmCesA-6* to *-8* cluster in a distinct subgroup in all our analyses and that these three genes showed very similar patterns in our general RT-PCR studies (Fig. 3).

If these groupings within clades represent clustering of proteins with functions related to deposition of either primary or secondary cell walls and/or in unique cell types, then certain predictions might begin to be made. First, since these groupings primarily arise from specific sequence pattern variation in the more conserved regions of the proteins and since promoter sequences are not taken into account in this analysis, it might be that some of these specific sequence patterns play some as-yet-unidentified role in the protein's function in these specific tissues. Second, one might be able to predict expression patterns for other members of a cluster based upon our current limited knowledge of expression of some of these genes/proteins. To begin to test this idea further, we have chosen the *ZmCesA-1* and *AtCesA-4* genes for further studies on tissue-specific expression, with the prediction that the former may be expressed in cells depositing primary cell walls and the latter in cells undergoing secondary wall formation. In addition, we have chosen *ZmCesA-8* for further study because we also have work in progress to analyze three independent mutant alleles generated by Mu insertions in this gene.

Using the same gene-specific probes described above, expression of *ZmCesA-1* and *-8* was assessed using the technique of in situ RT-PCR that employs the fluorescent substrate Oregon Green dUTP. Because of excessive autofluorescence and/or retention of un-polymerized substrate, not all tissues were found suitable for this technique. Figure 4 shows examples of expression patterns for *ZmCesA-1* and *ZmCesA-8* in longitudinal sections from roots of etiolated seedlings. For in situ RT-PCR, image intensities of observed fluorescence for confocal microscopy were adjusted to a level that showed essentially no background in control samples lacking the primers. Positive reactions seen at these intensities in samples with primers are indicated by the green fluorescence generated in the cytoplasm of cells where expression occurs. *ZmCesA-1* shows high expression in the region of cell elongation in both the pith and cortex (Fig. 4A), cells that are clearly undergoing primary wall deposition, a result that is consistent with the deduced protein falling within the same clade as AtCesA-1 (Rsw1). With the cautionary note that ex-

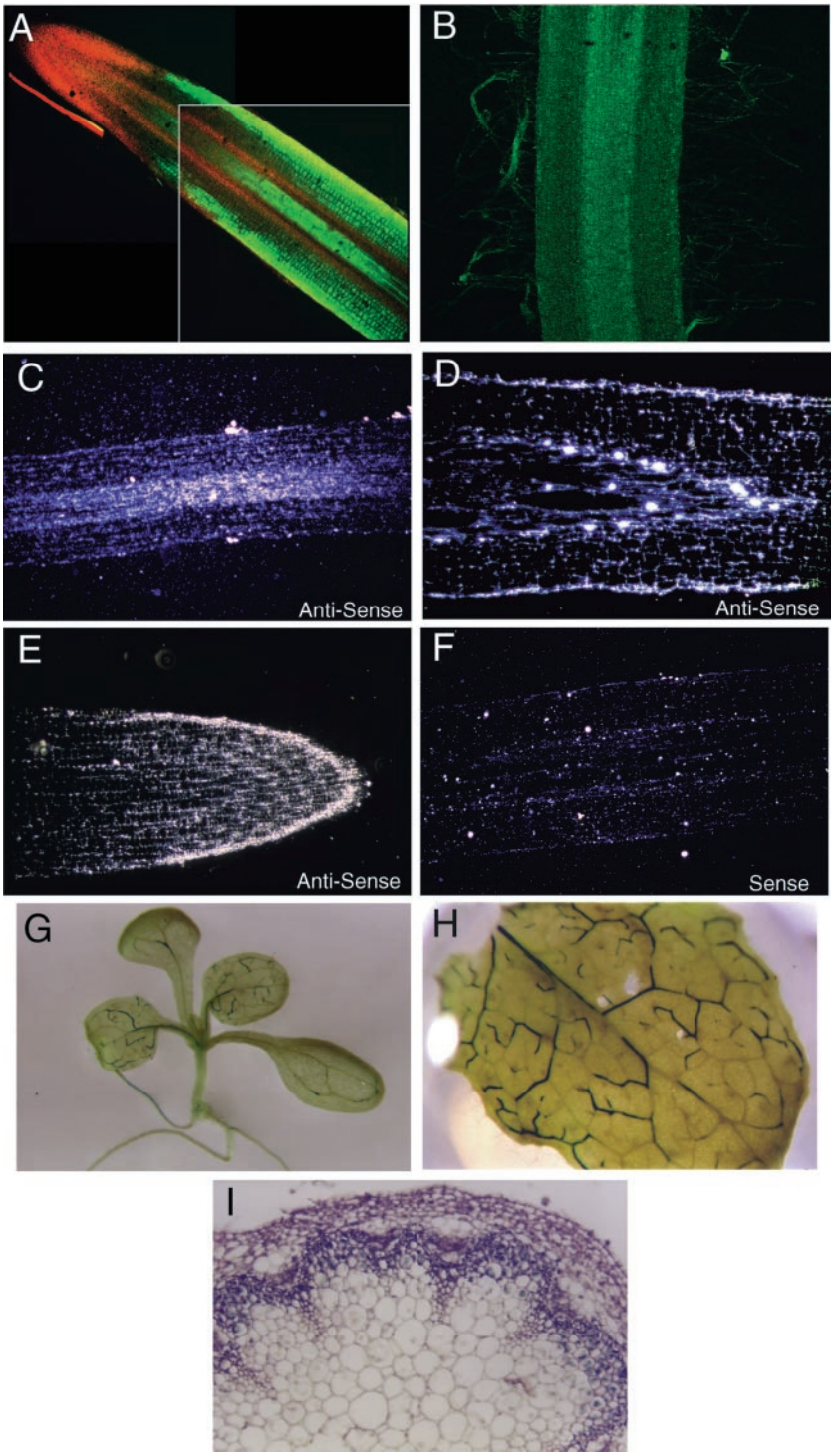

Figure 4. Examples of *CesA* gene expression in maize and Arabidopsis seedlings. A, RT-PCR for *ZmCesA-1* in root elongation zone; green color indicates expression overlaid on image of red fluorescence representing total tissue structure. B, *ZmCesA-8* expression in region of vascular development of root elongation zone; green color for expression only is shown. C–F, In situ hybridization using radioactive anti-sense probe for *ZmCesA-8* showing expression in epidermis and vascular region (C and D, the latter being a glancing section photographed at ×2 magnification as in C). E, Strong expression also seen in differentiating epidermis at root tip. F, Control using sense probe photographed at same light intensity and exposure time as that for C through E. G through I, In situ detection of GUS activity in transgenic Arabidopsis expressing the GUS gene under control of the *AtCesA-4* promoter. G and H, Expression is confined to vascular tissues in young leaves, hypocotyl, and roots. I, Cross-sections of inflorescence stem showing strong expression in the region where interfascicular fibers are developing.

pression has not been analyzed in many other plant parts, from this result one might predict that genes encoding other members of this clade (AtCesA-13, ZmCesA-2, and OsCesA-2) may also be expressed in cells involved in primary wall synthesis, a prediction that awaits further study. By contrast, the expression pattern for *ZmCesA-8* is observed in the region of developing vascular tissue (Fig. 4B). In these fresh longitudinal hand sections, the epidermis tended to peel away but also showed evidence of expression. In cross-sections, we also observed expression in the two layers of epidermis and exodermis in the zone of root elongation (not shown). Reaction in these cells was almost always stronger than in controls lacking primers, but we have found that these are cell types that have a tendency to retain in a variable fashion

the non-reacted fluorescent substrate. Because of these uncertainties, we have also used the alternate technique of in situ hybridization using a radioactive probe that is derived from the HVR-2 region of *ZmCesA-8*. These results in general confirm our finding in roots that expression is strong in the region where vascular tissue is developing and also in the epidermis (Fig. 4, C and D); near the root tip, we see an especially strong positive reaction in the region of developing epidermis (Fig. 4E).

Since *ZmCesA-3* shows expression in vascular tissues of young roots, it was of interest to see if it might also be expressed in the regions of vascular development in the stalks of greenhouse-grown older maize plants. This was examined using a ^{32}P probe derived from the same HVR-2 region with tissue prints of cut stalks blotted onto nylon. We do indeed see strong expression in the region where there is intense vascular bundle development (Fig. 5). Because the ^{32}P probe used is not finely localized on the films, we cannot say whether expression might also be occurring in the epidermis. As the stalk matured, expression was reduced to background levels (not shown).

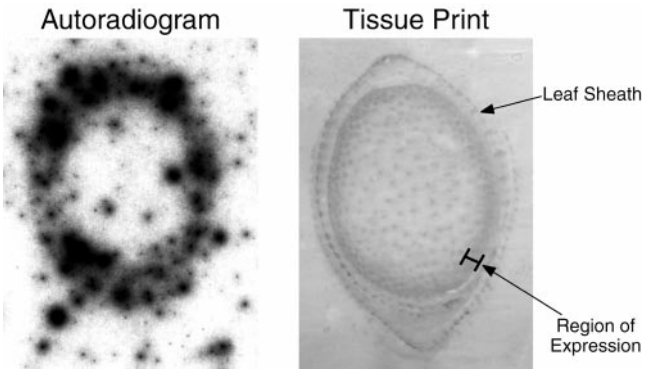

Figure 5. Tissue print showing expression of *ZmCesA-8* in region of developing vascular bundles/epidermis of young corn stalk. Tissue print used to generate autoradiogram was stained with India ink (right) to show details of stalk structure. Overlay of the radioautogram onto this print showed expression highest in the region indicated on the print. Probe used was derived from HVR-2 region of *ZmCesA-8*.

Expression Patterns for *AtCesA-4*

A genomic sequence (accession no. AB006703) for *AtCesA-4* has been identified on Arabidopsis chromosome 5. A P1 clone (MRH10) that covers the entire gene and its flanking sequences was used to clone the promoter region by PCR amplification. A 809-bp fragment of DNA immediately upstream to the coding region of *AtCesA-4* was placed in front of a promoterless *uidA* (β-glucuronidase [*GUS*]) gene. This construct was transferred to Arabidopsis by *Agrobacterium tumefaciens*-mediated stable transformation, and the expression patterns were analyzed by in situ assays of GUS activity of plants determined to be either heterozygous (F_1) or homozygous (F_2) for one copy of the GUS gene. For the 25 plants examined, we observed essentially the same patterns of expression as shown in Figure 4, and the pattern was similar in plants homozygous or heterozygous for the transgene. Expression was confined to the vascular tissues examined including roots, hypocotyls, and flowers (Fig. 4G; flowers not shown). In young leaves, the expression occurs in transient patches along the vascular system in patterns that might be considered to reflect the regions of localized expansion (Fig. 4, G and H). In the inflorescence stem, expression of *AtCesA-4* was observed initially in the vascular bundles of the primary xylem, whereas in maturing stems it was most pronounced in the region of developing interfascicular fibers (Fig. 4I). Thus, this gene also displays a pattern of expression exclusively in cells undergoing secondary wall cellulose deposition as predicted by its grouping in the cladogram with GhCesA-2 and PtCesA-1. Expression pattern for another Arabidopsis gene, *AtCesA-8*, is still unknown, but it might be predicted to be expressed in similar types of cells. (One wonders if this locus might encode the gene mutated in either the *irx1* or *irx2* mutant, genes that have not yet been characterized [Turner and Somerville, 1997].) We do note, however, that the gene *AtCesA-7* (*Irx3;* Taylor et al., 1999) does not fall into either of these two major clades representing known secondary wall genes, even though it is known to be expressed in developing vascular bundles. However, it is within the larger clade that contains ZmCesA-8 that we have shown is expressed in vascular tissue in this monocot. Thus, the data available at present for corn, poplar, Arabidopsis, and cotton do suggest that unique sequences within the protein, and not just promoter specificity alone, are correlated well with and, at least in the cases examined, may be indicative of a role in deposition of either the primary or the secondary cell wall. Extensive analysis of expression patterns for the other members of this family in Arabidopsis and maize is under way in our laboratories and should help clarify to what extent this generalization will hold true for the entire *CesA* family. Whether the sequence differences so far identified between members engaged in primary or secondary wall synthesis also reflect functional differences also remains to be determined. For example, they might reflect a requirement for interaction with a different set of accessory proteins, different patterns of rosette assembly, or even different reactions catalyzed—all possibilities that remain to be explored.

In many monocotyledonous plants including maize, a non-cellulosic β-1,3-β-1,4-glucan is a major component of some primary cell walls (Carpita, 1996). Xyloglucan that possesses a β-1,4-glucan backbone is also present but is much more abundant in walls of the dicots. Whether the backbone of either of these polymers, or of callose (β-1,3-glucan), may be synthesized by any specific member of the CesA proteins or by some other gene product is not yet

known (Buckeridge et al., 1999). Xyloglucan and the mixed-linked glucan are synthesized in the Golgi, where many glycosyltransferases are type II integral membrane proteins with a single membrane-spanning region and catalytic domain in the lumen (Nilsson and Warren, 1994). However, these are generally non-processive glycosyltransferases that transfer only one sugar residue, whereas enzymes synthesizing the mixed-linked glucan or the backbone of xyloglucan catalyze processive reactions (more than one sugar added in repetitive fashion) that are usually carried out by family 2 glycosyltransferases. The fact that Buckeridge et al. (1999) have found a Suc synthase associated with rice Golgi membranes suggests the possibility that a family 2 glycosyltransferase such as a CesA could be inserted in an orientation and topology similar to that predicted for CesAs in the plasma membrane—an orientation that would place the catalytic domain in the cytoplasm to interact with Suc synthase and allow secretion of the growing polymer into the lumen of the Golgi. Hydropathy plots (Kyte and Doolittle, 1982) for all nine CesA proteins from maize are strikingly similar with eight predicted transmembrane helices (not shown), and none show any obvious sequences that might denote retention in the Golgi, although we also recognize that Golgi retention signals for such proteins in plants are poorly understood. We also note that all of the maize and rice CesA proteins fall into major groupings with orthologs from Arabidopsis, a dicot that does not make the mixed-linked glucan (Fig. 1), and that no major lineage is strictly from monocots, although some subgroups are. At this point, we consider it unlikely (but certainly not impossible) that any of these CesA proteins would catalyze a unique reaction distinct from other CesA proteins such as that involved in xyloglucan or mixed-linked glucan synthesis. However, further studies will clearly be necessary to resolve this question.

Looking to the future, it will also be important to determine which, if any, of the *CesA* paralogs within a species may be expressed and functionally redundant within one cell type at one time in development. Redundancy might be predicted for pairs or triplets of genes that cluster closely together in both Arabidopsis and maize, and this situation may complicate the ability to assess phenotypes via gene knockouts unless all redundant genes are repressed or mutated within the same plant. In fact, our work in progress indicates that three independent alleles showing Mu insertions in exons of the *ZmCesA-8* gene display no gross alterations in vascular development, a major site of expression of this gene, suggesting that other *CesA* genes may be redundantly expressed in these cell types. Analyses such as those presented here may facilitate predictions that certain mutant combinations of genes encoding proteins that fall within a clade may have more severe phenotypes than single mutants; in the case of *ZmCesA-8*, one might predict that *ZmCesA-6* and/or *-7* could fulfill this function, a prediction that is currently being explored by us. Another challenge will be to develop appropriate approaches for assessing the pattern of expression and function for each of the members of this family in any particular plant species. Because many of the genes appear to be expressed in unique cell types that exist in a variety of different tissues, analyses of developmentally regulated expression patterns cannot be accurately assessed by the general techniques of northern, microarray/gene chip, or RT-PCR analyses, although these techniques should be suitable for studying effects of nutritional or environmental factors on expression. In situ expression studies are tedious and, in the long run, are not feasible for every type of tissue for each of the many genes. The use of promoter-reporter fusions, although having its own limitations, still offers one of the best approaches for studying patterns of expression, especially if coupled to reporter genes such as green fluorescent protein or luciferase, which might allow one to follow expression changes in real time during the development of a single plant. Finally, in discussing expression, this work has not considered other likely levels of control beyond the level of transcription that may well also play important roles in the ultimate control of CesA activity.

Assessment of function, i.e. confirmation of the specific reaction catalyzed by each of the genes, poses an even greater challenge. Mutant analysis alone, as currently being carried out, can tell one whether the gene is critical for a specific process such as cellulose synthesis, but it cannot define exactly which reaction is being catalyzed by the gene product under study. One would hope to express the genes in functional form in a heterologous host. Although we have recently found that the *GhCesA-1* gene can be expressed and inserted into membranes in predicted topological orientation in yeast cells, no rosette structures are formed nor is any glycosyltransferase activity detected (Y. Kawagoe, C.H. Haigler, and D.P. Delmer, unpublished results); a similar situation is emerging from our studies with expression in green monkey kidney (COS) cells (C. Grubb, A. Spicer, and D.P. Delmer, unpublished observations.) Thus, it appears that the expression of other necessary genes is also required to allow functional expression of these genes in heterologous systems. Nevertheless, determination of function in addition to the consequences of over- or under-expression of the individual genes is critical for development of strategies for engineering altered patterns of cellulose deposition in important crop plants.

MATERIALS AND METHODS

Isolation of Maize (*Zea mays*) *CesA* cDNA Clones

Total plant RNA was isolated from tissues/organs using the phenol-guanidine isocyanate method according to the

manufacturer's protocol (TRIzol, Life Technologies, Cleveland). Poly(A) RNA was isolated from this preparation using oligo(dT)-coated magnetic beads and the PolyATract system (Promega, Madison, WI). mRNAs were directionally cloned as cDNAs (5′>*Sal*I. . . . . *Not*I>3′) into the pSPORT1 vector using the SuperScript system (Life Technologies).

Mapping of Maize *CesA* Genes

CesA genes were mapped to their locations within the maize genome by associating an RFLP with each clone and determining its segregation within a characterized population, by what are now standard procedures as described by Helentjaris et al. (1985). Southern analysis was used to detect the initial RFLP (utilizing four different restriction enzymes, *Bam*HI, *Eco*RI, *Eco*RV, and *Hin*dIII) between parental lines for five populations that have been extensively characterized previously at Pioneer-Hibred using molecular markers. Appropriate probe and enzyme combinations were then used to determine segregation of the RFLP within segregating progeny, and associations between it and other characterized loci were determined by using MapMaker (Lander et al., 1987). Assignments to chromosomal bins were based upon the latest information contained in the Maize Genome Database.

Amino Acid Sequence Comparisons

The cladogram shown in Figure 1 was generated using PAUP 4.0 beta version (Phylogenetic Analysis Using Parsimony, version 4.0, Sinaur Associates, Sunderland, MD) with parsimony analysis employing a heuristic search algorithm with 100 replicates. Bootstrap analyses were carried out using parsimony as optimizing criterion analysis with bootstrap resampling (100 replicates); bootstrap values >50 are indicated in Figure 1. Prior to PAUP analysis, amino acid sequences were aligned using the Megalign/Clustal Program of Lasergene (DNASTAR, Madison, WI). HVR regions were discarded from the alignment prior to parsimony analysis. GenBank accession numbers for genes used to derive the amino acid sequences are: *AtCesA-1*, AF027172; *AtCesA-2*, AF027173; *AtCesA-3*, AF027174; *AtCesA-4*, AB006703; *AtCesA-5*, AB016893; *AtCesA-6*, AF062485; *AtCesA-7*, AF088917; *AtCesA-8*, AL035526; *AtCesA-9*, AC007019; *AtCesA-13*, AC006300; *ZmCesA-1*, AF200525; *ZmCesA-2*, AF200526; *ZmCesA-3*, AF200527; *ZmCesA-4*, AF200528; *ZmCesA-5*, AF200529; *ZmCesA-6*, AF200530; *ZmCesA-7*, AF200531; *ZmCesA-8*, AF200532; *ZmCesA-9*, AF200533; *GhCesA-1*, U58283; *GhCesA-2*, U-58284; *PtCesA-1*, AF081534; *PtCesA-2*, AF072131; *OsCesA-1*, AF030052; and *OsCesA-2*, D48636.

AtCesA-4 Expression

A P1 clone (MRH10; accession no. AB006703) that spans the genomic sequence for *AtCesA-4* was kindly provided by Y. Nakamura (Kazusa DNA Research Institute, Chiba, Japan). Two primers (5′ primer, 5′-ATC CGG TCG ACA TTA AAT CTT ATT TAC TAA CAA AAC-3′; and 3′ primer, 5′-AAA GGA TCC CAC GAA ATG TAC ATT ACG TTG AG-3′) were designed to amplify by PCR a fragment of DNA that spans the intergenic region between the *AtCesA-4* and the gene that precedes it upstream. This region starts immediately after the stop codon of the preceding gene and stops immediately before the start codon of *AtCesA-4*. The amplified fragment had a *Sal*I site at its 5′ end and a *Bam*HI site at its 3′ end that were nested within the amplifying primers. The 809-bp PCR-amplified fragment was cloned into pBluescript as a *Sal*I-*Bam*HI fragment. The *Sal*I-*Bam*HI fragment was then placed in front of the promoterless *uidA* (*GUS*) gene in binary vector pBI 101 to generate binary plasmid vector pDel-158. Seeds of Arabidopsis ecotype Columbia were placed at 4°C for 48 h prior to germination and then were grown in controlled growth chambers at 26°C with 16-h/8-h light/dark cycles. Transformation was done according to a modified vacuum infiltration protocol by dipping flowering stems in infiltration medium containing *Agrobacterium tumefaciens* EHA 105 harboring pDel-158 in binary vector PBI 101 as described by Clough and Bent (1998). Selection for transgenic plants was done on kanamycin plates. Seeds from kanamycin-resistant plants were collected and the F_2-segregating population was analyzed for GUS activity to select plants with a single copy per haploid genome and that are homozygous for the *pAtCesA-4-GUS* sequence.

For in situ analyses of GUS activity, young seedlings grown under conditions described above on agar plates containing one-half-strength Murashige-Skoog salts were harvested and pretreated with ice-chilled 95% (v/v) acetone for 15 min; they were then incubated at 37°C for 1 to 5 h in 25 mM potassium phosphate buffer, pH 7.0, containing 0.85 mg mL^{-1}, 0.25% (w/v) Triton X-100, 2.5 mM EDTA, and 1.25 mM ferrocyanide and ferricyanide. Tissues were cleared by immersing in 95% (v/v) ethanol prior to examination. For examination of cross-sections of the inflorescence stem, tissue was fixed in ethanol:glacial acetic acid:formaldehyde:water (50:5:10:35, v/v), dehydrated, embedded in paraffin, and sectioned.

RT-PCR on Total RNA

For determining general patterns of expression, RT-PCR was carried out using total RNA and gene-specific primers based upon the 3′-UTR of each *ZmCesA* gene. Total RNA was isolated from various parts of young maize seedlings (hybrid line 3489, Pioneer-Hibred, Des Moines, IA) using a method employing TRIzol Reagent (monophasic phenol/guanidine isothiocyanate mixture; Gibco-BRL, Grand Island, NY) according the manufacturer's protocol. The following forward and reverse primers, respectively, were used: for *ZmCesA-1*, 5′-TGA AGA GGC TCA ATC AAG ATC TGC-3′ and 5′-TAA AAC AAT AAA CTG CAC GCA TAA C-3′; for *ZmCesA-2*, 5′-TGA CTC CTT ATC TGA AGA GGC TC-3′ and 5′-GCA TTG CCA TAT AGT TCG TGT GAA TAG-3′; for *ZmCesA-3*, 5′-GCA GCA TGA AAC TTT GTC AAC TTA TG-3′ and 5′-AGA GAG AGA ATC CCT CAA TGT CAT TGA CC-3′; for *ZmCesA-4*, 5′-GGA AGT

GGA AGG TTT GTA CTT TG-3′ and 5′-GCGGACCCA-CAGGAGCGAGAAG-3′; for *ZmCesA-5*, 5′-GAT GAG CTG AAG ATA GTT AAA GAG TG-3′ and 5′-CAA AGC AGT ATA TAT ATT AAC TTA CGG-3′; for *ZmCesA-6*, 5′-TTG TCC CTC TGT AGA TTG AAA CAA G-3′ and 5′-GAT TGA TTG GTC GCG ATC ATC CC-3′; for *ZmCesA-7*, 5′-CTC AAT AAG GCA GGC AGG AAT G-3′ and 5′-ACC AAC TCA GAT GAT TAC AGT AC-3′; for *ZmCesA-8*, 5′-ATC TCG AAC GCG ATC AAC AAC G-3′ and 5′-GGC GAG GAA CGG GTC GAC GC-3′; and for conserved region of maize actins, 5′-TTC AGG TGA TGG TGT GAG CC-3′ and 5′-CCT GAT ATC AAC ATC ACA CTT C-3′. The RT step was carried out using 0.1 to 1 μg total RNA that was first incubated 2 min at 70°C with 2 μM oligo(dT) GAGA (Stratagene, La Jolla, CA) in a final volume of 5 μL; this mixture was then incubated on ice for 2 min and centrifuged 2 s at 10,000*g*. Following this, 2 μL of 5× Superscript buffer (Gibco-BRL), 1 μL of 20 mM dithiothreitol, 1 μL of a stock of 10 mM each dNTPs (final concentration 1 mM each), and 1 μL of Superscript II (200 units; Gibco-BRL) were added, and the mixture was incubated for 1 to 2 h at 42°C. The succeeding PCR step was initiated using 2 μL from the above RT reaction incubated in a final volume of 50 μL containing 0.2 mM each dNTPs, 1 μM each of CesA and actin primers, 1× Expand polymerase buffer, and 1.7 units of Expand High Fidelity Polymerase (Roche Diagnostics, Indianapolis). PCR conditions were 2 min at 94°C, followed by 10 cycles of 15 s at 94°C, 30 s at 50°C, 2 min at 68°C, and an additional 15 cycles of 15 s at 94°C, 30 s at 50°C, and 140 s at 68°C. Reactions were terminated by an additional 7 min at 68°C. Products were separated on 1% (v/v) agarose gels and stained with ethidium bromide. Bands were visualized and captured using the Stratagene Eagle Eye II Imager and quantified using the EagleSight software. Intensities were expressed relative to the intensity of the amplified actin band.

In Situ Gene Expression

For in situ RT-PCR, the method was similar to that developed by Chen and Fuggle (1997) and modified and described in detail by Ruiz-Medrano et al. (1999). Hand sections of fresh tissue from roots of 6-d-etiolated corn seedlings were used. Primers used were as described in the previous section for *ZmCesA-1* and *ZmCesA-8*. Sections were examined and digital images were created using a Leica (Wetzlar, Germany) confocal microscope.

In situ hybridization was carried out on sections of roots of 6-d-etiolated maize seedlings using the procedure of Dietrich et al. (1989). Forward and reverse primers for PCR generation of the radioactive probe were based upon the HVR-2 region of the *ZmCesA-8* gene and were: 5′-ATC CGG AAT TCA AAA CGA AGA AGC CAC CAT C-3′ and 5′-TCG TCT AGA AGA CTG CCC AAA TTT CTT CTC-3′. The fragment amplified was cloned into pBlueScript, and RNA probes were generated using either T3 or T7 polymerases labeled with [^{35}S]UTP.

Tissue printing to study expression of *ZmCesA-8* in developing stalks of corn was carried out by the basic method described by Ye et al. (1992). Printing was done using cross-cuts of young, developing stalks of greenhouse-grown plants in regions above the intercalary meristem. Blotting of the stems was onto MagnaGraph nylon transfer membranes (Osmonics, Minnetonka, MN) that were baked in a vacuum oven at 80°C for 2 h. The radioactive probe was generated by PCR using the *ZmCesA-8* cDNA as template with primers described above for the HVR-2 region using [^{32}P]dCTP as label. Hybridization and washing were as in Ye et al. (1992). Tissue prints were exposed for radioautography followed by staining with India ink.

ACKNOWLEDGMENTS

Significant acknowledgment goes to Xun Wang, who led the initial efforts at Pioneer-Hibred with researchers at Human Genome Sciences to develop the cDNA libraries along with their characterization, and to Robin Wineland, who performed the mapping analyses. For work at the University of California (Davis), we are grateful to Robert Kuzoff for help with PAUP analyses, to Pat Hogan, Ravit Eshed, Neelima Sinha, and Kelley Yee for advice and/or technical assistance, and to Samantha Barling-Silver for help with graphics.

Received March 16, 2000; accepted April 10, 2000.

LITERATURE CITED

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Ploymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter C (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science **252:** 1651–1656

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **5:** 403–410

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

Brown RM Jr, Saxena IM, Kudlicka K (1997) Cellulose biosynthesis in higher plants. Trends Plant Sci **1:** 149–156

Buckeridge MS, Vergara C, Carpita NC (1999) The mechanism of synthesis of a mixed-linkage (1,3),(1,4)-D-glucan in maize: evidence for multiple sites of glucosyl transfer in the synthase complex. Plant Physiol **120:** 1105–1116

Campbell JA, Davies GJ, Bulone V, Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J **326:** 929–942

Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. Annu Rev Plant Physiol Plant Mol Biol **47:** 445–476

Charnock SJ, Davies GJ (1999) Structure of the nucelotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. Biochemistry **38:** 6380–6385

Chen RH, Fuggle SV (1997) In situ PCR amplification of intercellular mRNA. Methods Mol Biol **71:** 123–132

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J **16:** 735–743

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

Delmer DP (1999) Cellulose synthesis: exciting times for a difficult field. Annu Rev Plant Physiol Plant Mol Biol **50:** 245–276

Dietrich RA, Maslyar DJ, Heupel RC, Harada JJ (1989) Spatial patterns of gene expression in *Brassica napus* seedlings: identification of a cortex-specific gene and localization of mRNAs encoding isocitrate lyase and a polypeptide homologous to proteinases. Plant Cell **1:** 73–80

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegman S (1985) Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol Biol **5:** 109–118

Helentjaris T, Weber D, Wright S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics **118:** 353–363

Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown RM Jr (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis.* Plant Cell **11:** 2075–2086

Kawagoe Y, Delmer DP (1997) Pathways and genes involved in cellulose biosynthesis. Genet Eng **19:** 63–87

Kawagoe Y, Delmer DP (1998) Recent progress in the field of cellulose synthesis. Trends Glycosci Glycotechnol **10:** 291–305

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol **157:** 105–132

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1:** 174–181

Michelmore RW, Meyers BC (1998) Cluster of resistance genes in plants evolve by divergent selection and a birth-and-death process. Genome Res **8:** 1113–1130

Nilsson T, Warren G (1994) Retention and retrieval in the endoplasmic reticulum and the Golgi apparatus. Curr Opin Cell Biol **6:** 517–521

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

Ruiz-Medrano R, Xoconostle-Cazares B, Lucas WJ (1999) Phloem long-distance transport of *CmNACP-1* mRNA: implications for supracellular regulation in plants. Development **126:** 4405–4419

Taylor NG, Scheible W, Cutler S, Somerville CR, Turner SR (1999) The *irregular xylem3* locus of Arabidopsis encodes a cellulose synthase required for secondary wall synthesis. Plant Cell **11:** 769–779

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

Ye Z-H, Song Y-R, Varner JE (1992) Gene expression in plants. *In* PD Reid, RF Pont-Lezica, eds, Tissue Printing. Academic Press, San Diego, pp 95–123