PROGRESS IN NORTHERN RED OAK GENOMICS IN THE SOUTHERN APPALACHIANS

Matthews, S.A.[a] Trout, B.K.[a] Schlarbaum, S.E[b] Bonds, W.D[a]
a Department of Chemistry and Physics, Western Carolina University, Cullowhee, NC 28723
b Department of Forestry, The University of Tennessee, Knoxville, TN 37901

Abstract

While published material regarding molecular biology of the genus *Quercus* is limited, considerable progress has been made recently in understanding the genetic basis of growth and development in model plant systems such as hybrid poplar and *Arabidopsis thaliana* (Chaffey, 2002). Techniques of comparative genomics are currently being used to analyze the northern red oak (NRO, *Quercus rubra*) genome using information derived from these smaller genomes. NRO was selected as a species for study here in the southern Appalachians for several reasons. Economically it is the most valuable hardwood in the region and the tree faces serious challenges to continued natural regeneration. Furthermore, an NRO seedling orchard is available within the area. In this article we will discuss techniques and strategies currently being used to isolate and sequence NRO genetic material; the inherent difficulties in conducting this type of research on forest trees; the results collected to date; and how these results will be used to further the biological knowledge of this majestic tree.

Introduction

The northern red oak is an ecologically and economically desirable hardwood species common in the southern Appalachian Mountains. Over 20,000,000 cubic feet of northern red oak is harvested annually in North Carolina at an annual value of over $30,000,000 US (FIA, 2003). The tree's ecological value is evidenced by large mast production, with good acorn crops yielding over 250,000 acorns per acre. Up to 40% of the mast is consumed by wildlife which become important vectors for the distribution of NRO acorns. In this region, NRO prefers moist, north facing slopes at altitudes up to 1,680 meters. Its wood is widely used for lumber and veneer (Hicks 1998). However, tree growth is limited by long periods of juvenility; acorn production in the third decade; limited competitive natural regeneration; infestation by such pests as red oak borer, canker worm, and gypsy moth; and fungal diseases such as oak wilt and shoestring fungus (Hicks, 1998).

The cellular, molecular, and developmental processes mediating wood formation in *Quercus* is poorly understood at best. There is only limited data concerning the influence of molecular mechanisms on differentiation in the secondary vascular system and the relationships between genetic variability and wood quality. Most molecular research on the family *Fagaceae* and the genus *Quercus* has been for the purpose of establishing phylogenetic relationships and not such practical targets as wood production. As of April, 2003, there were only nine genes from NRO listed in the GenBank nucleotide database (NCBI, 2003) while there are an estimated 20,000 genes present in angiosperms (Bradshaw, 2002). We have worked toward isolating NRO messenger RNAs (mRNAs) for the purpose of rapidly and inexpensively identifying gene coding sequences. The challenge lies not only in being able to isolate

and identify the genes, but in interpreting this information within the context of the cell and the tree as a living organism (i.e., functional genomics). While the exact size of the NRO genome (see note below) is not known, calculations based on 0.9 pg of DNA per cell suggest a nuclear compliment of approximately 800 million base-pairs of genomic DNA (Barreneche et al. 1998). The time and cost involved in the complete DNA sequencing of such a large genome (genomic library), coupled with the fact that protein coding regions probably comprise only 1-3% of this total, greatly influenced our project plan.

The construction of a complimentary DNA (cDNA) library allows the investigator to disregard the large majority of an organism's genome and focus solely on those regions that are transcribed into mRNA and ultimately protein. The construction of such a library begins with the tissue of interest and ends numerous steps later with a complimentary DNA representation of the mRNAs initially present. Segments of the DNA can then be sequenced and comparisons made to sequences in public databases. The construction of cDNA libraries from as many NRO tissues as possible, including shoot meristems, roots, leaves, acorns, pollen, developing bark, embryos and vascular cambium, will increase the complexity of the libraries and the number of unique or novel genes discovered.

Note: An organism's entire complement of nuclear DNA is known as its genome. In every living cell, segments of nuclear DNA are transcribed into lengths of mRNA which in turn are translated into proteins that perform the biochemical activities of the organism. The more frequently a gene is expressed, the greater the levels of mRNA exported from the cells' nucleus to the ribosomes for translation. Gene expression varies with cell type but can also be affected by exposure to external stimuli such as temperature, photoperiod, nutrient and water availability, and disease.

NRO Seed Orchard and Sample Collection

Samples were recovered from the Watauga Northern Red Oak Seedling Seed Orchard near Butler, Tennessee which was created from an open-pollinated progeny test established in 1973 by the Tennessee Valley Authority (TVA). One-year old (1-0) seedlings, grown at the TVA Nursery near Clinton, Tennessee, were used for the plantation. The seedling seed orchard was created in 1987 by thinning the smallest trees in each four-tree plot, leaving one or two trees (Lafarge and Lewis, 1987). Approximately 1150 trees were left in the plantation after thinning. The majority of families were retained (ca. 200), as nothing was known at the time about fruiting and management of NRO seed orchards. Another thinning occurred following the 1994 growing season, leaving approximately 750 trees (Schlarbaum *et al.*, 1998).

Working towards the objective of discovering genes in the wood forming region, tissues from the cambial zone were harvested by removing the outer bark with a hammer and chisel, visually identifying the boundary between bark and sapwood and then scraping the exposed surfaces of the sapwood with a chisel to a depth of 0-2mm. The tissue is collected in foil paper and snap frozen in liquid nitrogen.

Experimental: RNA Isolation

There are numerous methods for isolating RNA from plants. The chemical composition of the individual plant tissue is an important factor in selecting an extraction process. Initial stages of the project were difficult due to the lack of published material concerning isolation of genetic material from forest trees. Five different RNA isolation techniques were evaluated before selecting a method developed by Chang (1993). This method uses a cationic detergent and was selected because of

the ability to process large quantities of starting material and its ability to neutralize interfering compounds.

Samples are processed in the lab by grinding to a fine powder in a mortar and pestle under liquid nitrogen. A cationic detergent (CTAB) is used as the lysis buffer, followed by three chloroform extractions to remove polysaccharides and proteins. Centrifugation at 10,000 x the force of gravity expedites the organic and aqueous phase separation. The aqueous layer is removed leaving the unwanted proteins and polysaccharides in the organic phase. The RNA is precipitated out of the aqueous solution with LiCl and dissolved in an appropriate amount of water (Chang 1993). Once RNA has been isolated, the sample is separated on an agarose/formaldehyde denaturing gel to ensure quality and quantified using an UV spectrophotometer (260 nm) (Sambrook et al. 1989). Quality is assessed by the presence of distinct ribosomal bands in the gel. To establish a high quality cDNA library the RNA isolated from a source should be intact with as many full-length mRNAs as possible. The occurrence of two distinct ribosomal bands with a color intensity of the larger fragment twice as intense as the smaller fragment suggests the presence of intact, undegraded RNA. The ribosomal bands from the 28S and 18S subunits from NRO tissue can be clearly seen in Figure 1. Lane 1 contains a RNA marker used to estimate the size of unknown fragments.

The amount of RNA recovered from an individual sample depends on various things including the quality and tissue type of the starting material. Average RNA yields from leaves- 23 ug/g, root radicals- 40 ug/g, stem shoots- 38 ug/g, dormant cambial zone- 3 ug/g, and early spring cambial tissue- 11 ug/g have been obtained thus far.

Once an adequate amount of high quality total RNA from a specific tissue is obtained mRNA can then be isolated. Affinity chromatography is designed to manipulate a unique structural characteristic of mRNA. Full length mRNA messages have a poly-A tail composed of up to 250 adenosine residues (Brown 2002). Utilizing the affinity of the adenosine residues in the mRNA tail for complementary thymine residues on the chromatography column, the mRNA is isolated from the total RNA. 790ug of total RNA has been demonstrated to yield 10ug of NRO mRNA.

Tree cells also contain large quantities of RNA degrading enzymes known as RNAses. These enzymes are present as part of the innate immune system to protect the tree from viruses that are in the form of RNA. RNases are also released from the vacuoles of cells to degrade genetic messages that are no longer needed. To overcome the problems of RNAse contamination the work area must be kept painfully clean and all solutions are treated to remove RNAse activity. Tissue samples must be stored frozen at -80°C to reduce the activity of RNAses.

Conclusions

Messenger RNA is now being synthesized into complementary DNA utilizing Superscript II® reverse transcriptase (Invitrogen). The double stranded cDNA will then be ligated into a virus that infects a strain of the *E. coli* bacteria. The bacteria will produce millions of clones for each ligated message, creating a library of genetic messages unique to tissues of NRO. Clones will be randomly picked and regions of ligated DNA are excised and subjected to automated sequencing. After numerous clones are sequenced the quality of the sequence will be evaluated and compared to model organisms. Since the cDNA library is the end product of many individual steps its efficiency can be compromised by inefficiency at any step.

The preparation of full-length cDNA libraries is advantageous in that most clones will then contain the complete coding sequence, which accelerates sequencing, biocomputation, and protein expression (Carninci *et al.* 2000). The normalization of these libraries further enhances the quality and complexity. The mRNA content of a cell varies depending on expression. Some mRNAs are abundant in a cell with 5-10 species comprising at least 20% of the mass, while 500-2000 intermediately expressed species comprise 40%-60%. There can be as many 10,000-20,000 rare messages in a cell that account for less than 20%-40% of the mRNA mass (Carninci et al. 2000). Reddy *et al.* separately constructed standard and normalized cDNA libraries of rice for comparison. A redundancy of about 10% was reported for the first 200 clones of the non-normalized library, compared with 3.5% in the normalized library. The discovery of novel genes also increased with the normalized library compared to the non-normalized, being 28.2% and 5% respectively (Reddy *et al.* 2002).

Along with sequencing randomly chosen cDNAs, microarray technology offers the potential for the rapid identification of a large number of genes (DeRisi, Vishwanath, and Brown, 1997). Utilizing a microarray chip, fluorescently tagged cDNA's can be hybridized against single stranded DNA segments from thousands of known genes attached to a chip. Chips are presently available for *A. thaliana*, an angiosperm whose entire genome has been sequenced, and *Populus* chips have been spotted with more than 13,000 EST's, but are not as accessible as those for *A. thaliana* (Wullschleger *et al.* 2002). Those genes that have been conserved in NRO would hybridize with those on the microarray chip, and a scanner can then identify the degree of hybridization at each individual spot, allowing those genes that are present in the cDNA to be ascertained.

Knowledge of the transcriptional messages controlling reproductive development, wood formation and defense mechanisms will be essential for the effective integration of biotechnology into tree improvement work (Pena & Sequin 2001). The public database that will emerge from this project will facilitate the discovery of novel genes in northern red oak. By applying the described molecular techniques we hope to further the biological understanding of the northern red oak tree.

Work Cited

Barreneche, T., Bodenes, C., Lexer, C., Trontin, J., Fluch, S., Streiff, R., Plomion, C., Kremer, A. (1998) A genetic linkage map of Quercus robur L. based on RAPD, SACR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theoretical Applied Genetics* 97:1090.

Bradshaw, T. (2002). Geneticists branch out to sequence poplar genome. *Nature* 415:725.

Brown, T.A. (2002). Genomes. New York: John Wiley & Sons, 286.

Carninci, P., Shibato, Y., Norihito, H., Sugahara, Y., Shibato, K., Masayoshi, I., Hideaki, K., Okazaki, Y., Masami, M., and Hayashizaki, Y. 2000. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Research. **10**: 1617-1630.

Chaffey, N. 2002. Why is there so little research into the cell biology of the secondary vascular system of trees? *New Phytologist* 153:213-223.

Chang, S., Puryear, J., and Cairney, J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter.* 11: 113-116.

DeRisi, J.L., Vishwanath R.I., Brown P.O. 1997. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science* 278:680-685.

Hicks, R. Jr. (1998). Ecology and Management of Central Hardwood Forests. New York: John Wiley & Sons, 174-176.

LaFarge, T. and Lewis R.A. 1987. Phenotypic selection effective in a northern red oak seedlings seed orchard. In *Proceedings of the Southern Forest Tree Improvement Conference*, 16-19 June 1987, College Station, Texas. 200-207.

Merkle, S.A., Dean J.D. (2000). Forest Tree Biotechnology, *Current Opinion in Biotechnology* 11:298-302.

National Center for Biotechnology. Online: *http://www.ncbi.nlm.nih.gov/,2003*

Pena, L., Seguin A., (2001). Recent Advances in the Genetic Transformation of Trees. *Trends in Biotechnology* 19:500-506.

Plomion C., Leprovost G., Stokes A. (2001) Wood Formation in Trees. *Plant Physiology* 127:1513

Reddy, A., Ramakrishna, W., Sekhar, C., Nagabhushana, I., Babu, P., Bonaldo, M.F., Soares, M.B., and Bennetzen, J. 2002. Novel genes are enriched in normalized cDNA libraries from drought-stressed seedlings of rice. *Genome*. 45: 204-211.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Press, Cold Spring Harbor, New York.

Schlarbaum, S.E., Barber, L.R., Cox, R.A., Cecich, R.A., Grant J.F., Kormanik Paul P., LaFarge, T., Lambdin P.L., Lay S.A., Post, L.S., Proffitt, C.K., Remaley, M.A., Stringer, J.W., and Tibbs, T. 1998. Research and development activities in a northern red oak seedling seed orchard. In *Deversity and adaptation in oak species*. Proc. 2nd IUFRO Genetics of *Quercus* meeting, K. Steiner, ed. 185-192.

United States Forest Service. Forest Inventory and Analysis Timber Product Output Database Retrieval System 1997. Online: *http://srsfia.usfs.msstate.edu/rpa/tpo/*

Wullschlegar, S., Jansson, S., and Taylor, G. 2002. Genomics and Forest Biology. *The Plant Cell* 14: 2651-2655.

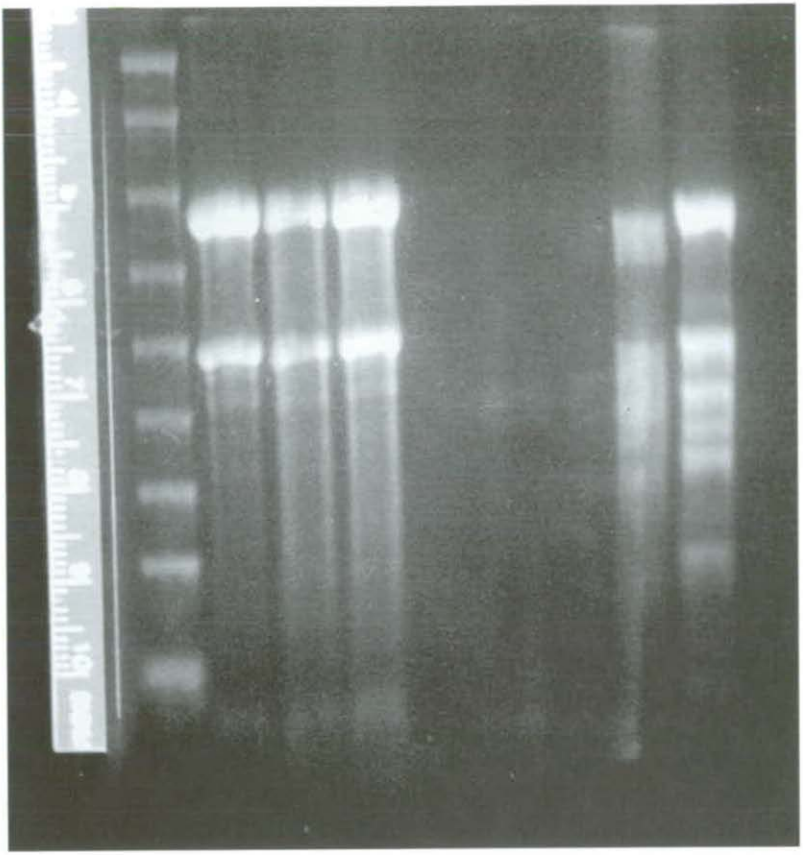

Figure 1. Formaldehyde/denaturing gel of RNA samples from stem shoots (lane 2) and root radicals (lane 3 & 4). Lane 1 contains an RNA marker used for size identification. (Sigma-Aldrich Company, St. Louis, Mo.)