Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 9693–9698, August 1998
Plant Biology

Analysis of xylem formation in pine by cDNA sequencing

ISABEL ALLONA*†, MICHELLE QUINN*, ELIZABETH SHOOP‡, KRISTI SWOPE‡, SHEILA ST. CYR‡, JOHN CARLIS‡, JOHN RIEDL‡, ERNEST RETZEL‡, MALCOLM M. CAMPBELL*§, RONALD SEDEROFF*, AND ROSS W. WHETTEN*

*Forest Biotechnology Group, Department of Forestry, North Carolina State University, Raleigh, NC 27695-8008; and ‡Computational Biology Center, University of Minnesota, Minneapolis, MN 55455-0312

Contributed by Ronald Sederoff, June 10, 1998

ABSTRACT **Secondary xylem (wood) formation is likely to involve some genes expressed rarely or not at all in herbaceous plants. Moreover, environmental and developmental stimuli influence secondary xylem differentiation, producing morphological and chemical changes in wood. To increase our understanding of xylem formation, and to provide material for comparative analysis of gymnosperm and angiosperm sequences, ESTs were obtained from immature xylem of loblolly pine (*Pinus taeda* L.). A total of 1,097 single-pass sequences were obtained from 5′ ends of cDNAs made from gravistimulated tissue from bent trees. Cluster analysis detected 107 groups of similar sequences, ranging in size from 2 to 20 sequences. A total of 361 sequences fell into these groups, whereas 736 sequences were unique. About 55% of the pine EST sequences show similarity to previously described sequences in public databases. About 10% of the recognized genes encode factors involved in cell wall formation. Sequences similar to cell wall proteins, most known lignin biosynthetic enzymes, and several enzymes of carbohydrate metabolism were found. A number of putative regulatory proteins also are represented. Expression patterns of several of these genes were studied in various tissues and organs of pine. Sequencing novel genes expressed during xylem formation will provide a powerful means of identifying mechanisms controlling this important differentiation pathway.**

Wood, a tissue unique to higher plants, contains a large fraction of the fixed carbon stored in the global ecosystem (1). Accumulation of secondary xylem, as wood in trees, is both ecologically and economically important. Trees dominate many ecosystems and exploit unique ecological niches. Wood is also a raw material for a major global industry. Despite the importance of wood, little is known of the genes involved in wood formation. The physical and chemical properties of wood are derived directly from the composition and morphology of the cell walls that are the substance of wood (mature xylem). The composition and morphology of the wood cell wall depend on the processes of biosynthesis and assembly. These processes are thought to be directed by specific genes that control macromolecular synthesis and determine structure (2). Some of these genes may be only rarely expressed in other plant tissues.

The ability of plants to produce different phenotypes under different environmental conditions has been the subject of several evolutionary and ecological studies (3). Morphological and chemical characteristics of wood can be modified dramatically by environmental influence. When a tree leans or twists, reaction wood is formed in response to gravitational and mechanical stimuli. In gymnosperms reaction wood takes the form of compression wood, formed on the underside of the tree, which is characterized by lower cellulose content and higher lignin content enriched in *p*-hydroxyphenyl subunits. Tracheid length is reduced, the cross-sectional profile becomes rounded, and the intercellular spaces become larger (4, 5). The wood formed on the upper side of the inclined stem (opposite wood) may also show some changes in cell wall structure and chemistry, but wood from the sides of the inclined stem is similar to wood from vertical stems in structure and cell wall chemistry (4). Changes in gene expression are likely to underlie the morphological and chemical differences observed between compression wood and normal wood (6).

Sequencing of cDNAs isolated from specialized tissues and organs has become a useful tool for identifying new genes (7–10). The majority of plant expressed sequence tag (EST) projects, however, have been carried out by using a mixture of organs and tissues from herbaceous angiosperms species. Early work on the sequence complexity of RNA pools in various organs of tobacco indicated that individual organs contain more than 6,000 mRNAs not present in other organs of the same plant (11). Sequencing of immature xylem cDNAs from a tree species would increase our knowledge about the mechanisms involved in wood formation.

Pinus taeda, the most important commercial softwood species in the United States, is the subject of ongoing molecular and quantitative genetic experiments in many laboratories (12, 13). We have sampled expressed genes from differentiating xylem forming compression and side wood, and also used subtraction techniques to isolate genes that may be differentially expressed between these two tissues. The use of stressed tissues could help us to find transcripts that might be difficult to find in normal tissue (14). In this paper, we present the results of first-pass sequence analysis for 1,097 sequences from differentiating xylem of loblolly pine. We identified 833 different expressed sequences, including many not yet identified in available databases, and many recognized to be involved in cell wall formation. This work will support research into the mechanisms of wood formation, supply many new cloned genes for genetic engineering, and increase our knowledge of the environmental influence on wood properties. Comparison of pine ESTs with sequences from angiosperms will add important information about the evolution of higher plants.

MATERIALS AND METHODS

Plant Material. Three 6-year-old loblolly pine (*Pinus taeda* L.) trees of different genotypes were induced to form com-

Abbreviations: CA, compression antisense; NA, normal antisense; XET, xyloglucan endotransglycosylases.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AA556146–AA557120 and AA566922–AA567040).

†To whom reprint requests should be addressed at the present address: Departamento de Biotecnologia, Escuela Tecnica Superior de Ingenieros de Montes, Universidad Politecnica de Madrid, Ciudad Universitaria s/n, Madrid 28040, Spain. e-mail: iallona@etsi.montes.upm.es.

§Present address: Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, England.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

pression wood by bending to a 45° angle and tying them to the ground during the growing season (May). Differentiating xylem was harvested separately from the bottoms and from the sides of the inclined stems at 3, 15, and 40 days after bending, frozen immediately in liquid nitrogen, and stored at −80°C until use. Compression wood formation was apparent upon visual examination of the undersides of the inclined stems, even at the earliest sampling time of 3 days. Samples of differentiating compression wood and side wood were taken from the same trees to minimize the effects of genetic or environmental variation on differential gene expression between these two tissue types. Previous work has demonstrated that genetic variation between individual trees can confound the interpretation of biochemical and molecular genetic experiments in loblolly pine (15, 16).

Tissue and organ samples for Northern blot analysis of gene expression patterns were taken during the growing season and frozen in liquid nitrogen, then stored at −80°C. Shoot-tip samples are the terminal 1–2 cm of succulent stem tissue from elongating branches and terminal leaders. Immature xylem was harvested with a vegetable peeler from the surface of woody stems after removal of the bark, and includes cells with primary cell walls derived from the cambial zone through the zone of radial cell expansion. Compression wood and side wood immature xylem RNAs were those used for preparation of the libraries, whereas vertical immature xylem denotes samples taken from stems of control trees growing vertically. The tissue described as "phloem" was collected from the inner surface of the bark of vertical stems and actually includes periderm and other tissues as well as active phloem. Planings were collected from the surface of vertical woody stems with a wood plane after immature xylem was harvested and includes cells that have formed secondary cell walls strong enough to resist harvesting by the vegetable peeler. Needle samples were partially expanded juvenile needles.

RNA Isolation and Northern Analysis. Total RNA for cDNA library construction was prepared following the method of Chang *et al*. (17). mRNA isolation was carried out with PolyATtract Systems IV (Promega), according to the manufacturer's instructions. Total RNA for Northern blots was separated in 1.2% agarose gels with 2.2 M formaldehyde and transferred to Zetaprobe nylon membranes (Bio-Rad). cDNA probes were ^{32}P-radiolabeled with Ready-To-Go Labeling Beads (Pharmacia) and used as probes. The prehybridization and hybridization in the presence of 50% deionized formamide were carried out according to published procedures (18). Membranes were washed at high stringency and used to expose x-ray film or PhosphorImager screens (Molecular Dynamics).

cDNA Library Construction. Two cDNA libraries were constructed by using the Stratagene cDNA cloning kit (lambda ZAP II vector). The side wood and compression wood cDNAs were directionally cloned with the 5′ extreme at the T3 side of the polylinker. These libraries were named N (morphologically normal differentiating side wood) and C (differentiating compression wood).

We produced two subtraction libraries to enrich for specific clones involved in either compression or side wood formation, using the method of Hesse *et al.* (19). Briefly, this method uses a lambda ZAP vector with a modified polylinker (lambda PAZ), such that the orientation of directionally cloned cDNAs is opposite between lambda ZAP and lambda PAZ. Mass excision of single-stranded phagemid DNA from libraries made in these two vectors yields vector DNA of the same orientation, but cDNA inserts of complementary orientations. The single-stranded DNA from the PAZ library is photobiotinylated and used to subtract complementary sequences from the ZAP library. The same N and C cDNA preparations were ligated into the vector lambda PAZ II (kindly provided by H. Hesse, Institut für Genbiologische Forschung, Berlin, Germany), the plasmids were mass-excised and photobiotinylated and then used to subtract sequences from mass-excised N and C libraries. The subtracted library enriched in transcripts from the N library was called CA (compression antisense) and the subtracted library enriched in transcripts from the C library was called NA (normal antisense).

Nucleotide Sequencing. Randomly chosen single-phage plaques were picked from the N and C primary libraries and from the NA and CA subtracted libraries, and the pBluescript plasmid was individually excised following the manufacturer's protocol (Stratagene). DNA plasmid miniprep were prepared by using the AGCT 96-well DNA Miniprep Kit (Advanced Genetic Technology Corporation). Each plasmid was named according to the 96-well culture plate, the library of origin, and the well identity, e.g., clone 7C8A is from well 8A of plate 7 of the C library.

Sequencing was carried out on plasmid DNA preparations, in the sequencing facilities of Iowa State University and the University of Missouri on automated sequencers (Applied Biosystems). Partial sequence from the 5′ end of each clone was obtained.

Sequence Processing and Analysis. The following processing and analysis steps were performed on each sequence by using computational tools developed at the University of Minnesota, Computational Biology Center (20, 21).

Sequence artifact removal consisted of detecting leading and trailing vector in the clone sequence and trimming it off. The quality checks were: (*i*) determine how many unknown or "N" base calls were in the sequence and trim leading and trailing high-N sections to obtain the best subsequence where the number of Ns was 4% or less of the total number of base pairs, and (*ii*) tag each short (<50-bp) sequence so that further processing was discontinued.

The compositional complexity analysis was as follows. (*i*) Translate each resulting high-quality DNA sequence into the possible peptide sequences for all six reading frames and scan those peptide sequences for regions of low compositional complexity (LC regions). LC regions can affect similarity search results by producing similarities of high mathematical probability but low probability of being biologically meaningful (22). (*ii*) If LC regions were found, they were masked out of the translated peptide sequence.

The similarity analysis for each sequence used the BLAST database search programs (23). The steps in this analysis were: (*i*) execute a BLASTN search of each sequence against GenBank; (*ii*) execute a BLASTX search of each sequence against a peptide sequence database composed of the National Center for Biotechnology Information (NCBI) GenPept database, release 97.0, and the Protein Information Resource (PIR) protein sequence database, release 49.0 (24); (*iii*) for each translated protein sequence containing LC regions that are masked out, execute a BLASTP search against the combined GenPept plus PIR protein sequence database.

The results from the similarity analysis were transformed from text files into a format suitable for storage in a database management system (DBMS). The data were placed into the DBMS and analyzed by using Structured Query Language (SQL) queries. Each analysis was carried out via SQL queries on the DBMS.

The results of the processing and analysis for each sequence are stored in a set of World Wide Web pages, one for each sequence. These pages can be searched collectively by using words or phrases of interest. In addition, users can conduct BLASTN or TBLASTN searches of a DNA or protein sequence against all of the cDNA sequences. These tools can be accessed via the University of Minnesota WWW site at http://www.cbc.umn.edu/ResearchProjects/Pine.

Score, *P* value, and percent identities from the BLASTX results were considered when determining significant similarities. Strong similarities, or hits, are those with scores greater than 150, *P* values less than 0.005, and identities greater than

or equal to 40%. Marginal hits are those that did not qualify as strong hits, but had scores greater than 80, *P* values less than 0.010, and identities greater than or equal to 30%.

The partial cDNA sequences were clustered together based on their similarity to each other, using BLASTN searches. Each similarity with a score greater than 800 and a *P* value less than 0.001 was considered a strong hit, and each marginal hit had a score between 600 and 800 and a *P* value less than 0.001. Both strong and marginal hits were used to cluster the sequences together.

RESULTS AND DISCUSSION

cDNA Libraries and Nucleotide Sequencing. Immature xylem from bent trees was used to construct the cDNA libraries. Directionally cloned lambda ZAP II libraries were made from compression and morphologically normal side wood, designated C and N. Both primary libraries contained 1×10^6 recombinant phage and, therefore, were considered to have an adequate representation of the expressed genes. An average insert size of 1.5 kb was calculated based on 60 randomly picked clones. The presence of poly(A) tails in all of these clones was confirmed by sequencing the 3′ ends, as a test of the quality of the libraries.

A subtraction procedure was used to obtain more specific or rare clones. We produced two subtraction libraries, one enriched in compression transcripts (NA) and the other enriched in normal transcripts (CA). The method used, described by Hesse *et al.* (19), allows verification of the origin of subtracted clones, because a failure of subtraction results in cDNA clones that are in the opposite orientation relative to the vector. The numbers of clones sequenced from each library are C = 577, NA = 72, N = 324, CA = 124. The clone names indicate the culture plate, the library, and the individual well location of each clone.

Partial DNA sequences were determined from the 5′ end of 1,097 clones, randomly picked from the primary libraries or the subtracted libraries. All the sequences were trimmed with a computer program to remove sequencing artifacts and vector sequence as described in *Materials and Methods*. The average sequence length was 510 nt, after trimming.

Characterization of the EST Sequences. A current analysis of the immature xylem loblolly pine database sequences can be obtained from the World Wide Web server at http://www.cbc.umn.edu/ResearchProjects/Pine. These sequences have been submitted to the NCBI database of ESTs (dbEST). The plasmids corresponding to the ESTs have been deposited in the American Type Culture Collection (Manassas, VA), and ATCC accession numbers are given in the dbEST records for the sequences.

Similarity searches of the pine cDNA sequences against each other allowed clustering using criteria described in *Materials and Methods*. The cluster analysis revealed that 736 of the 1,097 sequences were unique or nonredundant, whereas 361 were members of 107 distinct clusters (table 1 in web page). Because the clones were picked from primary libraries and were excised individually, the clusters represent either independent clones of the same transcript, allelic sequences, or different members of multigene families. The largest cluster contained 20 sequences, and more than 60 of the clusters contained only 2 sequences. The number of unique sequences may be overestimated, because some sequences could be nonoverlapping regions of the same cDNA.

Among the 1,097 sequences, 460 (42%) showed strong similarity to at least 1 sequence in the public databases, and another 148 (13%) showed marginal similarity (table 2 in web page). In some cases the putative function of a sequence could not be inferred from similarity analysis, but the cluster analysis placed the clone in a group with others that had either strong or marginal scores. The combination of cluster analysis within the pine EST dataset and similarity searching of GenBank increased the total number of sequences for which we can infer a putative function to 651 (59%). The genome structures of extant species suggest that conifer and angiosperm genomes have evolved by different mechanisms (25). Angiosperms are believed to have diverged from conifers about 270–300 million years ago (26). This large evolutionary distance makes difficult the isolation of pine homologues of angiosperm genes by hybridization with heterologous probes or by PCR using degenerate primers, but does not seem to limit our ability to recognize similarity in ESTs.

Sequences that have neither strong nor marginal scores may represent nonconserved regions of known genes or completely new genes. Sequences that show strong similarity average 565 nt in length, those with marginal similarity have an average length of 533 nt, and the average length of those with no significant similarity is 447 nt. This suggests that length and quality of cDNA sequence are correlated with the ability to identify similar sequences in public databases. However, many long, high-quality sequences show neither strong nor marginal similarity to sequences in the database. Full-length sequence of these clones will be needed for more complete similarity searches.

Although a putative function can be deduced from sequence similarity, the true function can only be verified by biochemical and genetic approaches (14). Different approaches exist to test the function of pine xylem genes. For example, mutant analysis in pine is favored by the high frequency of mutant alleles found in natural populations (27). A cell culture system using *Zinnia elegans* mesophyll cells, which can trans-differentiate into tracheary elements *in vitro*, would constitute a good system to test the function of xylem related genes if an efficient transformation system were developed (28).

Putative Genes in the Pine Xylem Dataset. The largest group of sequences with putative functions, assigned by results of similarity searches, are genes encoding proteins important in plant cell wall formation. This is not unexpected, because pine xylem is characterized by massive cell walls. The use of cDNA libraries made from tissues or organs specialized for metabolic processes of interest has been reported previously (10). Structural proteins present in the plant cell wall, and enzymes involved in production of cell wall polymers, comprise more than 10% of the sequences with putative functions. Cell wall associated carbohydrate metabolism proteins are represented with 23 sequences, including cellulases and xyloglucan endotransglycosylases. There are 50 clones corresponding to cell wall structural proteins, including extensin-like proteins (1 cluster of 20 sequences) along with proline-rich proteins, arabinogalactan-like proteins, and glycine-rich proteins. Work is now in progress to isolate representative full-length clones for each class of cell wall protein and to purify the corresponding proteins from pine xylem cell walls. Robertson *et al.* (29) analyzed N-terminal amino acid sequences of cell wall proteins isolated from suspension-cultured cells of five angiosperm species and concluded that "a significant proportion of wall proteins have not been previously described."

Most of the genes known to be involved in the lignin biosynthetic pathway also are represented, including PAL, C4H, OMT, 4CL, and CAD (30). There are six clones with similarity to diphenol oxidase (laccase) and only one clone similar to peroxidase, both types of enzymes thought to be involved in lignification. The difference in apparent abundance of these two classes of enzymes suggests that diphenol oxidases may be more important than peroxidases in lignin biosynthesis in pine xylem (31, 32). Only one of the 16 clones similar to genes involved in lignin biosynthesis was from the N library; all others were from the C or NA libraries. Compression wood is known to have higher levels of lignin and of lignin biosynthetic enzymes than side wood or normal wood [reviewed by Timell (4)].

Two other large clusters represent enzymes involved in amino acid metabolism. Transcripts encoding methionine synthase (16 sequences in 1 cluster) and glycine hydroxymethyltransferase (9 sequences in 2 clusters, plus 2 singlets) are abundant. Both of these sequences occur at about 10-fold higher frequency in the pine xylem dataset than in the TIGR *Arabidopsis* EST assembly database (http://www.tigr.org/tdb/at/atest.html). *Arabidopsis* ESTs similar to methionine synthase or glycine hydroxymethyltransferase occur at frequencies of about 0.1% in the TIGR database, whereas pine ESTs similar to these two enzymes occur in the pine xylem dataset at frequencies greater than 1%. Methionine is involved in methyltransferase reactions as S-adenosylmethionine, whereas glycine hydroxymethyltransferase participates in methyl transfer reactions through regeneration of 5,10-methylene tetrahydrofolate catalyzing the interconversion between serine and glycine. These two cofactors are required for a number of biosynthetic reactions in plant cells, and the abundance of these sequences in the pine xylem dataset may reflect the demand for methyl transfer reactions in differentiating xylem. Also, glycine is one of the most abundant amino acids in cell wall structural proteins. Alternatively, these proteins may have other functions. Methionine synthase, for example, has been suggested to play a role in protein folding or signal transduction in *Chlamydomonas* (33).

Genes involved in regulation of cellular metabolism are also represented abundantly in this pine EST dataset. There are several clones similar to protein kinases, and several transcription factors such as MADS box, zinc finger proteins, homeodomain, and LIM domain proteins. These proteins could be involved either in the normal development of xylem, or in the response to a specific stress. Plants can be subjected to a large number of stresses, such as heat or cold shock, drought or water stress, or attack by a wide variety of pathogens, which induce the transcription of genes whose expression is specific to these conditions (14). In our studies, for the induction of compression wood, the environmental stimulus is mechanical and gravitational. Many genes involved in signal transduction and regulation of gene expression are expressed at a very low level, and these genes may be hard to find even by using a normalized library (14).

Differential Expression of Genes in Different Tissues and Organs of Loblolly Pine. Northern blots were used to study the relative abundance of transcripts corresponding to sets of clones from this study. One set potentially is involved in cell wall structure and metabolism, whereas another set is related to regulatory proteins. The objective of these experiments was to gain additional information about the abundance of transcripts in various tissue and organs of pine and to test the hypothesis that sequencing in a highly specialized tissue would yield a high frequency of clones corresponding to differentially expressed genes. Most clones do detect widely differing levels of transcripts in the different RNA samples tested, although the highest transcript levels detected by an individual clone are not always found in the tissue from which that clone was isolated.

The cell wall-related clones include two different members of the xyloglucan endotransglycosylase family and two putative cell wall proteins (Fig. 1). Xyloglucan endotransglycosylases (XET) are enzymes thought to participate in restructuring of cell wall crosslinks, and a family of related genes encoding these enzymes has been identified and characterized in *Arabidopsis* and other herbaceous plants (34–36). A total of four different putative pine XET were identified. The two XET clones chosen for testing were the most divergent from the known angiosperm sequences. The 1C8A transcript is more abundant in vertical xylem than in side wood, whereas the 3N5D transcript is more abundant in side wood than in vertical xylem (Fig. 1*B*). The differences between these genes in patterns of transcript abundance suggest that these two members of the XET family have specialized functions in modification of xylem cell wall structure. Two putative cell wall proteins were also analyzed. One of these proteins is similar to proline-rich proteins and the other is similar to glycine-rich proteins (37–39), but both pine proteins differ from the most similar angiosperm counterpart in the repeat patterns of predicted polypeptides. These cDNAs also detect different patterns of transcript abundance, suggesting functional specialization of the corresponding proteins (Fig. 1*C*). Both cDNAs detect transcripts that are relatively more abundant in side wood than in other tissues, although the 6N5F cDNA also detects abundant transcripts in the "phloem" RNA sample, suggesting that this protein may be involved both in the stress response of xylem and in the normal development of phloem or related tissues.

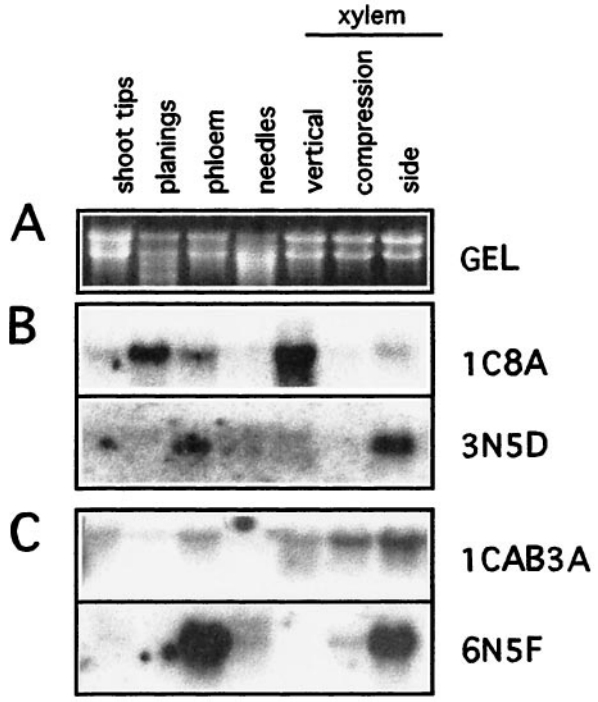

FIG. 1. Northern blot results using cell wall-related cDNA clones identified during this sequencing project. (*A*) Total RNA samples separated in formaldehyde-agarose gels and stained with ethidium bromide to estimate total loading. (*B*) Results of hybridization of the immobilized total RNA samples to clones 1C8A and 3N5D, which encode different putative xyloglucan endotransglycosylases (XET). (*C*) Results of hybridization of total RNA samples to two cDNA clones similar to cell wall proteins. 6N5F is more similar to glycine-rich cell wall proteins, whereas 1CAB3A is more similar to proline-rich cell wall proteins.

Several cDNAs that show similarity to regulatory proteins show tissue specificity (Fig. 2). Three of these cDNAs are related to proteins involved in regulation of calcium homeostasis or cellular responsiveness to calcium; these clones are of interest because calcium and calcium-binding proteins are implicated in plant responses to gravity (40, 41). Clones similar to calreticulin and calnexin detected transcripts in all samples, although somewhat higher levels are present in the compression wood and side wood samples, suggesting a role for these proteins in the stress response of immature xylem. A clone similar to calcium-dependent protein kinase detected much higher levels of transcripts in needle RNA than in any other tissue, suggesting that this particular gene may play a more important role in needles than in differentiating xylem. The other three cDNAs, 5N3D, 5C7A, and 6C12H, show similarity to proteins containing either homeodomain or LIM-domain regions (Fig. 2*C*). Homeodomains are implicated in DNA-

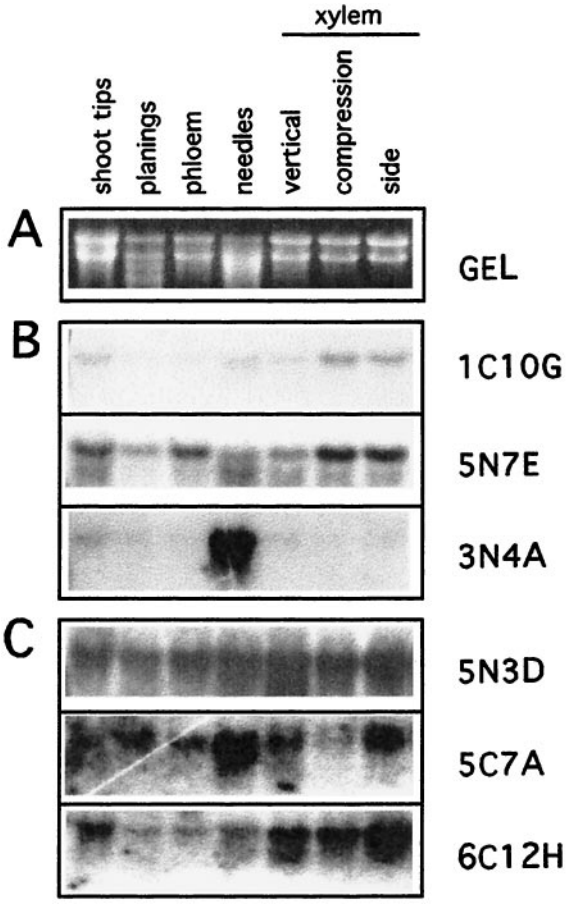

FIG. 2. Northern blot results using cDNAs similar to regulatory proteins. (*A*) The same total RNA stained with ethidium bromide, provided to allow lane identification. (*B*) Results of hybridization with clones similar to calcium-related proteins. 1C10G is similar to calreticulin, 5N7E is similar to calnexin, and 3N4A is similar to calcium-dependent protein kinase. (*C*) Results of hybridization with clones similar to homeodomain or LIM domain proteins. 5N3D and 5C7A are similar to homeodomain proteins, and 6C12H is similar to LIM domain proteins.

binding activity in a wide variety of animal and plant species (42, 43). The LIM-domain is a specialized form of zinc finger that has been implicated in protein–protein interactions and signal transduction in animal systems (44, 45), although few representatives of this protein family have been characterized extensively in plants. These cDNAs were chosen for Northern analysis to test the hypothesis that regulatory proteins obtained from specialized tissue might show tissue-specific patterns of transcript accumulation. The 5N3D clone (homeodomain-related) detected transcripts in all RNA samples, whereas the 5C7A (homeodomain-related) and 6C12H (LIM-related) clones showed a less uniform pattern of transcript accumulation. The patterns of message abundance are different for all 10 clones tested, and in many cases the highest levels of message are not in the tissue from which the clone was originally isolated.

CONCLUSIONS

The use of a cDNA library of a specific tissue, immature pine xylem, has proven to be an excellent source of genes to study wood formation in conifers. The large number of cDNAs found in the pine xylem dataset that encode putative cell wall structural proteins will provide an opportunity to study different classes of cell wall proteins in a single tissue of a single species. The biosynthetic and hydrolytic enzymes that are up-regulated during xylem formation are all potential commercial targets, and promoter analyses of the relevant genes may identify useful xylem-specific promoters to reveal the signal transduction cascade that leads to xylem formation (28). We conclude that cDNA sequencing in pine is a useful approach to identifying pine homologues of known genes, as well as discovering novel genes not previously described in other organisms. Many sequences can be identified as ESTs but would have been difficult to obtain by hybridization with heterologous probes or homologous probes prepared by PCR using degenerate primers.

The authors thank Dr. Holger Hesse for providing the lambda PAZ II vector, Reenah Schaffer for assistance with figures, and the Iowa State University and University of Missouri DNA Sequencing Facilities for DNA sequencing. Isabel Allona was supported by Fulbright fellowship FU94-05378494, and research support was provided by U.S. Department of Agriculture Grant 95-373000-1591 and the National Aeronautics and Space Administration Specialized Center of Research and Training in Gravitational Biology at North Carolina State University.

1. Dixon, R. K., Brown, S., Houghton, R. A., Solomon, A. M., Trexler, M. C. & Wisniewski, J. (1994) *Science* **263,** 865–872.
2. Higuchi, T. (1997) *Biochemistry and Molecular Biology of Wood*, ed. Timell, T. E. (Springer, Heidelberg).
3. Pigliucci, M. (1996) *Tree* **11,** 168–173.
4. Timell, T. E. (1986) *Compression Wood in Gymnosperms* (Springer, Heidelberg).
5. Fukushima, K. & Terashima, N. (1991) *Wood Sci. Technol.* **25,** 371–381.
6. Zhang, X.-H. & Chiang, V. L. (1997) *Plant Physiol.* **113,** 65–74.
7. Adams, M. D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M., Utterback, T. R., Nagle, J. W., Fields, C. & Venter, C. (1992) *Nature (London)* **355**, 632–634.
8. Hofte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M., Moisan, A., Jourjon, M. F., Charpenteau, J. L., Berthomieu, P., Guerrier, D., *et al.* (1993) *Plant J.* **4,** 1051–1061.
9. Lim, C. O., Kim, H. Y., Kim, M. G., Lee, S. L., Chung, W. S., Park, S. H., Hwang, I. & Cho, M. J. (1996) *Plant Physiol.* **111,** 577–588.
10. Van de Loo, F. J., Turner, T. & Somerville, C. (1995) *Plant Physiol.* **108,** 1141–1150.
11. Kamalay, J. C. & Goldberg, R. B. (1980) *Cell* **19,** 935–946.
12. Ahuja, M. R., Boerjan, W. & Neale, D. B. (1996) *Somatic Cell Genetics and Molecular Genetics of Trees* (Kluwer, Dordrecht, The Netherlands).
13. Neale, D. B. & Sederoff, R. R. (1996) in *Genome Mapping in Plants*, ed. Paterson, A. H. (R. G. Landes Company and Academic, San Diego), pp. 309–319.
14. Cooke, R., Raynal, M., Laudie, M., Grellet, F., Delseny, M., Morris, P. C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G., *et al.* (1996) *Plant J.* **9,** 101–124.
15. Voo, K. S., Whetten, R. W., O'Malley, D. M. & Sederoff, R. R. (1995) *Plant Physiol.* **108,** 85–97.
16. O'Malley, D. M., Porter, S. & Sederoff, R. R. (1992) *Plant Physiol.* **98,** 1364–1371.
17. Chang, S., Puryear, J. & Cairney, J. (1993) *Plant Mol. Biol. Rep.* **11,** 113–116.
18. Brown, T. & Mackey, K. (1997) in *Current Protocols in Molecular Biology,* eds. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Siedman, J. G., Smith, J. A. & Struchl, K. (Wiley, New York), pp. 4.9.1–4.9.12.
19. Hesse, H., Frommer, W. B. & Willmitzer, L. (1995) *Nucleic Acids Res.* **23,** 3355–3356.
20. Shoop, E., Chi, E., Carlis, J., Bieganski, P., Riedl, J., Dalton, N., Newman, T. & Retzel, E. (1995) *Proceedings of the 28th Annual Hawaii International Conference on System Sciences* **5,** 52–61.
21. Shoop, E. (1996) Ph.D. thesis (University of Minnesota).
22. Claverie, J. M. & States, D. (1993) *Comput. Chem.* **17,** 191–201.
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. (1990) *J. Mol. Biol.* **215,** 403–410.
24. Barker, W. C., George, D. G., Mewes, H. W., Pfeiffer, F. & Tsugita, A. (1993) *Nucleic Acids Res.* **21,** 3089–3092.
25. Kinlaw, C. S. & Neale, D. B. (1997) *Trends Plant Sci.* **2,** 356–359.

26. Savard, L., Li, P., Strauss, S. H., Chase, M. W., Michaud, M. & Bousquet, J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5163–5167.
27. Allendorf, F. W., Knudsen, K. L. & Blake, G. M. (1982) *Genetics* **100,** 497–504.
28. McCann, M. C. (1997) *Trends Plant Sci.* **2,** 333–338.
29. Robertson, D., Mitchell, G. P., Gilroy, J. S., Gerrish, C., Bolwell, G. P. & Slabas, A. R. (1997) *J. Biol. Chem.* **272,** 15841–15848.
30. Whetten R. & Sederoff, R. (1995) *Plant Cell* **7,** 1001–1013.
31. Bao, W., O'Malley, D. M., Whetten, R. & Sederoff, R. R. (1993) *Science* **260,** 672–674.
32. O'Malley, D. M., Whetten, R., Bao, W., Chen, C.-L. & Sederoff, R. R. (1993) *Plant J.* **4,** 887–892.
33. Kurvari, V., Qian, F. & Snell, W. J. (1995) *Plant Mol. Biol.* **29,** 1235–1252.
34. Medford, J. I., Elmer, J. S. & Klee, H. J. (1991) *Plant Cell* **3,** 359–370.
35. Xu, W., Purugganan, M. M., Polisensky, D. H., Antosiewicz, D. M., Fry, S. C. & Braam, J. (1995) *Plant Cell* **7,** 1555–1567.
36. Xu, W., Campbell, P., Vargheese, A. K. & Braam, J. (1996) *Plant J.* **9,** 879–889.
37. Keller, B. (1993) *Plant Physiol.* **101,** 1227–1130.
38. Showalter, A. M. (1993) *Plant Cell* **5,** 9–23.
39. Kieliszewski, M. J. & Lamport, D. T. A. (1994) *Plant J.* **5,** 157–172.
40. Stinemetz, C. L., Kuzmanoff, K. M., Evans, M. L. & Jarrett, H. W. (1987) *Plant Physiol.* **84,** 1337–1342.
41. Young, L. M. & Evans, M. L. (1994) *Plant Growth Reg.* **14,** 235–242.
42. Schena, M. & Davis, R. L. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3894–3898.
43. Baima, S., Nobili, F., Sessa, G., Lucchetti, S., Ruberti, I. & Morelli, G. (1995) *Development* **121,** 4171–4182.
44. Sanchez-Garcia, I. & Rabbitts, T. H. (1994) *Trends Genet.* **10,** 315–320.
45. Arber, S., Halder, G. & Caroni, P. (1994) *Cell* **79,** 221–231.