DEVELOPING CODOMINANT PCR MARKERS IN PINES

D. E. Harry and D. B. Neale[1]

Abstract.--Genetic markers which are easy to use and reveal abundant allelic variability provide powerful tools for many forest genetics applications. Because of their simplicity, methods using the polymerase chain reaction (PCR) are being developed for many plants. One class of PCR-based marker takes advantage of random amplification of polymorphic DNA (RAPD), but RAPD markers suffer several disadvantages related to their dominance and to their lack of comparability across pedigrees and between laboratories. To circumvent such problems, we are developing codominant PCR-based markers for pines. Our initial efforts have focussed on a subset of the cDNA clones previously used for constructing genetic maps. Nucleotide sequences from cDNA clones are used to design PCR primers for amplifying targeted fragments of genomic DNA from loblolly pine *(Pinus taeda L.)*. Our pilot study demonstrates the feasibility of generating PCR markers in this way, but this method is complicated by the abundance of gene families in pine genomes.

Keywords: *Pinus taeda,* polymerase chain reaction, genetic map, genetic marker, small subunit

INTRODUCTION

Genetic mapping has been widely used in basic and applied studies of plants and animals. Among their many applications, genetic maps provide descriptive information on genome organization and they can assist in isolating and cloning genes of interest (Neale and Williams 1991). Genetic maps can also help to interpret data collected in population and evolutionary studies. Moreover, by tracking specific chromosomal segments from generation to generation, genetic maps and their associated markers are becoming particularly important in plant and animal breeding (Tanksley et al. 1989; Neale and Williams 1991). We are engaged in a long-term project to develop genetic maps and associated map-based technologies for pines (Devey et al. 1991; Neale et al. 1992). We have focused on loblolly pine because it is the mostly widely planted commercially important forest species in the United States. However, we are keenly aware of the demand for map-based applications in other pines and conifers. To-date, we have largely worked to develop genetic markers based on restriction fragment length polymorphisms (RFLPs) because they are reliable and informative, and because the information they provide is comparable across pedigrees and species.

---

[1] Molecular Geneticists, Institute of Forest Genetics, USDA Forest Service, Pacific Southwest Research Station, Albany, CA 94710

Many factors can influence selecting genetic markers for a given purpose. Isozyme markers cannot be surpassed for their cost, simplicity and ease of use, but they are limited to relatively small numbers of informative (i.e. polymorphic) loci. RFLPs are essentially unlimited in number, but intricacies associated with their use limits them to technically sophisticated laboratories. Recently, PCR based markers such as RAPDs (Williams et al. 1990, Welsh and McClelland 1990) have received much attention. Unfortunately, RAPD markers tend to be difficult to reproduce between laboratories and between pedigrees. Ignoring problems of reproducibility, RAPDs are typically dominant, so much of the high levels of heterozygosity found in forest tree species (e.g. Hamrick et al. 1981) cannot be detected unless haploid tissues (in conifers) are sampled or other special circumstances apply (e.g. Grattapaglia et al. 1992). Despite their obvious utility, RAPDs have limited applicability for forest trees.

Forest trees are different from nearly all other crop plants in that even the most commercially important species are relatively undomesticated. As was pointed out some time ago (Libby et al. 1969) this feature brings both opportunities and responsibilities. While we are very much interested in breeding applications stemming from our mapping efforts (e.g. Williams and Neale 1992), there is a need for technology to support conservation and stewardship activities on both public and private forests (e.g. Millar et al. 1990 and associated papers). DNA markers clearly have a role to play in this arena (Schaal et al. 1991, Wagner 1992). Ideally, such markers would be codominant, multiallelic, simple to use and assay, reproducible, reliable, and to the extent possible, useful in related species. Hence, given the breadth of the questions to address, and yet recognizing limited resouces within the forestry community, we are striving to develop marker technologies that are simple to use and yet as broadly applicable as possible.

Most of our mapping efforts to-date have used RFLPs detected using anonymous cDNA clones. On average, 10-20% of these cDNAs about reveal mappable polymorphisms. A distinct advantage of using cDNA clones is their utility for identifying RFLPs in related conifer species (Neale and Williams 1991, Colby et al. 1993, and Ahuja et al., in press). This suggests that PCR primers designed on the basis of a cDNA from loblolly pine may also be able to amplify genomic DNA from related pines and perhaps other conifers.

An unusual feature of pine genomes, at least as compared to other higher plants, is the frequency and extent of multigene families (Gerttula and Kinlaw, 1993, and Kinlaw, unpublished). We postulated that such apparent complexity might be caused by genes with large or many introns (Kinlaw et al. 1990), but since a few conifer genes have now been characterized (Hutchison et al. 1990, Kojima et al. 1992, Spano et al. 1992, and Harry et al. in preparation) this now appears unlikely. Such features of conifer genomes are important to consider because the ability to amplify genomic DNA and to detect polymorphims depends on the size and nature of the amplified PCR fragments.

We are now studying the feasibility of using PCR primers designed from cDNA clones to detect allelic variability in loblolly pine. Here we report the identification of a PCR-based polymorphism that maps to the same location as a RFLP marker detected the using same cDNA probe from which the PCR primers were designed.

METHODS

We selected a subset of 27 cDNA clones that had previously been used as probes to generate either of two genetic maps using RFLPs (Devey et al. 1991, Groover et al. 1993). Our primary selection criterion was that the clones revealed well-resolved RFLP banding patterns on Southern blots of loblolly pine genomic DNA. Typically, we selected clones

revealing few RFLP bands, but we also selected a few clones with more complex banding patterns (e.g. Gerttula and Kinlaw, 1993).

After selecting the clones, we determined the nucleotide sequence for 150-250 bases into each end of the cloned inserts (i.e. one sequencing reaction and gel, using standard dideoxy sequencing methods). Nucleotide sequences were then compared against published sequences contained in the GenBank electronic database to determine whether they appeared similar to any previously characterized sequences. If a tentative identification was made, then additional information was sought concerning the nature of the identified sequence and any corresponding genes that had been characterized. If the latter search found any plant genes with known gene structures (e.g. location and size of exons and introns), then this information was used to help determine suitable locations for PCR primers. Finally, we used the computer program PRIMER (version 0.5, M. Daly, S. Lincoln, and E. Lander, Whitehead Institute, unpublished) to assist in designing PCR primers. Generally, we choose primers to be 18-21 nt long, with a Tm near 60° C, and GC content of 40-60%.

Reaction conditions for the PCR amplifications were standardized after testing a range of concentrations for such components as template DNA and Mg++ (Perkin-Elmer protocols). Genomic DNA samples were prepared for the PCR using standard CTAB extraction procedures (Devey et al. 1991), supplemented by an additional organic extraction and ethanol precipitation. Final DNA concentrations were estimated with a fluorometer (Hoefer) using calf thymus DNA as a standard. The following conditions are described for $50 \mu_l$ reactions, but 25-100 $_A$l reactions were used with equal success after adjusting all components proportionately. Reactions were set up containing 30 ng of genomic DNA and 1.5 mM Mg++ in 40 ⌐ volumes (other components followed standard Perkin-Elmer recommendations) without Taq polymerase, overlain with mineral oil, and subjected to an initial denaturation step of 100°C for 5 min, then cooled to 85°C. At this point, 1.25 U Taq polymerase was added in 10 μl of lx PCR buffer supplemented with gelatin and DMSO for a 1% (v/v) final concentration. Immediately afterwards, temperature cycling began using the following conditions: 4 cycles of 1 min, 96° denaturation; 1 min, 55° annealling; 2 min 72° extension; followed by 30 cycles in which the denturation step was lowered to 94°. A 5 min extension at 72° was added as the final step, and then the samples were cooled to 4°. PCR fragments were subjected to agarose gel electrophoresis in 1-2% agarose (BRL) and visualized by staining with ethidium bromide.

RESULTS AND DISCUSSION

Similarity Searches

Based on our sample of 27 cDNA clones, and including preliminary results from a larger sample (Colby et al. 1993, and Kinlaw, unpublished) we find that about 25% of the cloned cDNAs contain sequences that are similar (60% similarity or better) to previously characterized sequences. We also confirmed that two pairs of clones that were presumed to be similar to one another based on similar RFLP patterns on Southern blots indeed are similar at the level of nucleotide sequence.

Among the identified sequences we found two photosynthetic genes--those encoding the chlorophyll a/b binding proteins and others encoding the small subunit of ribulose bisphosphate carboxylase (rubisco). The latter are represented by the cDNA clones pPt1IFG2025 and pPtIFG1635 (henceforth abbreviated as 2025 and 1635). We now focus our attention on PCR products derived using primers derived from these clones.

Small Subunit Gene Family

Nucleotide sequences for the small subunit of rubisco (rbcS) have been characterized from a variety of plants. RbcS genes typically belong to a small gene family (summarized by Dean et al. 1989) showing a conserved gene structure. Individual genes contain a sequence encoding a transit peptide, and most angiosperm genes contain two introns with identical locations (Dean et al. 1989). Small subunit clones have also been isolated from conifers, including a cDNA from *Pinus thunbergiana* (Yamamoto et al. 1988) and a genomic sequence from *Larix laricina* (Hutchison et al. 1990). The structure of the larch gene appears similar to angiosperm genes in that it encompasses a sequence for a transit peptide and also has two introns (Figure 1).

The nucleotide sequences from 2025 and 1635 were easily aligned with those from larch and with each other (Figure 1). The 5' ends of both cDNAs begin near the middle of the first exon in larch. The nucleotide sequences from the two sequencing reactions done for 2025 overlapped and so could be joined, but those for 1635 could not because it contains a longer insert. In fact, no 3' coding sequence was detected in 1635 using only the one sequencing reaction.

Four PCR primers were designed for the rbcS cDNAs using the following rationale. A single "forward" primer was chosen near the 5' end of the cDNAs in a region conserved in 1635, 2025, and also in larch (Figure 1). This is, therefore, a "generic" rbcS forward primer that may prove useful for amplifying DNA from a number of conifer species. In a similar vein, a reverse primer was selected near the 3' end of the coding sequence in 2025, also in a region conserved in pines and larch. In addition, two reverse primers were designed to correspond to non-coding regions unique to either 1635 or 2025. We reasoned that such 3' specific primers might preferentially amplify a subset of the rbcS gene family members.

Figure 1. Alignment of loblolly pine cDNAs pPt1IFG2025 and pPt1IFG1635 with a gene from larch. Two PCR primers (indicated by forward and reverse arrows) were designed to align with coding sequences, and two additional primers were designed to be specific to the non-coding regions at the 3' ends of the cDNA clones. Exons (representing coding regions) are wide boxes with cross-hatches. Introns are indicated by narrow boxes with stipples. Exons and introns (from larch) are drawn to scale.
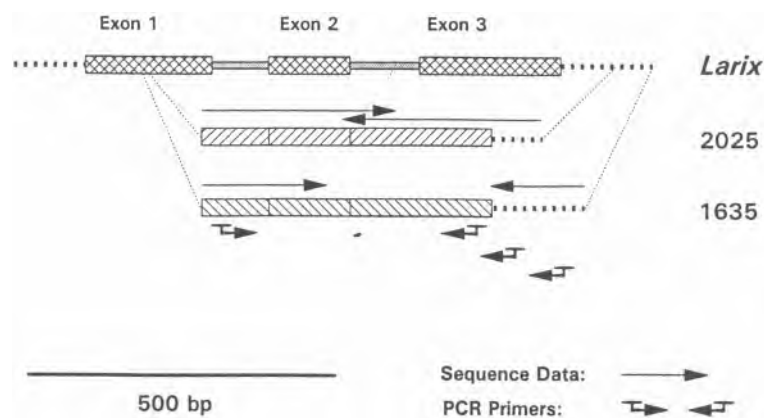
PCR Amplification using RbcS Primers

As expected, PCR amplifications done using the forward and reverse coding sequence primers amplified both 1635 and 2025 cDNAs, as well as genomic DNA from loblolly pine. Only one fragment was seen from the genomic sample. The size of the PCR fragment from the genomic sample was about 220 bp longer than from the cDNAs, suggesting the presence of one or more introns with a total length of 220 bp. This is consistent with the larch gene, which contains two introns with a combined length of 203 bp. Furthermore, different members of the rbcS family, if they are indeed represented in this pool of PCR fragments, have very similar gene structures.

Both 3' specific primers also amplified genomic DNA, and the size of single PCR fragment from these reactions was consistent with expectations predicted in Figure 1: the 2025-specific 3' primer produced a fragment of about 780 bp, whereas the 1635-specific 3' primer produced a larger fragment, about 870 bp. To determine whether the 3' specific primers might selectively amplify different rbcS genes, both were used for PCR reactions using both 2025 and 1635. The 2025-specific primer only produced a PCR fragment using 2025 template, and likewise for the 1635-specific primer. Even so, it appears that both 3' specific primers amplify more than one gene since restriction digests of the amplified products reveal DNA fragments whose combined size is larger than the undigested products (problems associated with incomplete digests can be excluded).

Allelic Segregation

To identify allelic polymorphisms, PCR fragments were generated using all three pairs of rbcS primers to amplify genomic templates from individuals in a 3-generation pedigree (Devey et al. 1991). No length polymorphisms were observed among samples of undigested DNA. To reveal variation in nucleotide sequence, the PCR fragments were subjected to digestion using an array of restriction enzymes with 4-base recognition sequences. To-date we have found one polymorphism that segregates in a Mendelian fashion.

PCR fragments generated using the 3' primer specific to 2025 were digested with *BstNI.* Most of the PCR fragments produced in this way lacked a *BstNI* site, but for one of the four grandparents of the pedigree, this digestion produced two novel fragments, of about 360 and 420 bp (i.e. summing to the 780 bp of the undigested product). The F1 parent produced by this grandparent also contained this fragment, as did 7 of the 21 tested F, progeny. In other words, variation for the presence of a *BstNI* site appears to behave as a Mendelian marker.

Because of the allelic configuration in this cross, we have not yet observed an individual that is homozygous for the allele carrying the *BstNI* site. However, because of the nature of the rbcS gene family, we believe that such an individual would have a phenotype identical to a heterozygote. That is, the 780 bp fragment would still be present because it also represents PCR fragments that were generated from other genes, and these products also lack the restriction site. We plan to test this prediction by observing the PCR phenotypes of haploid gametophytes from seeds of the heterozygous parent.

Co-segregation with a 2025 RFLP Marker

To determine whether this PCR-based marker appears linked to the mapped 2025 RFLP marker, we cross-classified individuals whose phenotypes had been scored for both. For the RFLP marker, the parents had been classified as Al A2 x A1A3, and we observed

the four genotypic classes segregating approximately 1:1:1:1 (Figure 2). Only two PCR phenotypes could be distinguished, but the "b" allele (carrying the *BstNI* site) was completely coincident with the allele A3 (Figure 2). These data suggest that the RFLP allele A3 is linked with the PCR-marker, and indeed probably represents a DNA fragment with an internal *BstNI* site that the A 1 and A2 RFLP alleles lack. To our knowledge, this is the first example whereby a mapped RFLP polymorphism has been converted to a mapped PCR-based polymorphism in conifers.

Figure 2. Cross-classification of RFLP and PCR genotypes among individuals segregating for an rbcS polymorphism. RFLP genotypes were scored from Southern blots probed using the loblolly pine cDNA clone pPt1IFG2025. PCR genotypes were scored using PCR primers designed from the same cDNA clone.

RFLP Genotypes

		A1 A1	**A1 A2**	A1 **A3**	**A2 A3**	**Total**
PCR Genotypes	**aa**	1 0		0	0	1 4
	ab	0	0	3	4	7
	Total	1 0	4	3	4	21

Parental Genotypes:
RFLP: Al A2 x Al A3   PCR: aa x ab

We are encouraged by these results in many respects. We have learned a great deal about the nature of the rbcS gene family in pine with a few relatively simple experiments. For example, the size of the fragments produced by digesting the *BstNI* suggests that the polymorphic site is contained in the second intron. Base substitutions are expected to be more frequent in introns because of the constraints imposed by natural selection in coding sequences. Moreover, by selecting PCR primers in non-coding regions flanking expressed genes we can reduce the molecular complexity in the pool of amplified products, and thereby simplify genetic interpretations for any polymorphisms that are observed.

However, we have seen that it may be a difficult tast to detect length polymorphisms among PCR fragments generated using primers from expressed genes, at least unless specific steps are taken to do so. Nevertheless, we are also encouraged by the results of Morgante and Olivieri (1993), who located a number of simple sequence repeats in the introns of plant genes. Allelic variants of simple sequence repeats contain different numbers of repeat units, and therefore represent length variants. Experimental strategies designed to identify such repetitive sequences in the introns of loblolly pine genes may prove to be productive. We are about to begin such a series of experiments.

CONCLUSIONS

This preliminary survey leads us to the following tentative conclusions:

1. PCR primers selected from sequencing mapped cDNAs can reveal useful and mappable genetic polymorphisms.

2. Polymorphisms that reveal themselves as length variants among PCR fragments are not common. Instead, other methods are needed to detect variation nucleotide sequence among the fragments. Such methods could be as simple as restriction digests, but more powerful techniques may also be required.

3. Gene families in pines will complicate gene interpretations if multiple members of the family are amplified simultaneously.

ACKNOWLEDGEMENTS

LITERATURE CITED

Ahuja, M.R., M.E. Devey, K.D. Jermstad, and D.B. Neale. 1993. DNA probes from loblolly pine detect restriction fragment length polymorphisms in other conifers. Can. J. For. Res. (in press).

Colby, S.M., A. Groover, C.S. Kinlaw, D.E. Harry, and D.B. Neale 1993. Advancing towards a transcriptional map in loblolly pine. Proceedings, Southern Forest Tree Improvement Conference, Atlanta (this meeting).

Dean, C., E. Pichersky, and P. Dunsmuir. 1989. Structure, evolution, and regulation of *RbcS* genes in higher plants. Ann. Rev. Plant Physiol. Plant Mol. Biol. 40: 415-439.

Devey, M. E., K. D. Jermstad, C. G. 'Dauer, **D. B.** Neale. 1991. Inheritance of RFLP Loci in a Loblolly-Pine 3-Generation Pedigree. Theor Appl Genet 83: 238-242.

Gerttula, S.M., and C.S. Kinlaw, 1993. Complex gene families in pines. Proceedings, Southern Forest Tree Improvement Conference, Atlanta (this meeting).

Groover, A. and others, 1993. Mapping quantitative trait loci affecting wood specific gravity in loblolly pine. Proceedings, Southern Forest Tree Improvement Conference, Atlanta (this meeting).

Grattapaglia, D., J. Chaparro, P. Wilcox, S. McCord, D. Werner, H. Amerson, S. McKeand, F. Bridgwater, R. Whetten, D. O'Malley, and R. Sederoff. 1992. Mapping in woody plants with RAPD markers: Application to breeding in forestry and horticulture. Applications of RAPD Technology to Plant Breeding, Joint Plant Breeding Symposia Series, Minneapolis, MN, pp. 37-40.

Hamrick, J. L., J. B. Mitton, and Y. B. Linhart. 1981. Levels of genetic variation in trees: Influence of life history characteristics. pp. 35-41. in M. T. Conkle (ed.), Proceedings of the Symposium on Isozymes of North American Forest Trees and Forest Insects, July 27, 1979, USDA Forest Service Pacific Southwest Forest and Range Experiment Station Gen. Tech. Rep. PSW-48.

Hutchison, K.W., P.D. Harvie, P.B. Singer, A.F. Brunner, and M.S. Greenwood. 1990. Nucleotide sequence of the small subunit of the ribulose-1,5-bisphosphate carboxylase from the conifer Larix laricina. Plant Mol. Biol. 14(2): 281-284.

Kinlaw, C.S., D.E. Harry, and R.R. Sederoff. 1990. Isolation and characterization of alcohol dehydrogenase cDNAs from Pinus radiata. Can. J. For. Res. 20: 1343-1350.

Kojima, K.; Yamamoto, N.; Sasaki, S. 1992. Structure of the Pine *(Pinus thunbergii)* Chlorophyll a/b Binding Protein Gene Expressed in the Absence of Light. Plant Mol Biol 19: 405-410.

Libby, W. J., R. F. Stettler, and F. W. Seitz. 1969. Forest genetics and forest tree breeding. Ann. Rev. Gen. 3: 469-494.

Millar, C.I., F.T. Ledig, and L.A. Riggs. 1990. Conservation of diversity in forest ecosystems. Forest Ecol. Managmnt. 35: 1-4.

Morgante, M., and A.M. Olivieri. 1992. PCR-Amplified microsatellites as markers in plant genetics. The Plant Journal 3(1): 175-182.

Neale, D.B. and C.G. Williams. 1991. Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. Can. J. For. Res. 21: 545-554.

Neale, D.B., M.E. Devey, K.D. Jermstad, M.R. Ahuja, M.C. Alosi, and K.A. Marshall. 1992. Use of DNA markers in forest tree improvement research. New Forests 6: 391-407.

Schaal, B.A., S.L. O'Kane, abd S.H Rogstad. 1991. DNA Variation in Plant Populations. Trends Ecol Evolution 6: 329-333.

Spano, A.J., Z.H. He, and M.P. Timko, 1992. NADPH - Protochlorophyllide oxidoreductases in white pine *(Pinus strobus)* and loblolly pine (R taeda): Evidence for light and developmental regulation of expression and conservation in gene organization and protein structure between angiosperms and gymnosperms. Mol. Gen. Genet. 236: 86-95.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: new tools for an old science. Bio/Technology 7: 257-264.

Wagner, D.B. 1992. Nuclear, chloroplast, and mitochondrial DNA polymorphisms as biochemical markers in population genetic analyses of forest trees. New Forests 6: 373-390.

Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitary primers. Nucleic Acids Res. 18: 7213-7218.

Williams, C.G., and D.B. Neale 1992. Conifer wood quality and marker-aided selection: a case study. Can. J. For. **Res.** 22: 1009-1017.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18: 6531-6535.

Yamamoto, N., Y. Kano-Murakami, M. Matsuoka, Y. Ohashi, and Y. Tanaka. 1988. Nucleotide sequence of a full length cDNA clone of ribulose bisphosphate carboxylase small subunit gene from gree dark-grown pine *(Pinus thunberii)* seedling. Nucleic Acids Res. 16(4): 11830.