DNA SEQUENCE DIVERSITY IN ALCOHOL DEHYDROGENASE GENES FROM PINES

D. E. Hare [1/], K. S. Mordecai [1/], C. S. Kinlaw [2/], C. A. Loopstra[3/], and R. R. Sederoff [3/]

Abstract.--Molecular clones that represent ADH coding sequences (cDNA clones) from Monterey pine have been been isolated and characterized. These clones have been subsequently used to identify DNA fragments containing homologous sequences in samples of DNA prepared from needles of mature loblolly and Monterey pine trees. Relative to analogous studies of ADH from angiosperms, the number of fragments in pines is large. Just how these homologous sequences are related to each other is currently under investigation. Genomic libraries have been constructed in a bacteriophage vector (EMBL3) for both loblolly and Monterey pines. Several million recombinant phage were screened using our ADH cDNA clones as molecular probes. We identified a large number of clones containing sequences homologous to our probes, a result that is consistent with our studies of restriction fragments. Preliminary analyses of DNA purified from these clones suggest that many contain DNA sequences from different portions of the genuine. In combination, these data suggest that ADH in pines may belong to a gene family that contains more members than suggested from studies of isozymes (i.e. 2-4 loci). Alternatively, pine ADH genes may themselves be very large, containing many or large intervening (non-coding) sequences in addition to a normal complement of coding sequences.

Keywords: Pinus taeda, Pinus radiata, ADH, molecular cloning, multi-gene family, genetic engineering, gene structure, genomic library.

INTRODUCTION

A basic understanding of gene structure and expression is a prerequisite for genetic engineering. Optimum utility from engineered novel (or foreign) genes that have been transferred into plants will come only if their expression can be regulated in a desired fashion. Likewise,

[1/]Department of Forestry, University of Illinois, Urbana, Illinois.
[2/] Pacific Southwest Forest and Range Experiment Station, Berkeley, California.
[3/]Department of Forestry, North Carolina State University, Raleigh, North Carolina.

endogenous genes identified as having important roles in plant growth can be subjected to genetic engineering only if their structure and expression are understood. The structure and regulation of a number of genes from angiosperms are being studied, including genes encoding alcohol dehydrogenase (ADH). We are studying the structure and expression of alcohol dehydrogenase (ADH) genes in loblolly (Pinus taeda) and Monterey (P. radiata) pines because information from angiosperms will help us to more readily understand similar properties of conifer genes.

ADH is the enzyme that catalyzes the reversible conversion of ethanol to acetaldehyde. When roots of many plants are deprived of oxygen, as would occur following flooding, ADH activity increases rapidly and dramatically (Bennett and Freeling 1988). The function of ADH under these circumstances is not clear, but may include the generation of cellular energy or the maintenance in intracellular pH (Bennett and Freeling 1988). In woody plants, ADH activity may be constitutive in stems and in leaves, perhaps because of anaerobic conditions in the vascular cambium (Kimmerer and Stringer 1988).

Much of the increased ADH activity following flooding is caused by increased transcription of ADH genes (see reviews by Freeling and Bennett 1985, Harry et al. 1988). Specific DNA sequences involved in the activation of these genes are being identified, including sequences in the promoter region (Walker et al. 1987) and in introns (Callis et al. 1987). Such studies have included transferring DNA sequences of ADH isolated from one species into evolutionarily similar and dissimilar species. For example, when regulatory sequences of ADH from corn (a monocot) were transferred into tobacco (a dicot), they failed to function properly, even though regulatory sequences of ADH from pea (another dicot) functioned properly in tobacco under identical circumstances (Howard et al. 1987, Walker et al. 1987). This has important implications for applying genetic engineering to conifers because genes that are engineered for optimal expression in angiosperms may not function properly, if at all, in gymnosperms.

ADH activity is induced by anaerobic stress in pines as it is in other plants, with the activity of one ADH gene being greatly increased relative to a second ADH gene (Harry et al. 1988, Harry et al., in preparation). As an initial step towards the characterization of pine ADH genes, we used an ADH cDNA from maize (pZML793, Dennis et al. 1984) to identify cDNA clones from Monterey pine (Harry et al. 1988, Kinlaw et al. 1988, and Kinlaw et al., in preparation). None of the identified clones contain the entire coding region from ADH. Three of the longer clones, designated RCS1025, RCS1019, and RCS1029 (hereafter abbreviated simple as 1025, 1019, and 1029) have been fully characterized (Kinlaw et al. in preparation). Based on its alignment with the maize sequence, the 531 bp insert of 1025 spans 518 bp of the protein coding region, including the ATG translational start site. The approximately 400 bp insert of 1019 overlaps with about 250 bp at the 3' end of 1025, continuing further downstream for another 150 bp. 1029 and 1019 are nearly identical. Comparisons of DNA sequences among pine, monocot, and dicot ADHs shows that 1025 and 1019-1029 probably represent two genes whose ancestral form diverged from the lineage leading to angiosperms shortly before monocots and dicots diverged (Kinlaw et al. in preparation). When

these pine cDNA clones were used as probes against Southern blots of genomic DNA from Monterey and loblolly pines, many homologous genomic sequences were identified (Harry et al. in preparation). This result was surprising since ADH belongs to a small gene family in the angiosperms studied to-date.

As part of our continued efforts to characterize ADH in pines, we have screened genomic libraries from Monterey pine and from loblolly pine. Consistent with our analyses of genomic Southern blots, we identified many genomic clones that share DNA sequence homology with our pine ADH cDNAs. We present here preliminary analyses of same of the many clones we have identified.

MATERIALS AND METHODS

Total cellular DNA was isolated from Monterey and loblolly pines using an extraction procedure described by Loopstra et al. (in preparation). For Monterey pine, we used cells from the suspension culture line we had used earlier as a source of our pine cDNA clones (Kinlaw et al. 1988, and Kinlaw et al., in preparation). This cell line has been maintained in culture for several years and was a gift from Dr. Robert Teasdale, Bond University, Australia. For loblolly pine, needles were collected from clone 7-56 in the NC State Cooperative Tree Improvement Program. For both species, the majority of the purified DNA was over 100 kilobases (kb) in length, as judged by gel electrophoresis (unpublished).

Libraries for both species were prepared using standard methods in the bacteriophage cloning vector EMBL3 (see Kaiser and Murray 1985 for a description of the methods). The Monterey pine genomic library was prepared under contract by Clontech, Palo Alto, California, while the loblolly pine was prepared by one of us (CAL). Briefly, DNA was subjected to partial digestion by restriction endonucleases (Sau3A or Mbol), size fractionated by centrifugation, and ligated to BamHI digested "arms" of the bacteriophage cloning vector EMBL3. Recombinant phage were packaged in vitro before being mixed with E. coll host cells (strains K802 or K803) for plating. The Monterey pine library contains approximately 2.7 million independent recombinants while the loblolly pine library is several--fold larger.

The libraries were plated at an average density of approximately 50,000 recombinant phage (or plaque forming units, pfu) per 150mm petri dish. Two to four filter lifts (Hybond--N, Amersham) were made from each of 20-30 plates, but in all cases, duplicate filters were used in each subsequent DNA hybridization reaction. The filters were probed with [321-P labeled DNA from purified inserts of the cDNA clones 1025 or 1029. A typical hybridization reaction consisted of 30 filters in a jar with 350 ml hybridization solution (6X SSPE or SSC, 5X Denhardt's, 200 ug/ml carrier DNA [salmon sperm, Sigma Chemical), and 0.5% w/v SOS, 65 C, Maniatis et al. 1982) with labeled DNA from 300 ng of template. For large hybridization reactions such as these, we have found that placing nylon mesh (e.g. nylon window screens cut into circles) between the filters helps to eliminate background. After hybridization (18-24 hours), filters were washed in at least four changes of 0.5X SSPE (or SSC) with 0.1% SDS, 65 C.

Following autoradiography, plaques showing positive hybridization signals on both replicate filters were identified and subjected to a second round of plating and screening. Secondary screens were done using a lower plating density (1,000 to 5,000 pfu per 150mm plate), and plaques that again showed positive hybridization signals to ADH cDNA probes were subsequently purified. Once a homogeneous phage stock was obtained, phage DNA was prepared from liquid lysates using a PEG "miniprep" procedure modified from Ausubel et al. (1989). Purified phage DNA was digested with various restriction enzymes (BRL, using the manufacturer's recommendations), subjected to electrophoresis (0.7-1.2% agarose in 1X TBE buffer containing 0.5 mg/1 ethidium bromide, Maniatis et al. 1982). Fragments were visualized and photographed on a UV transilluminator (302 nm). Fragment lengths were estimated visually by comparing their migration distances to DNA standards included in the gel. Subsequently, DNA was blotted to to membrane filters (Hybond-N), probed with labeled ADH cDNAs, and the hybridizing DNA fragments were determined by autoradiography. Hybridizations and washes were done using high stringency conditions when pine cDNAs were used as probes, but the the stringency of the washes was reduced (4X SSC, 0.1% SDS, 65 °) when the maize clone pZML793 was used as a probe. Blots were probed first with either 1025 or 1029, then stripped (following the manufacturer's recommendations) and probed a second or third time. When pZML793 was used as a probe, it was always used last.

RESULTS AND DISCUSSION

We identified a large number of clones containing sequences homologous to our cDNA clones (Table 1). **We** considered as positive only those plaques with clear hybridization signals on replicate filters. Using this strategy, our frequency of false positives is very low (2-5% or less). The frequency of positive clones (number of positives per pfu screened) we identified from each library is remarkably consistent, as is the fraction of clones that show hybridization to both probes. Because 1025 extends further towards the 5' end of the gene than does 1029, we speculate that some of the additional DNA sequences identified by 1025 share homology with this region. We are not certain why both 1025 and 1029 hybridize to some genomic clones under the high stringency conditions we used in these experiments. The pine cDNAs share about 85% DNA sequence homology and clearly show different patterns of hybridization on identical Southern blots of genomic DNA. Nevertheless, results from hybridization experiments using plaque lifts (or blots of phage DNA following electrophoresis) and genomic DNA are not comparable.

Purified DNA prepared from homogeneous phage stocks has now been prepared from genomic clones of both Monterey and loblolly pines. As an initial step to characterize these clones, we have digested phage DNA with selected restriction endonucleases. Sall separates insert DNA (genomic DNA from pine) from that of the phage vector. Cloned pine DNA is visualized as one or more fragments, depending on the number of Sall cleavage sites contained in the insert. In addition, EcoRI or BamHI were used in combination with Sall because phage vector contains no recognition sites for either of these enzymes. In this way, vector and insert. DNA are easily

distinguished, and fragment lengths of the insert DNA can be conveniently determined. It is important to distinguish DNA fragments formed by cleavage at the Sall site of the vector adjacent to the insert DNA from fragments formed by cleavage at internal recognition sites. Such fragments have been truncated in the cloning process and will be shorter than homologous sequences in vivo or that are represented as internal fragments in independent genomic clones.

Table 1. Number of positive plaques identified while screening genomic libraries of Monterey pine and loblolly pine using probes from two ADH cDNA clones from Monterey pine.

Species	No. pfu Screened (x 10)	Probe 1025	Probe 1029	Positive to Both Probes
Monterey Pine	1.0	30	Not Done	NA
	1.5	30	9	3
Loblolly Pine	2.5	77	35	12

Results of such analyses are summarized in Figure 1 for 24 clones of loblolly pine identified using 1025 as a probe. We present here data for Sall/EcoRI double digests because pine DNA contains many recognition sequences for EcoRI and we limit our discussion to insert fragments that contain at least 3 recognition sites for these enzymes (i.e. the insert is cleaved into 4 fragments). In these cases, two of the insert fragments are "full-length". No two of these clones are identical (Figure 1), but this is perhaps not surprising because overlapping clones were intentionally generated by partial digestion with Sau3A in the construction of the library.

Clones containing similar sized restriction fragments should be compared to determine whether they contain overlapping sequences of genomic DNA. To facilitate such analyses, clones were sorted by the length of their hybridizing fragment(s) (Figure I). Clones designated B and C are the best example of a pair of clones likely to contain overlapping sequences. Both clones contain a 1.4 kb fragment that hybridizes to 1025, and they contain 0.4 and 1.5 kb fragments that do not hybridize. Two other pairs (J-K and M-N) each contain one other similar-sized fragment in addition to the similar-sized hybridizing fragment. From these limited data, there is no convincing evidence that other clones overlap. These results are now being extended by digesting DNA with Sau3A, which has a 4-base recognition sequence and is expected to cleave insert DNA into many smaller fragments.

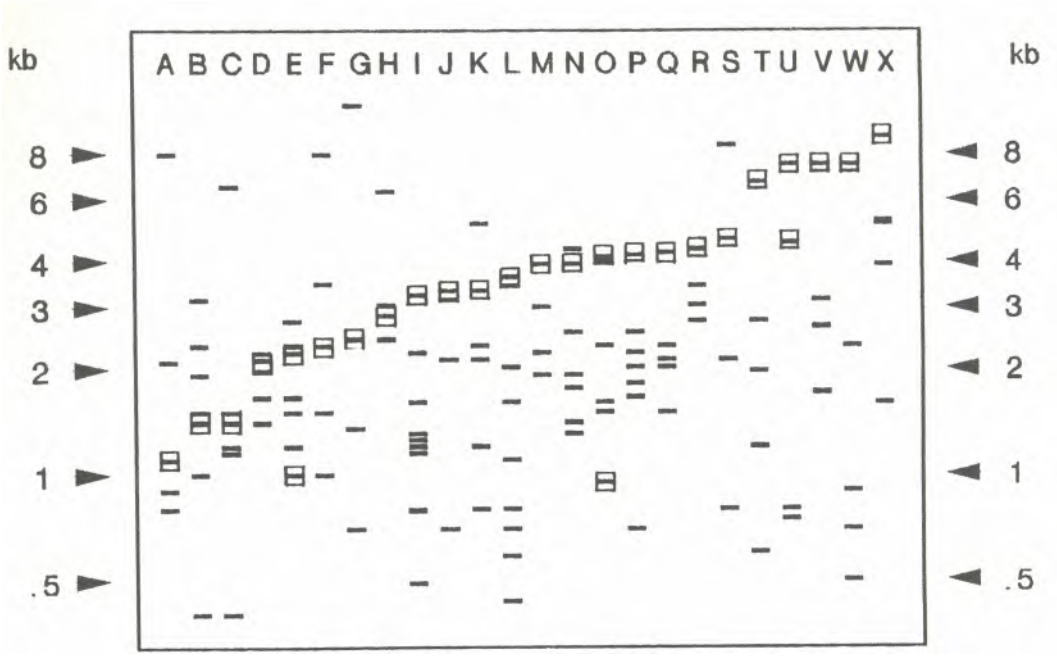

Figure 1. Diagrammatic representation of restriction fragments from genomic clones of loblolly pine, as they wouLd appear on a gel. Shown for each. clone are fragment lengths of insert. DNA digested with both EcoRI and Sall. Fragments that hybridize to ADH cDNA probes are indicated by a box. Clone ID is coded along the top of each "lane". Arrangement of the clones is by size of the hybridizing fragment. Clones that represent overlapping regions of genomic DNA should share DNA fragments of the same size. Clones B and C probably contain such overlapping genomic DNA.

We are particularly interested in identifying clones that contain an entire ADH gene. The most direct strategy to identify such clones is to determine those clones that hybridize to DNA homologous to regions at both the 5' and 3' ends of the gene. While our pine ADH cDNAs contain sequences near the 5' end of the gene, they do not contain sequences at the 3' end.

As an indirect measure to-date, we have probed blots of the genomic clones with purified insert DNA from the full length maize ADH cDNA pZML793. So far, we have not identified insert fragments that hybridize to the maize cDNA that have not already been observed to hybridize to pine **ADH. We are** now repeating these experiments with a subclone of pZML793 that represents the 3' portion of the coding sequence that does not overlap with our pine cDNAs.

SUMMARY AND CONCLUSIONS

Pine genomes contain many DNA sequences homologous to the coding sequence of ADH. Many different genomic clones have now been isolated and they appear to represent different portions of the pine genome. Whether any of these clones contain an intact ADH gene is not yet known. We cannot account for the large number of ADH sequences we observe in pines by extrapolating from the structure and organization of ADH genes and gene family in angiosperms. One possibility is that pine ADH genes are large, containing many introns or containing large introns. Another possibility is that the ADH gene family contains several members in addition to those already characterized by isozyme analyses. Whether ADH in pines turns out to be typical of other pine genes remains to be seen. If so, this will have important ramifications for such diverse tasks as gene identification and isolation, and genome analysis and mapping by RFLPs.

ACKNOWLEDGEMENTS

This research has been supported by a U.S.D.A. Forest Biology Grant (86-FSTY-9-0190) to DEH, CSK, and RRS, and by Cooperative Agreement PSW-88-0008CA with the U.S. Forest Service.

LITERATURE CITED

Ausubel, F.M. et al. (eds) 1989. Current Protocols in Molecular Biology, Greene Publishing Associates, John Wiley, New York.

Bennett, D.C., and M. Freeling. 1988. Flooding and the anaerobic stress response. in D. Newman and K. Wilson (eds) Models in Plant Physiology and Biochemistry, Vol. 3, CRC Press (in press)

Callis, *J.*, M. Fromm, and V. Walbot. 1988. Introns increase gene expression in cultured maize cells. Genes and Devel. 1: 1.183-1200.

Dennis, E.S., W.L. Gerlach, A.J. Pryor, J.L. Bennetzen, A. Inglis, D. Llewellyn. M.M. Sachs, R.J. Ferl, and W.J. Peacock. 1984. Molecular analysis of the alcohol dehydrogenase (Adh1) gene of maize. Nucleic Acids Research 12(9): 3983-4000.

Freeling, M., and D.C. Bennett. 1985. Maize Adh1 Ann. Rev. Gen. 19: 297-323.

Harry, D.E., C.S. Kinlaw, and R.R. Sederoff. 1988. The anaerobic stress response and its use for studying gene expression in conifers. pp. 275-290 in J. Hanover and D. Keathley (eds) Genetic Manipulation of Woody Plants, Plenum Press, New York.

Howard, E.A., J.C. Walker, E.S. Dennis, and *W.J.* Peacock. 1987. Regulated expression of an alcohol dehydrogenase 1 chimeric gene introduced into maize protoplasts. Planta 170: 535-540.

Kaiser, K., and N.E. Murray. 1985. The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. pp. 1-47 in D.M. Glover (ed.) DNA Cloning, A Practical. Approach. IRL Press, Oxford.

Kimmerer, T.W., and M.A. Stringer. 1988. Alcohol dehydrogenase and ethanol in the stems of trees: Evidence for anaerobic metabolism in the vascular cambium. Plant Physiol. 87: 693-697.

Kinlaw, C.S., D.E. Harry, D.D. Sleeter, and R.R. Sederoff 1988. Using heterologous probes to isolate and characterize conifer genes. pp. 9-18 in Cheliak, W.M., and A.C. Yapa (eds). Molecular Genetics of Forest Trees. Proc. Second IUFRO Working Party on Molecular Genetics. Petawawa Nat. For. Instit., Chalk River, Ont.. Canada.

Maniatis, T., E.F. Fritsch, and *J.* Sambrook. 1982. Molecular Cloning, A Laboratory Manual. Cold Spring Harbor.

Walker, J.C., E.A. Howard, E.S. Dennis, and W.J. Peacock. 1987. DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase 1 gene. Proc. Natl. Acad. Sci. USA 84: 6624-6628.