MICROSATELLITE MARKERS FOR LOBLOLLY PINE

Christine G. Elsik and Claire G. Williams'

We are developing PCR-based codominant microsatellite markers for loblolly pine. Until now there has been a shortage of fully informative PCR-based codominant markers for loblolly pine. In outbred pedigrees, heterogeneity of mating types at each marker increases the number of segregating progeny required for mapping. Multi-allelic markers are more likely to be fully informative, and reduce the number of progeny required in outbred pedigrees. Micro satellite, or simple sequence repeat (SSR), markers produce the highest proportion of fully informative genetic markers, which are especially advantageous to QTL detection and population studies in outcrossing species. The multi-allelic nature of SSR markers can reduce bias in detecting quantitative trait loci for loblolly pine by increasing the proportion of heterozygous parents. Furthermore, since SSR markers require little DNA and no radioactivity, they can be automated for high-throughput data analysis.

Microsatellite sequences were recovered from our first loblolly pine genomic plasmid library enriched for (CCT)n, (GGT)n, (CGT)n and (GCT)n. Library screening revealed that 32% of the clones were positive for trinucleotide repeats. Of the positive clones, 236 have been sequenced and all have been found to contain trinucleotide repeats, up to 70 repeat units in length. Of the clones sequenced, 62 are unique, indicating that our enrichment technique resulted in 74% duplication of sequences. Of the 74% duplication, 65% resulted from two sequences, most likely a result of PCR bias. This problem is now being alleviated by decreasing the number of PCR amplification steps and number of cycles per PCR step in the enrichment procedure. Of the unique clones with sequences flanking the microsatellites sufficient for primer design, 16% contained large repeated motifs in regions proximal to the microsatellites, rendering it impossible to design unique primers. This may be a unique characteristic of the pine genome.

BLAST (Basic Local Alignment Search Tool) results suggest that a high proportion of these microsatellites are proximal to coding regions. The BLAST algorithm measures local sequence similarity based on a matrix of similarity scores for all possible pairs (Karlin and Altschul 1990; Altschul et al. 1990), and detects aligned pairs of sequence segments which are "high scoring segment pairs" (HSPs). Probability of a random match is estimated using the Poisson probability, with a probability, P(N), of 0.05 as the cutoff value for significant HSPs. Our 62 unique sequences were compared to sequences in nucleotide and amino acid sequence databases using BLAST. Microsatellites and other repetitive regions in the query sequences were eliminated by low-complexity filters. BLASTN, which searched for significant HSPs between the query nucleotide sequence and nucleotide sequences in the Genbank, EMBL, DDBJ, and PDB nucleotide databases, indicated similarity to eukaryotic sequences for 21 of the microsatellite clones. BLASTX was used to find significant HSPs between translated nucleotide query sequences in all six reading frames and target amino acid sequences in Genbank, CDS, PDB,

---

'Graduate Student and Associate Professor, Department of Forest Science, Texas A&M University, College Station, Texas.

Swisprot and MR protein databases, revealing significant HSPs between 13 of the microsatellite clones and eukaryotic proteins. Four of the clones matched sequences using both BLASTN and BLASTX. Therefore, 48% of the clones matched nucleotide and/or amino acid sequences of eukaryotes in the databases. Among the matches in BLASTN is a sequence that is similar to a *P. taeda* promoter for an arabinogalactan-like protein gene with P(N) = 4.3 x 10' ^{7}. BLASTX revealed two sequences that are similar to retrotransposons with P(N) = 5.7 x 10^{-28} and 8.8 x 10' ^{23}, and sequences similar to *Nicotiana tabacum* extensin and *Zea mays* CRINKLY 4 precursor, with P(N) = 0.00016 and 0.011, respectively.

Microsatellite markers that are in close proximity to coding regions are especially valuable. Our working hypothesis is that many SSRs are directly involved in gene expression and function. For example, variation in repeat number may cause quantitative changes in gene expression (Kashi et al. 1997). Potential applications of this type of marker are chromosome landing, QTL mapping, and molecular systematics.

ACKNOWLEDGMENTS:

We would like to thank the Texas Coordinating Board of Higher Education, TAES, Georgia Pacific, and the USDA Forest Service for support in this study. CGE would like to acknowledge her dissertation committee members: Claire G. Williams, Claire S. Kinlaw, J. Spencer Johnston and David M. Stelly.

LITERATURE CITED

Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman. 1990. Basic alignment search tool. J. Mol. Biol. 215: 403-410.

Karlin, S. and S.F. Altschul. 1990. Methods of assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. 87: 2264-2268.

Kashi, Y., D. King and M. Soller. 1997. Simple sequence repeats as a source of quantitative genetic variation. Trends Genet. 13:74-78.