**Quarterly/Final Progress Report**

**For:** QTL and Candidate Genes for Growth Traits in *Pinus taeda* L.

**Covering Entire Granting Period:** April 1, 2000 to October 1, 2002

**Recipient:** Claire G. Williams
**Award Number:** DE-FC07-00ID13877

**Consortium:** International Paper, The Timber Company, Weyerhaeuser, Union Camp Corporation, PE-Celera,

**Contact:** Dr. Claire G. Williams, Texas A&M University, 305 Horticulture Building, College Station Texas 77843-2135. Phone: (979) 862-3745, fax (979) 845-6049, email: claire-williams@tamu.edu

**Project Objective:** Develop low-copy microsatellites for *Pinus taeda* L. and conduct a molecular dissection of growth traits.

**Summary:**

Phase 1: Marker development.

Our enriched-copy libraries for developing microsatellites, made in 1999 prior to the DOE award, were based on isolating the low-copy or undermethylated portion of the *Pinus taeda* genome. These library methods made microsatellite development far more cost-effective. The protocols have been published.

Developed 245 nuclear microsatellite markers and released them to the public domain through 1) a Web site (http://forestry.tamu.edu/genetics), 2) an international workshop sponsored by USDA, NSF and Forest Service grants and 3) a training manual (Auckland et al. 2002 Conifer Microsatellite Handbook, Corporate Press, Raleigh NC 57 p). This effort was also aided by a short-term industrial marker consortium, a USDA-Forest Service Agenda 2020 grant and grant funds from the Texas Higher Education Coordinating Board. This phase of the project was published.

The importance of this microsatellite marker achievement is that other conifer groups worldwide have not been able to get any large number of microsatellite markers. There are three reasons for the lack of microsatellite markers:

1) the public sequences databases, another source of sequences bearing repeat motifs, have many EST sequences from coding regions but this type of pine sequence rarely has any polymorphic microsatellite repeat motifs,

2) the public sequence databases so not have the right knid of sequences for finding polymorphic repeat motifs. There are less than 100 genomic sequences from known, fully characterized

conifer genes and yet noncoding regions are where polymophisms (including microsatellites) near genes occur,

3) conifer microsatellite development from anonymous genomic sequences has been stymied by the preponderance of highly repetitive DNA (76-86% of the genome).

An unexpected bonus is that many of these microsatellites developed from our low-copy and undermethylated are polymorphic in other hard pine species in Europe, Middle East and North America including Mexico. These results have been published or will be presented at an upcoming national conference.

Phase 2: Computer programs developed for QTL analysis in large, outbred pedigrees
Software programs were written to facilitate marker data analysis: 1) an error-checking program for single marker loci which detects for non-Mendelian inheritance and 2) a program which traces founder-origin probability for chromosome segments or haplotypes in outbred pedigrees from offspring to a specific grandparent or founder and 3) a QTL detection program for outbred pedigrees which we have named SATORI. A visiting professor, Dr. Humberto Reyes, has written the code for SATORI and it is based on his original contributions to QTL mapping theory for outbred pedigrees. The founder-origin probabilities manuscript is now in press.

Phase 3: Linkage map construction
Using these markers, over 24,000 datapoints were generated for linkage map construction. Of the 245 markers, 95 were polymorphic for the public RFLP map and of these, 44 were not linked at LOC score 4.0. Contrary to expectations, the linkage map was not clustered. A protocol for high throughput marker collection was published.

Phase 4: Inheritance of growth rate trajectory
There are few reports of inheritance patterns for our trait of interest, growth curves. We estimated the quantitative inheritance patterns by estimating genetic parameters of growth curves from a series of old and young genetic tests planted throughout the P. taeda range. All parameters in the growth curve have inheritance patterns that parallel any age-specific height. These results have been published.

**Phase 5: QTL analysis and founder-origin probabilities**
Detection of major factors affecting height growth rates was tested using the new computer programs and founder-origin probabilities.

We tested for genome-wide distribution of QTLs influencing height from ages 2 to 6 years using a Weyehaeuser three-generation pedigree (BASE map). The linkage map was constructed using CRI-MAP program. The interval regression analysis program (SATORI) developed for QTL mapping in outbred pedigrees was used to estimate QTL number, magnitude of QTL effects and relative position in the linkage map.

A large QTL in linkage group 6 showed consistently large effects (10 to 14%) at all ages 2-10 years of age. These results have been accepted for publication.

We have now asked the question in retrospect: could we have used data-mining methods to search for pedigrees with large QTLS? The results show that two of the five data mining methods detected this large QTL and that there appear to be even better full-sib crosses in the same diallel designs than this one full-sib cross.

**Phase 6: Population survey for P. taeda**
We surveyed the entire range of P. taeda to determine allelic frequency patterns. This began as a preliminary study to determine if the Lost Pines, a disjunct *P. taeda* population has sufficient linkage disequilibrium to serve as founder population for population-level mapping. The linkage equilbrium estimates are high; there is no merit at this level to disequilibrium mapping. We have shown that *P. taeda* populations east of the Mississippi River Valley have significantly higher mean allele numbers (10.0) than populations west of the Mississippi River (8.5). These values have been adjusted for sample size and geographic area. One explanation is that prevailing winds have moved pollen west to east during pollination since the last glacial maximum or the past 21,000 years. Another explanation is that the western range of *P. taeda* has expanded then contracted eastward with climate changes over the last 5000 years (a bottleneck).This work has been published and is now part of a doctoral dissertation. The doctoral student was supported on funds independent of this project but the data collection was supported by Agenda 2020. This phase was accomplished without additional cost to DOE.

# MANUSCRIPTS SPONSORED BY DOE AWARD

1. Gwaze, D.P., Y. Zhou, M. H. Reyes-Valdes, M. A. Al-Rababah and Claire G. Williams. Haplotypic QTL mapping in an outbred pedigree. *Genetical Research* (accepted)

2. Reyes-Valdés, M.H. and Claire G. Williams. A haplotypic approach to founder-origin probabilities and outbred QTL analysis. *Genetical Research* (in press).

3. Gwaze, D.P., F.E. Bridgwater and Claire G. Williams. 2002. Genetic analysis of growth curves for woody perennials: a case study for *Pinus taeda* L. *Theoretical and Applied Genetics* 105: 526-531.

7. Zhou, Y., T. Bui, L.D. Auckland and Claire G. Williams. 2002. Undermethylation as a source of microsatellites in large plant genomes. *Genome* 34: 91-99.

8. Zhou Y, T. Bui, L.D. Auckland and Claire G. Williams. 2002. Direct fluorescent primers are superior to M13-tailed primers for automated genotyping of *Pinus taeda* microsatellites. *Biotechniques* 32: 46-52.

**Submitted Manuscripts**

Zhou, Y., D. P. Gwaze, T. Bui, M.H. Reyes-Valdes and Claire G. Williams. No clustering for linkage map based on low-copy and undermethylated microsatellites (Manuscript in review).

Gwaze, D.P. and Claire G. Williams. Early selection criteria as viewed through analysis of quantitative trait loci (Manuscript in review).

Gwaze, D.P., R. Pong-Wong and Claire G. Williams. Testing phenotypic data mining methods with a major QTL allele. (Manuscript in review).