The Loblolly Pine Genomics Project

A community-developed research plan intended to:

- **Guide and prioritize R&D efforts in pine genomics**
- **Assist with the development of research collaborations**
- **Provide rationale and background for improving the funding status of genomics research in loblolly pine in the United States**

Third Edition
September, 2004

Prepared by
Nicholas C. Wheeler, Molecular Tree Breeding Services
and
David B. Neale
Institute of Forest Genetics
USDA Forest Service

Table of Contents

Executive Summary

Loblolly pine (*Pinus taeda)* provides ca 16% of the world's annual timber supply, and grows on nearly 58 million acres of plantation and natural forest in the southeastern United States. Annually, only corn exceeds timber in farmgate economic value in America. Forest trees are the ecologically dominant life-form on ca 275 million acres in the US, and play a critically important role in carbon sequestration. In short, forests are vital to the economic and ecologic landscape of this country and loblolly pine is the single-most valuable species in those forests.

A genomics approach to describing and understanding the genetic and molecular basis of all biological processes controlling economically and ecologically relevant traits in pine is both feasible and desirable. Excellent progress in gene discovery, marker development and QTL mapping has been made in loblolly pine and near term breakthroughs in functional genomics and physical mapping are anticipated. Furthermore, pine is nearly ideal for conducting association genetic studies; current efforts in this area are among the most progressive in the plant genomics community. There is a ready market for application of genomics tools. Progress in genetic improvement of many forest tree species over the last half century has been notable (~10% increase in volume growth per generation), but traditional breeding and testing programs are expensive, time-consuming (15 to 25 years/generation), and very restrictive in the number of traits addressed in a given population. Furthermore, trees possess an abundance of natural variation that makes the potential for tree improvement large, but progress using traditional means slow. Few, if any, crops would benefit more from the development of genomic technologies which would enhance our understanding of biological processes.

Support for genomics research in conifers (cone bearing trees; gymnosperms) has lagged significantly behind most major agricultural crops and model species. Initiatives are required from the pine genomics community and those who would benefit from that research (public and private land managers/owners, forest, paper, and energy industries) to enhance research support and speed progress. The purpose of this document is to assist with initiative development by providing useful and complete information on the status of genomics research in loblolly pine today and suggest means for meeting 5 year R&D targets. It is anticipated this information will a) help guide and prioritize future R&D proposals, b) help scientists identify useful research collaborations, and c) provide rationale and background to improve funding efforts.

The overall benefit of the application of genomics to loblolly pine will be a vastly improved understanding of the biological and molecular basis of adaptive and economic traits of the most dominant genus of forest trees in the world[1] and significantly improved precision of genetic improvement practices that will surely reduce the time and cost of these activities.

[1] There are over 110 species of pines and comparative genomics studies show they are remarkably similar, genetically.

Introduction

Species Economic and Ecologic Importance

Loblolly pine is one of the most important crop species, and clearly the most important commercial timber species, in the US. Its native range spans 14 states from southern New Jersey south to central Florida and west to Texas where it makes up more than half of the standing pine volume in the region. In 1998, 75% of the 1.6 billion seedlings planted in the United States were loblolly pine (Moulton and Hernandez, 2000). It is the dominant tree species on 11.7 million ha of native forest (Baker and Langdon, 1990) and is established on over 12 million ha of plantation (Byram, 1999). The southern states provide 58% of the timber in the US and 16% of the world's timber (Wear and Greis, 2002). Collectively, timber is among the most highly valued commodities in America.

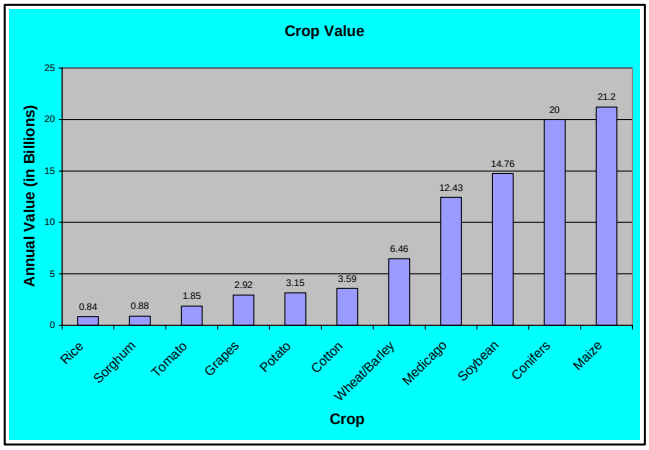


Figure 1 Annual crop value (in $billions)

Viewed as an agricultural crop, only corn currently exceeds timber in farmgate value on an annual basis (**Fig. 1**; extracted from http://www.usda.gov/nass/ and USFS reports). The value of finished wood products exceeds 200 billion dollars a year! Furthermore, as the dominant plant species on millions of hectares, loblolly pine provides a huge and renewable/sustainable resource for carbon sequestration for a significant portion of the US, and is therefore critical in the larger picture of climate change.

Programs to enhance loblolly pine productivity have historical roots. Loblolly pine tree improvement programs began in the southern states in the 1950s but even the most advanced programs have completed just 2-3 generations of breeding and testing (Li et al., 1999). The long generation time of pine, combined with phenotypic selection of mature-tree traits, results in a very slow rate of genetic domestication. Consequently, loblolly pine would benefit tremendously from the development of any genomics technology that could accelerate breeding and improvement efforts.

It is important to note that what is learned in loblolly pine will likely have utility across much of the Pinaceae, a family containing great ecological value world-wide as well as most of the commercially important species of the world. The genus *Pinus* alone contains more than 110 species, or about 20% of all known gymnosperms (Richardson and Rundel, 1998). Comparative genetic mapping among conifers has demonstrated high levels of genetic similarity among Pinaceae genomes, facilitating comparative genomic analysis in this important plant family (Brown et al., 2001; Krutovsky et al., 2004).

Evolutionary and Biological Significance of Working with Pine

Gymnosperms are evolutionarily ancient, having arisen as much as 300 million year bp (see **Figure 2**, courtesy of http://www.ornl.gov/sci/ipgc/home.htm). The conifers, an order to which the pines belong, separated from flowering plants (Angiosperms) approximately 100 million years ago and by most measures, have evolved very slowly and conservatively. The pine genome is huge in base pair (bp) content (>2 x 10^{10} bp), exceeding that of Arabidopsis (1 x 10^{8} bp), poplar (5 x 10^{8} bp) and human (3 x 10^{9} bp) genomes by orders of magnitude. However, the vast majority of the pine genome appears to be repetitive (>99%), and is characterized by large gene families with many pseudogenes.

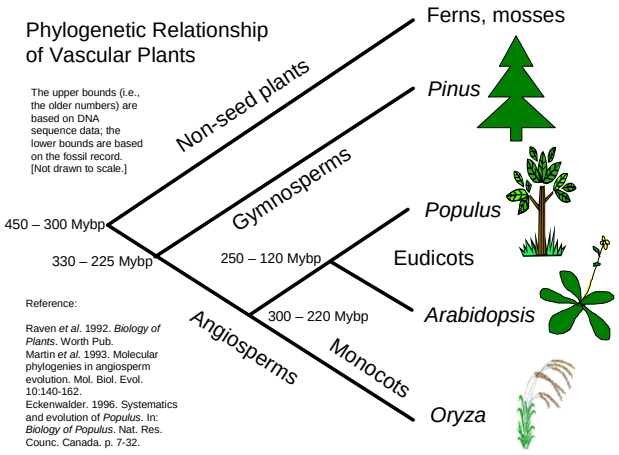


Figure 2

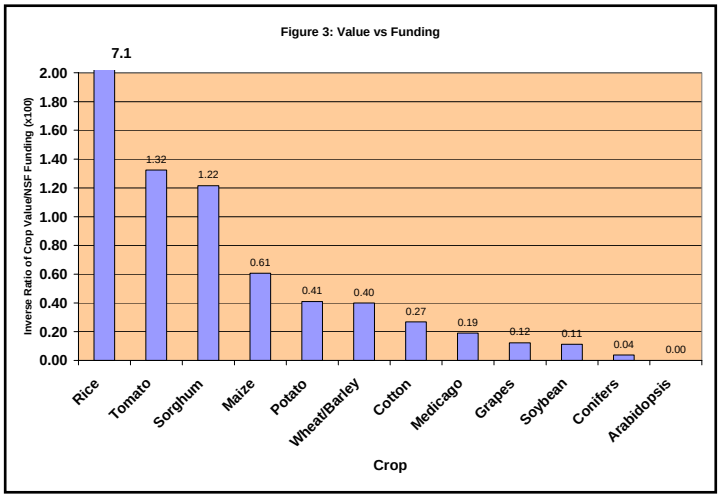
Though whole-genome sequencing of the pine genome would be prohibitively expensive and time-consuming at this time, prospects for gene-space sequencing are attractive.

Large public EST discovery projects, though by no means complete, are revealing a great deal about the pine genome. Significantly more effort is required to identify a complete unigene set.

While pines offer unique challenges for some aspects of genomics research, they possess many characteristics that make them good models for such studies. They are long-lived and can be easily cloned, making permanent field trials in variable environments feasible for the study of genotype/phenotype relationships. Pines are among the most genetically diverse plants on earth (H_o ~ 3.0), and they typically occur in large, out-crossing, natural populations with high gene flow and little population substructure. They possess a large, haploid megagametophyte which makes direct haplotype determinations possible. Additionally, a suite of other traits makes pine ideal for association genetics: they possess high levels of nucleotide diversity, limited linkage disequilibrium (<2500 bp) and large detection and verification populations can easily be generated.

Status of Funding for Loblolly Pine Genomic Research

Funding for genomics studies of pines (trees in general) have lagged significantly behind agricultural crops relative to their overall value and contribution to the countries economy. A contrast of crop value to total NSF funding between 1998 and 2003 demonstrates this relationship (Figure 3). The figure graphically illustrates the ratio of the crops total funding for genomics research from NSF (7.5 million dollars US) to its annual farmgate value (~20 billion dollars US) x 100 (= 0.04 for conifers).



Total funding for pine genomics research from all sources since 1988 is slightly over $18.0 million dollars (**Appendix 1**). While this is a substantial amount of funding, it represents 30 separate grants to 11 different institutions, often with overlapping or non-complementary research goals, delivered over 16 years. In short, funding levels and continuity have not been conducive to making rapid progress in genomics research of pines in the US. Members of the pine genomics community in the US have determined that a collective, integrated approach to conducting research and acquiring federal funding is required to improve the rate of scientific discovery and progress in this important crop species. They are proposing the creation of a virtual organization called: ***The Loblolly Pine Genomics Project.***

The Loblolly Pine Genomics Project (LPGP): Purpose and Goals

Purpose: The purpose of the LPGP is 3-fold:

- *To assist with the development of research collaborations among members of the genomics community*
- *To help guide and prioritize R&D efforts in pine genomics*
- *To provide rationale and background for improving the funding status of genomics research in loblolly pine in the United States*

The creation of the LPGP is a work-in-progress, with membership and participation informal and voluntary. There are no organization officers. At present there are four working groups that have been charged with the creation of short reports as elements of a research plan (see subsequent sections of this report). The LPGP has grown out of a pair of informal workshops, the first hosted by the US Forest Service / University of California at Davis, CA in May of 2003 and the second held at Jekyll Is, GA in June, 2004 (**Appendix 2: participant list**). The first meeting, attended by some 20 academic

and government scientists, produced a document that summarized the state-of-the-art in pine genomics in the US, identified critical needs to move forward, and listed action steps for the next 5 years. In addition, a website (http://dendrome.ucdavis.edu/lpgp/) was created to post the report and a list of participating scientists / organizations. The second meeting sought to expand participation of the scientific community, engage the timber industry as full partners in the development of this critical field of science, and to update the status of the previous report. This document is meant to serve the latter function as well as to address the overall purpose of the group.

Goals: Ultimately, the scientific goal of the LPGP is

to describe and understand the genetic and molecular basis of all biological processes controlling economically and ecologically relevant traits in pine by applying and advancing state-of-the-art technologies and, through open scientific collaborations and exchanges of information.

The technology transfer and outreach goal is

to reduce the time and expense, and increase the genetic gain potential, of tree improvement programs by providing requisite tools and molecular reagents to the industrial community in a timely manner.

Each of the separate working groups has prepared a brief synopsis of the state-of-the-science for their specialty area, and defined specific needs that must be met over the next 5-10 years if progress in pine genomics is to be competitive with other crops, and useful to the industrial community. Prioritized needs are summarized in **Appendix 3.**

Literature Cited

Baker, J.B. and O.G. Langdon. 1990. Silvics of North America. Vol. 1: Conifers. Agric. Handbook 654. US Dept. Agric., Washington, D.C.

Brown, G.R., E.E. Kadel III, D.L. Bassoni, K.L. Kiehne, B. Tesmegen, et al. 2001. Anchored reference loci in loblolly pine (*Pinus taeda* L.) for integrating pine genomics. Genetics 159: 799-809.

Byram, T.D., W.J. Lowe, and G.D. Gooding. 1999. Western gulf forest tree improvement program gene conservation plan for loblolly pine. For. Gen. Res. No. 27, FAO, Rome, Italy. Pp. 55-59.

Krutovsky K.V., M.Troggio, G.R. Brown, K.D. Jermstad, and D. B. Neale. 2004. Comparative mapping in the Pinaceae. Genetics (in press).

Li, B., S.E. McKeand, and R.J. Weir. 1999. Tree improvement and sustainable forestry impact of two cycles of loblolly pine breeding in the U.S.A. For. Gen. 6:229-234.

Richardson, D.B. and P.W. Rundel. 1998. Ecology and biogeography of Pinus: an introduction. In: Richardson, D.M. (ed), Ecology and Biogeography of Pinus. Cambridge Univ. Press, Cambridge, UK. pp. 3-46.

Wear, D.N. and J.G. Greis. 2002. Southern forest resource assessment: summary of findings. J. For. 100: 6-14.

Working Group Reports: Status and Needs

Four scientific working groups (gene discovery, functional genomics, maps and markers, and genetic stocks and germplasm) convened at the second LPGP workshop to update and revise the 5-year action plans identified the previous year. Additionally, a working group comprised of industrial participants met to discuss how they might be able to expedite, guide and assist the progress of the scientific community. A sixth topical area, bioinformatics, though not treated separately at the workshop, was discussed at length in the other working groups.

Gene Discovery and DNA Sequencing

Working Group Members: Jeffrey Dean (Chair), University of Georgia; John Cairney, IPST at The Georgia Institute of Technology; Marie-Michèle Cordonnier-Pratt, University of Georgia; Daniel Peterson, Mississippi State University; Ronald Sederoff, North Carolina State University

The long-term goals of this working group are to locate all genes in the transcriptome and provide full-length cDNA sequence for 90% of all superscripts, estimated to be about 35,000 in number and obtain a reference DNA sequence by first sequencing gene-rich regions and then a complete sequence.

General questions:
The gene discovery and DNA sequencing working group identified a series of questions that defined the current status and needs of the pine genomics community. Specifically,

- What is the structure of the pine genome? Are genes arrayed in islands within the largely repetitive motifs or are they relatively well distributed throughout the genome?
- How does the genome of pine compare with the genomes of other organisms? Is it more like Arabidopsis or maize, or is it novel? Is it representative of gymnosperms or conifers in general?
- What genome component is functional? For instance, how many genes are there? What is the true nature of gene families? What level of abundance of pseudogenes exists in the pine genome? Are there relevant and meaningful patterns of methylation in the pine genome? Is alternative splicing important? How is intron frequency, size and distribution characterized in the pine genome?

Current Status: Existing or anticipated near-term resources
Tremendous progress has been made in gene discovery of loblolly pine over the last 5 years, primarily through NSF supported projects at North Carolina State University, University of Georgia, and IPST at Georgia Tech (Appendix 1). In addition, a newly supported NSF proposal for BAC and Cot library construction will contribute significantly to genomic resources in the community over the next few years. Specifically, a tally of public resources includes:

1. cDNA libraries for roots, wood-forming tissues, and reproductive tissues. As many as 720,000 cDNAs have been picked and archived at UGA alone. These represent 36 different tissue libraries challenged with physical, chemical, biological and nutritional stress factors.
2. Public EST accessions numbering ca. 250,000 should exist by September 2005.
3. A 10x BAC library is scheduled for completion ca. September 2006.
4. Cot libraries (high-, moderate and low) are anticipated to be constructed within the next year. Sampling of 10,000 clones (= 99.5% coverage in HRCot, 0.08% in MRCot, 0.10% in SLCot libraries).

5. Pilot probes of repetitive and unique sequence distribution across the genome (BACs) will be completed by September 2007.
6. FISH mapping of 120 single-copy markers (10/linkage group) to pine chromosomes, and hybridization to the BACs is anticipated to be completed by September of 2007. This will provide connection between physical maps and linkage groups.
7. Data repositories currently exist at the Univ. of Minnesota, UGA, TIGR, and GenBank with varying levels of integration. See TIGR, GenBank and UGA for Pine Gene Indices and Unigene sets (SuperScripts and UniScripts at UGA)

It is worth noting that very large pine EST libraries exist in the private sector (>500,000 ESTs), though these are currently not available to the public, and were generated largely from radiata pine. There is reason to believe these might be made available to the academic community in the near future, but no specific plans currently exist for their release.

Needs:
The working group feels there is a critical need for the following resources:

Genome

1. Increased amount of genomic DNA sequence (Cot and methyl filtration subgenomic libraries, BAC ends)
2. Placement of ESTs, SSRs, and other markers on BACs (provides binning for QTL mapping)
3. Techniques for high-throughput placement of ESTs on chromosomes (by FISH) and on BACs

Transcriptome

4. Deeper EST sequencing
5. Creation of full-length cDNA sequences
6. Transcriptome scanning (SAGE? MPSS?)

Bioinformatics

7. Increased/continued integration across datasets
    1. More accurate clustering of all datasets; integration of data derived from EST, SAGE, MPSS and others expression tags that might be available
    2. Deeper data mining of sequences for domains to improve clustering and use in discovery of gene expression control
8. Data mining tools targeting smooth integration of different data types such as sequence data, expression data (whether from microarray or digital expression), map data, genotyping and phenotyping. Interfaces should provide easy visual interpretation of data types integrated and complex querying capabilities (for example, what genotypes offer new genes and/or expression profiles that correlate with drought resistance)

5-year Targets

The gene discovery and sequence working group identified the following targets, to be reached over a 5 year period.

1. Obtain 3' and 5' sequences from an additional 500K ESTs from a broader array of tissues. We propose sequencing 100 libraries 5000 clones deep. The estimated cost for this is ca. $3-4 million US.
2. Complete a collection of full-length cDNAs representing 90% of all Superscripts (ca. 35K sequences, ca. $3-4million)
3. Complete increased, deep analysis of moderate and low-Cot libraries.
4. Conduct pilot studies of methyl filtration libraries.
5. Placement of a suite of ESTs/gene families (up to 50) on the BAC grids
6. Initiate placement of BACs and markers on chromosomes
7. Perform a transcriptome scan in a variety of important tissues using SAGE and/or MPSS in parallel with EST discovery (ca. cost of $1 million)
8. Continue bioinformatics development in parallel with and supportive of the preceding wet-lab targets

In addition, the working group identified the following items as 10 year targets:

1. The completion of a physical map of loblolly pine.
2. The completed integration of the physical map with genetic map
3. Additional genomic sequencing targeted to BAC islands of biological interest (e.g., specific QTLs)
4. Increased linkage of genes (SuperScripts) to the BAC physical /genetic maps
5. Extended distributed-data-mining capabilities

Functional Genomics:
Working Group Members: John Davis, Chair, University of Florida; Sarah Covert, University of Georgia; Ulrika Egertsdotter, Virginia Tech; Len van Zyl, ExpressArrays; Gary Peter, University of Florida; Joe Nairn, University of Georgia; Lee Pratt, University of Georgia; Bailian Li, North Carolina State University; Will Rottmann, ArborGen.

Introduction

Functional genomics of loblolly pine has the overall goal of understanding the genetic control of growth, development, metabolism, adaptation, and evolution in this model species. A fully successful functional genomics effort will identify the relationship between genotype and phenotype for every loblolly pine gene. As functional genomics is carried out in loblolly pine, desirable scientific and practical advancements will result in an improved understanding of pine biology (including adaptive variation and ecosystem function), and an acceleration of its domestication for specific end uses.

Genomic Resources: Current status

It is anticipated that 250,000 ESTs from a variety of tissues (stem, embryos, needles, roots), representing approximately 30,000 different genes will be submitted to public databases by mid-2005.

Creation of microarray resources including 1) ca 14k element cDNA array (UGA), 2) a 2200 element xylem array (NCSU), and 3) an ca 3500 element embryo array (GTU) should be completed by the end of 2004.

A 10X coverage BAC library with shallow sequencing for characterization is currently underway at Mississippi State University; molecular markers assembled into maps include ~100 SSRs (Claire Williams, Dana Nelson), >170 RFLPs, and numerous RAPDs. Mapping of ~300 SNPs (Neale) is anticipated within the next year. These same maps have placed nearly 100 candidate genes along side dozens of QTL for wood properties.

Database capacity exists for a variety of applications, but more work is required to link these resources and to identify permanent sources of support. The Dendrome (TreeGenes) database provides organization for all manner of mapping information, but lacks institutional support. A series of databases exist and/or are already funded to make publicly available phenotypic data (disease and drought resistance, Forest Biology Research Coop (UF); wood quality, IFAFS (NCSU); and wood quality, DOE Agenda 2020 (UGA)), and genetic information (xylem ESTs (UMN); root ESTs (UGA); microarray (UGA, NCSU, UF)).

Large, replicated field trial resources with hundreds of clonally replicated genotypes exist with associated phenotypic data in cooperatives and in corporate tree improvement programs.

Proven quantitative genetic and statistical expertise exists in the community for dealing with real-world phenotypic data collected from long-term field trials (NCSU, UF)

Loblolly pine transformation rates are comparable to that of corn and are potentially available the research community on a service basis.

An enormous plantation estate containing genetically defined germplasm exists for loblolly pine. The genetic level of resolution in this estate varies from open pollinated mixs to half-sib and full-sib crosses, to clones.

Loblolly pine is amenable to cloning through rooted cuttings, somatic embryogenesis and by grafting.

<u>The loblolly pine functional genomics community could benefit from:</u>

- support for innovations in exploring the unique biology of loblolly pine, including but not limited to 1) discovering genes unique to novel cell types and developmental stages such as epithelial cells, vascular cambium and tracheids, and 2) creation of high throughput precision phenotyping of tree processes such as wood properties and whole-tree architecture.

- development of a publicly available glass slide microarrays. There is a need to make technical improvements in representation of all genes, and sensible QA/QC by centralizing the creation of a common integrated array for all researchers. Integrating microarray data into existing databases, improving data mining tools to integrate across experiments, better annotation of microarray data are all needed to improve this technology.

- a sensible, agreed-upon means for integration (including annotation) of EST data from multiple groups.

- development of full-length cDNA sequences to better understand the function of ESTs, and to ultimately assist in annotation of genomic sequence.

- participation of the larger community of pine ecologists, physiologists, pathologists to creatively phenotype trees – including phenotypes that emerge at the stand level and are therefore not observed in isolated trees

- public availability of loblolly pine transformation as a research tool for reverse and forward genetic approaches to functional genomics

- sequence of all the loblolly pine genes (or as many as possible) including their allelic forms

- meaningfully linked relational databases for phenotype data, including phenotypes from microarrays, metabolic profiles, proteomics, and whole-plant

properties. The phenotypic databases must be coupled with the allele sequence information that will facilitate association genetics, forward genetics and reverse genetics.

- improved statistical methodologies for improved precision in phenotyping and association genetics

- development of good physical maps of the genes expressed in loblolly pine, using approaches such as BAC contig assembly and/or cytogenetic approaches to integrate with the genetic map.

- improved understanding of the regulatory factors that control expression of pine genes, including cis-elements in promoters and the trans-acting factors such as transcriptional regulators and proteins in signal transduction cascades.

Five Year Plan

The following activities are recommended to meet functional genomics goals in the next 5 years.

- Develop and submit a proposal to have pine considered for genome sequencing.

- Convene a working group to investigate means of successfully integrating all current pine genomic/phenotype databases to facilitate functional genomics research.

- Develop an affordable loblolly pine gene chip that can be used for genotyping and gene expression phenotyping in single experiments.

- Create and make available to the public, a replicated (clonally propagated) reference population of 500-1000 individuals for association genetics.

- Obtain sequence information on a substantial number (>500) of promoter elements, and the regulatory cascades that interact with these elements, to reveal new insights into what makes a pine tree.

- Develop the capacity (land, management activities, etc) to establish and maintain field test sites for genetic studies; these sites are required for conducting phenotypic analysis of genotypes generated through conventional crossing or transgenesis.

- Seek improvements in integrating phenotypic data across all levels of biological organization, including metabolic, proteomic, cellular, whole-tree, and stand levels. This will involve new partnerships for our research community to include

for example physiologists, ecologists, silviculturists and engineers. The working group working on database integration could address this issue.

- Develop new partnerships to enhance computational approaches at a genomic scale in order to meaningfully integrate vastly different types of phenotypic data, and ensure their use by the research community.

- Provide technology transfer opportunities to the pine genomics community and their collaborators to take advantage of advancements in cell biology to enhance resolution of phenotypic analysis. [For example, high throughput *in situ* hybridization and laser capture micro-dissection of individual cells].

- Generate a substantial number of full-length cDNA sequences (>10,000) in order to complement genome annotation and functional genomics research efforts.

- Develop a publicly accessible transformation service to facilitate unrestricted research applications.

o

Mapping and Markers

Working Group Members: Dana Nelson, Chair — Southern Institute of Forest Genetics (SFIG); David Neale, U.C. Davis / USFS Institute of Forest Genetics (IFG); Matias Kirst, University of Florida; Henry Amerson, North Carolina State University, and Jill Wegrzyn, U.C. Davis

We discussed both genetic and physical mapping, although 80% of our effort and time was spent on genetic mapping. This was due to the group's interest and expertise as well as the overarching need to complete a highly transferable reference/consensus map for loblolly pine. Towards that end we felt a concerted effort on SSR and SNP based marker development was needed as well as unrestricted access to a model (i.e., reference) mapping population. The genetic mapping discussion is summarized below starting with markers, moving to mapping populations, and ending with a proposal for a stock center. For physical mapping we summarized the key aspects of the recently funded NSF loblolly pine genome proposal and brainstormed on the most efficient path forward to complete a draft genome sequence.

Genetic Maps

Preferred or Recommended Markers

Microsatellites: Microsatellites or simple sequence repeat (SSR) markers are the preferred markers for all future mapping operations in loblolly pine. Advantages to these markers include:

- they are PCR based
- primer sequences are/should/can be made available on treegenes web site (http://dendrome.ucdavis.edu/treegenes.html)
- they are highly polymorphic
- they are transferable across pedigree (sometimes across species)

Emphasis should be on SSR marker development and mapping using the IFG reference populations (BASE and QTL). The Southern Institute of Forest Genetics (SIFG) has a goal of mapping 300 (200 from genes + 100 from genomic clones) SSRs using one or the other IFG mapping populations. The 200 SSRs from genes is an ongoing Agenda 2020 project with Mississippi State University (MSU). In addition SIFG will mine SSRs from Peterson's (MSU) NSF Plant Genome Project. Overall, 1000 SSRs should be mapped for applications in quantitative trait locus mapping and marker assisted breeding.

In addition it is very desirable to develop a 12-16 marker set for routine use in genotype identification. A species-wide allele frequency database of at least 1000 individuals will be needed for this marker set.

Gene markers: Gene markers need to be sequence-based, such that they can be made readily available on the treegenes database. It is likely that these markers will be SNP

based and developed detection methods also will need to be made available on treegenes.

IFG has a goal to map all the contigs from the various EST projects (on the order of 30,000 unigenes). SIFG and University of Florida (UFL) are planning to use the IFG reference populations to map SNPs discovered in a newly created Southern Pine Beetle (SPB) resistance project.

Others are encouraged to map their genes with IFG reference populations and make the data available on treegenes. The goal is to map all the genes of loblolly pine.

Mapping populations

Currently the LPGP community has two mapping populations available—BASE and QTL, courtesy of Weyerhaeuser Company (WeyCo) and IFG. They are both complete three-generation pedigrees, with 90 and 96 progeny DNA samples available. DNA sample requests can be made through Treegenes. All trees in the pedigrees are owned by WeyCo and have been immortalized by grafting at a company site. The LPGP needs to make sure these are available for research purposes on a continuing, non-interrupted basis. This seems best done by a MTA transferring the material as grafts from WeyCo to the SIFG. SIFG would be a natural host as it has ample land in the native growing region and the interest and expertise to accomplish the goal. Some funding will be needed and SIFG (& IFG) will seek funds through SRS and WO of the US Forest Service. The MTA would allow SIFG to graft the material for immortalization purposes and distribute tissue and/or DNA samples to others for research purposes. Both SIFG & IFG will pursue this with WeyCo.

In addition, several other populations are in wide use and should be considered as LPGP community resources. These are listed below.

- Loblolly Pine Association Population (~500 clones, provided by NCSU-I and WGFTIP Cooperatives) should be immortalized. IFG has MTA in place for the purpose of a single experiment. MTA would be needed to allow grafting (for immortalization) and distribution of tissue or DNA samples. The same model for SIFG to host the immortalized material as discussed above should be followed. In addition we discussed the possibility of doubling this to 1000 (by adding 500 loblolly pines from the USFS tree improvement program).

- Cclones (~1000 clones) should be considered for immortalization (at least with massive quantities of DNA, if not grafted to clone bank). Questions about MTA are necessary and will be asked of FRBC (UFL).

- Development of a 3rd (4th or 5th) three-generation reference population should be considered in cooperation with one or all of the Tree Improvement Coops or USFS (SRS and/or R8).

- Populations beyond these especially for the purpose of QTL mapping or fine mapping were discussed and determined to be outside the scope of LPGP. However, the LPGP group as a whole thought this might be reconsidered.

Loblolly Pine Stock Center

The establishment of a Loblolly Pine Stock Center should be a high priority. It would be run by IFG and housed at UC Davis. Two other stock centers are being developed at UC Davis and IFG is working on adding pine. Reagents would be available on a cost recovery basis and requested through Treegenes. Reagents should include

- DNA samples from all reference mapping populations (SIFG would host the live, grafted material and provide needle samples for IFG to make and distribute DNA).
- EST clones from all pine EST projects
- All marker sequences (SSR primers, SNP primers, technique for generating marker data)

Physical Mapping

Physical mapping and genome sequencing in loblolly pine received a much needed boost this summer with the NSF funding of Dan Peterson's project at MSU. The funded work includes the following:

- 10X library produced by MSU and distributed by Clemson University Genome Institute (CUGI).
- Chromosome specific karyotype and cytogenetic markers produced by SIFG (at Texas A&M) available through the Pine Stock Center & Treegenes.
- Integrated reference genetic maps and karyotype (MSU, SIFG, IFG).
- Characterized BAC library by Cot (reassociation kinetics) class.
- Characterized distribution of Cot classes (genespace vs. repeat space) by fluorescent in situ hybridization (FISH).
- Complete genome sequence for High Repetitive Cot Class (HR).
- Sample genome sequence for Middle and Low Repetitive/Single Copy Cot Class (MR and LRSC).

We brainstormed a bit on what might follow, in particular on how to most efficiently get to a draft genome sequence. The following list these steps.

i. Complete genome sequencing of MR and LRSC using Cot Based Cloning and Sequencing (CBCS) technology.

Number of clones needed to get 90% of the genome sequence using CBCS.

Class	avg clone size	nr clones	cost @$4/clone
HR	350	500	$2,000

MR	350	1,118,,000	$4,472,000
LRSC	750	14,500,000	$58,000,000

Note that CBCS provides a ~5-fold decrease in the effort/cost required to sequence the pine genome (http://www.msstate.edu/research/mgel/other_pdf/cbcs_table%20.pdf) relative to shotgun sequencing methods.

ii. Ordering the BACs—BAC-end sequencing (try to fund through DOE at Joint Genome Institute) & BAC fingerprinting to create a minimum tiling path (the physical map).

iii. Integrate physical map with genetic map (& karyotype) through cytomolecular mapping with FISH.

iv. Order Cot clones by hybridizing to physical map and chromosomes, providing the genome sequence.

Genetic Stocks and Germplasm

Working Group Members: Barry Goldfarb -Chair, North Carolina State University; Tim Mullin and Steve McKeand, North Carolina State University; Tom Byram and Larry Miller, Texas A&M University; and Joe Weber, Boise Cascade.

The ultimate goals of the 'genetic stocks and germplasm' working group are to 1) identify, immortalize, and make available to collaborators, DNA and living plant material of reference populations used for mapping and association genetics studies, DNA libraries and pooled DNA samples, and supporting information on all collections and 2) develop 1 or more permanently funded germplasm centers that hold these reference collections.

Resources available

Genetic stocks

Several laboratories have compiled collections of DNA samples. Existing DNA collections fall into two categories—DNA of cloned genes or genomic DNA of populations of trees. In most cases, collections are maintained by individual laboratories, however, in some cases, collections have been duplicated and stored in more than one laboratory.

Examples of DNA collections of cloned genes include:

- NCSU/Sederoff et al.: 65,000 ESTs, largely from xylem (NSF)
- UF and UGA, Davis & Covert et al.: ~400 genes differentially expressed during disease interactions, with plans to expand to ~5,000 in the near future (IFAFS project – ADEPT)
- UGA/Dean et al. ~720,000 clones picked largely from roots of plants exposed to different levels of drought stress, other tissues/conditions to be added as become possible(NSF)
- IPST/Cairney et al. ~15,000 ESTs, from zygotic and somatic embryogenesis.

Examples of DNA collections from populations of trees include:

- IFG/Neale – 2 reference mapping populations: 96 3rd-gen trees along with 2 parents and 4grandparents for each;
- IFG/Neale et al. DNA from 1400 pine tree clones (from Forest Biology Research Coop CCLONES study), derived from ~60 full-sib crosses;
- Goldfarb/Neale association population—Megagametophytes and diploid tissue (DNA not yet extracted) from 500 unrelated individuals from across the loblolly pine range (from NCSU and Western Gulf Tree Improvement Programs);

USFS/Texas Forest Service/TAMU Founder Population—foliage samples for DNA isolation have been collected from all available plus tree selections in the three breeding programs (see below).

Germplasm Loblolly pine is one of the few species for which ermplasm is available both in widely distributed natural populations and in economically based breeding programs. This is largely the result of the relatively undomesticated status of the species, but it represents a unique scientific opportunity.

Natural populations of loblolly pine still exist, although most of the region underwent significant disturbance as a result of agricultural land clearing. The natural populations that exist now are largely the result of seeding in on abandoned farms. As time passes, these natural populations are getting smaller and fewer. The availability of natural populations in certain regions may be threatened. Because of the overall large size of the natural population, it is unlikely that many important alleles have so far been lost.

Commercial breeding populations are maintained by the members—both private companies and state agencies—of the three major breeding cooperatives, NCSU-ICTIP, WGTIP (Texas FS), and CFGRP (UF). These populations have been primarily selected for rapid growth, straight boles, and to some extent, fusiform rust resistance. They do not constitute a random sample of pre-selection populations.

The trees with higher breeding values (commercial worth) are generally well archived for genetic conservation. However, less commercially valuable trees are less well archived and some have already been lost. Formal genetic conservation strategies vary among the coops and their members, but more effort is expended on commercially valuable material than on material without known commercial value.

It is highly probable that most, if not all, of the original gene diversity in loblolly pine is still represented in these selected populations. Common alleles should be well conserved, while the fate of rare alleles (< 1%) is less certain.

Seed storage: Most of the seed of loblolly pine is stored by individual companies and state agencies and, again, there is a heavy emphasis on commercially valuable seed by today's standards. Some seed is stored at the facilities of some of the breeding coops, although these tend to be small collections during active breeding and testing or specialized research collections. There is an indication that some loblolly pine seed is stored at the National Seed Center, however, the extent and nature of the curated collection is not generally known to the pine genomics community.

Ability to conserve and produce ermplasm: Seedlings- many organizations, including private companies and state agencies operate commercial nurseries that produce bare-root loblolly pine seedlings (>1 billion produced annually). In addition, on a smaller scale, there are commercial containerized seedling producers and numerous companies, agencies and universities produce seedlings (usually containerized) for research studies. Thus, given high-quality seed, there are numerous options for producing quality seedling stock.

Cloning — One very useful tool for studying genomics of loblolly pine is the ability to clone individual genotypes. By replicating trees of a given genotype, a much more precise estimate may be made of the genetic vs. environmental contributions to gene expression and phenotype. Currently, there are two relatively common methods for cloning loblolly pine trees—rooting stem cuttings and somatic embryogenesis. Both methods rely on starting from juvenile material; mature genotypes may not be cloned. In addition, it is possible (and relatively routine) to multiply the shoot system (but not the roots) of mature trees by grafting (not considered further here)

- Rooted cuttings: The technology to root stem cuttings is not exceptionally difficult, although it does require some specialized facilities and experience. Currently, this technology is well established at NCSU and in several companies. The ability to propagate loblolly pine by rooted cuttings extends to most genotypes. However, in a short period of time, relatively few copies (ramets) of a given genotype can be produced. In addition, after clonal hedges are maintained for a number of years, maturation may affect gene expression and phenotypic expression.

- Somatic embryogenesis: This technology is more specialized and requires tissue culture facilities and extensive know-how. Currently, while there is research underway in the public domain to improve the process, there are no public facilities or labs that can routinely produce clones by this method. There are however, at least two private companies with extensive embryogenesis facilities and expertise. In general, the limitation to embryogenesis from a genomics perspective is the difficulty and expense of producing large numbers of embryogenic lines, thus limiting the application to population studies. On the other hand, fairly large numbers of ramets of a given clone can be produced, once a clone has successfully initiated an embryogenic culture. Moreover, cultures can be stored essentially indefinitely in liquid nitrogen. This can be useful for ermplasm storage purposes and for controlling maturation in clones to be used for genomic studies over long periods of time.

Scientific study populations: Currently, three populations of trees are available for various population genomic studies: (1) CCLONES (FBRC/ADEPT)--~1400 clones from approximately 60 full-sib crosses, originally from the Lower Gulf Elite Breeding Population—a joint effort of the three breeding coops. These clones have been established in replicated blocks (8) at 6 field test sites. (2) The USFS/ Weyerhaeuser (Neale) loblolly pine reference populations which consist of 96 achieved pine clones maintained in living clone banks in Arkansas, and (3) NCSU/USFS (Goldfarb/Neale) association population—500 unrelated individuals from across the entire range of loblolly pine were set up as an association mapping population. Seed was obtained from the NCSU and Western Gulf coops. Currently, these exist as seedlings at NCSU, but they will soon be pruned into hedges (stock plants for rooted cutting production) and a limited number of ramets produced. These clones will be planted in

several locations in the South. However funding for this effort has not yet been obtained and intellectual property issues must be resolved.

What resources are needed?

Genetic stocks

Ideally, there would be an infrastructure for a centralized, curated collection. Critically important, would be the distribution mechanism, which is an ongoing and resource-consuming task. It might not be ideal for this to be done privately, as the costs would need to be high to provide a profit margin and there would be no guarantee about stability of a private firm. Therefore, a publicly funded entity would be best. Because of the service function, this might not be appealing to a research laboratory, unless there was sufficient funding for it to be self-sustaining.

Germplasm

Ideally, a common set, or several common sets of plant material would be publicly available for research purposes. This would facilitate cross-referencing and integration across different studies and among different investigators. Germplasm can be maintained as seeds, as grafts (shoot system only), as hedged stock plants for rooted cuttings, or as cryopreserved somatic embryogenic cultures. The latter two options allow the possibility of the genotypes being propagated for new field or other phenotypic studies.

With the NSF-funded physical mapping project beginning at Mississippi State, there is need for a single, reference genotype for BAC cloning. The DNA and sequence information from this tree should be in the public domain, with unrestricted public access. It should also be a tree that is well-archived as grafted plants in a number of locations to prevent loss. Ideally, pedigree information (grandparents, parents, and/or progeny) about the tree should also be available and it would be tied into various other study populations.

Propagation/Transformation

Rooted cutting propagation: While several private companies have the capability for producing large numbers of rooted cuttings, and universities could develop such a capability, currently only NCSU has the facilities and expertise to accomplish this type of plant production for research studies. Because of the time required to produce cloned trees of specific genotypes, advance planning and infrastructure support would be an asset.

Somatic embryogenesis: Routine use of this technology by the research community would be advantageous for two principal reasons: (1) it would allow for indefinite storage of genotypes (cryopreservation) in a form capable of producing new plants (embryogenic cultures) and (2) it is a platform for genetic transformation (see next section). Many advances in embryogenesis know-how have been made by the private

sector. Greater collaboration between commercial entities and the public research community would be beneficial.

Genetic transformation: The lack of a routine, publicly available, genetic transformation system for loblolly pine remains a serious impediment to functional genomics research progress, as well as the ability of individual researchers to attract grant funding. However, grant funds from federal, competitive agencies to develop transformation technology are not likely. As with somatic embryogenesis, substantial advances have recently been made in the private sector. Closer cooperation among public researchers studying or using transformation and greater collaboration with private firms with transformation expertise would benefit the field.

Information access.
Research on loblolly pine genomics would be facilitated by increased and more efficient access to information. Comprehensive relational databases would allow researchers working on the same genotypes to exchange information and make the availability of plant material more accessible to potential researchers. Because there are various ownerships of the ermplasm, relational databases would need to protect confidential information, while making public information freely available. Public researchers would benefit from having more availability to relevant plant material and private owners of the ermplasm would benefit from increased information about their plant material.

Essentials of Five-Year Plan

Genetic stocks.
A curator should be nominated to maintain and distribute DNA stocks. Initial talks have focused on UC-Davis, but funding from a public agency needs to be obtained.

Germplasm

- The 96-clone WeyCo population should be reproduced at a 2nd and/or 3rd site(s). Access and permission to transfer the material should be obtained from Weyerhaeuser Co.

- The 500-clone association population should be propagated and planted in several locations across the South. Likely locations include the USFS in Saucier, MS and university forests in North Carolina, Florida and Texas.

- A reference genotype for physical mapping and future sequencing should be obtained immediately.

- Future, additional mapping or other scientific study populations can be produced to meet future scientific objectives. Breeding, seed and pedigree information should be arranged through the tree improvement cooperatives

and their members. Clonal propagation, if desired, can be arranged given adequate time and resources.

- Collaborations with private companies for somatic embryogenesis, cryopreservation and high-throughput transformation should be established and promoted.

Information

Bioinformatics databases should be modified to include phenotypic data of sample trees. There is a wealth of data on many trees from the tree improvement breeding and testing activities. These data will prove extremely useful to genomic scientists if appropriate access controls can be arranged.

Facilitate the use of Material Transfer Agreements by drafting a general format appropriate for pine materials that clarifies appropriate conditions for publication of genotype identification, commercialization of genotypes, genes, or gene products, and other intellectual property issues among public researchers and private owners of genetic material.

Private Sector Contribution

Working Group Members: Maud Hinchee, Chair - ArborGen, LLC; Phil Cannon – Boise Cascade; Jim Rakestraw – International Paper; Bob Purnell – International Paper; Suzanne Bertrand – ArrayXpress; Fred Raley – Texas Forest Service

General Comments

Industry is supportive of loblolly genomics research, but it is not likely to directly fund a large, long term research project. The lumber, pulp and paper industries have gone through significant mergers and consolidation, and have trimmed funding for genetic improvement. Support for their internal research programs is directed to optimizing and applying technology that will have near term positive impact on the productivity of their forests. The trend is to outsource more of their research. This provides a greater opportunity for university researchers to generate proposals that will be attractive to industrial research organizations. Industry is more likely to fund research that can be highly leveraged against grant supported activities.

Tree improvement at many pulp, paper and timber companies in the United States is focused on breeding to develop elite families, which in turn can be used to generate planting stock through seeds or rooted cuttings, or a selection population from which clones may be developed. Forestry companies have participated in pre-competitive and collaborative tree breeding and rooted cutting cooperatives. These groups have developed technology and *germplasm* which individual member companies have been able to use or amend within their own research program.

Significant advances and acceleration of the genetic improvement of Loblolly pine will require the application of new breeding and selection tools that are the products of genomics - based research. The development of such tools is costly, but the potential to accelerate improvement in genetic gain in a single generation is highly valuable.

The industry is very positive about the potential benefits of utilizing genomics-derived breeding and selection tools in their tree improvement programs. Therefore, industry would like to see a community-wide loblolly genomics project developed, that would be able to obtain a high level of funding from government agencies (e.g. Forest Service and DOE - Agenda 2020).

The following comments are based on the overarching recommendation that the Loblolly Pine Genomics Project develop a research proposal that would be directed towards obtaining significant government funding.

Recommendations:

The near and long term benefits of Loblolly pine genomics research needs to be appreciated by the highest levels of management in forestry companies. Currently, genomics research is considered long term and costly, with unclear benefits to operational tree improvement programs. To address this, the private sector working group recommends:

1. Development of a strategic and tactical plan to explain in layman's terms how genomics research can directly impact a company's bottom line. Components of such a plan might be:

a) Preparation of an article on genomics in forestry for "Key Decison Makers" (CTOs, VPs, etc) published in an industry trade journal that addresses new technology

b) A workshop tailored to highlight the technology and its relevance to current tree improvement and clonal development programs (how it affects the bottom line), provided for upper management of forest products companies.

2. Investigators seek supplemental industrial support for grant-supported projects to insure private sector dollars are highly leveraged. (Industry is simply not willing or able to provide financial contributions to support very large and costly proposals for pine genomics. Industry, however, is interested in contributing financial support to initiatives that are funded through research grants from the government.

3. Creation of a task force to develop a proposal to Agenda 2020, to include industry and academia, presenting a compelling proposal for a large-scale loblolly pine genomics project. Specific recommendations for such a proposal would be to:

a) seek long-term funding through agenda 2020,

b) work with AFPA FSTC to gain buy-in and support across and within member companies,

c) go after "big money" and involve a comprehensive collaboration between all loblolly pine genetics researchers,

d) make it clear that loblolly researchers know how to get the most out of the funding by taking scientifically smart short-cuts, and by building upon existing genomics capacity developed for humans, etc.,

e) describe the project by it's expected outcomes, not the genomics process itself (example:. Accelerating Genetic Gain through Development of Advanced Genetic Tools), and

f) highlight how expected outputs and outcomes could directly be applied to genetic improvement of loblolly pine with examples of proof-of-concept.

4. The above noted funding initiative should be guided to support research that is considered pre-competitive and broadly applicable across the industry. Complete industry alignment is needed on the scope and deliverables anticipated in research proposals funded via the Initiative. The relevant universities (to include tech transfer and administration) need to be in alignment on, and supportive of, the initiative and its intellectual property requirements. Industry needs unfettered access to the technology developed, especially using their germplasm. (An example of a concern was the patent application made on the use of FP detection of a Cad null SNP. It is unclear if industry will be able to use this tool). Industry

would also be concerned if the universities were able to provide the use of information, trees Ids and pedigrees, etc going to non-participating 3rd parties. Industry would provide germplasm to an open-ended research program if industrial rights were protected.

5. Generate an executive summary of the proposal to be sent to industrial partners for review and buy-in. This document might include

a) economic justification (i.e. acceleration of genetic gain by 20 years creates X value by improved furnish to the mills),
b) proof of concept examples of how genomics-derived molecular tools and information can be applied to tree improvement within the next 5-10 years (ie use of SNPs in breeding and clonal selection),
c) description of how the breeding cooperatives and company breeding programs can utilize the tools,
d) value of improved trait description and understanding through the linkage of functional genomics to breeding programs.

6. Industrial members of the breeding cooperatives should consider and address issues associated with germplasm donations for genomic based research that is considered pre-competitive and non-exclusive. The issues that industry currently anticipates are:

a) the breeding cooperatives have differing bylaws concerning ownership and intellectual property rights, and this could potentially prevent some loblolly genotypes being utilized in the research program,
b) rights to future inventions derived from the donated genotypes also needs to be addressed.

7. Universities conducting research on loblolly genomics that is considered pre-competitive by the industry (and considered in the scope of the Loblolly Genomics Research Funding Initiative) should align their granting of intellectual property rights to meet the pre-competitive objectives of the funding program. It is important that all forest companies contributing to the initiative are not blocked or financially constrained from using the technology. The industry preference is that intellectual property generated under this funding program at universities be offered non-exclusively.

8. Industry consider providing a financial contribution to the funding program that might be based on the number of seedlings that they plant, or some other mechanism such that the industry is automatically providing money based on the importance of improved seedlings to their business.

9. pre-competitive research might include:

a) development of SNPs through association genomics to identify alleles associated with economic phenotypes that can be applied in pine breeding and clonal selection programs,
b) completion of EST libraries for all expressed genes,

c) development of a full-transcriptome pine microarray chip,
d) completion of cytogenetic and physical maps locating molecular markers, SNPs, and ESTs on pine chromosomes,
e) BAC libraries and sequencing.

Potential types of support from Industry

Industrial partners might be able to provide support to the pine genomics community in the following ways.

1. By making available a reference loblolly pine genotype for the BAC sequencing project (noted in the gene discovery section of this document) that fits the required criteria.
2. By making available other reference genotypes for other research activities. Such genotypes could represent loblolly clones that are transformable, cryopreserved, and capable of large-scale reproduction from somatic embryos.
3. By provision of mapping populations and reference genotypes.
4. By lobbying for funding through Agenda 2020.
5. By assisting with the establishment of research priorities.
6. By assisting the tree improvement cooperatives in aligning their intellectual property rights for germplasm with the needs of the aforementioned Initiative.
7. By providing transformation and SE Services to support research proposals funded by the Initiative
8. By providing matching funds to Government funding for the Initiative.
9. By providing loblolly pine EST sequences not currently in the public domain.

Appendix 1
Loblolly Pine Genomics Support[2]

Year	Title	PI	Institution	Source	Award (K $)
1988	Construction of a saturated linkage map for loblolly pine	Neale	USFS, UC Davis	USDA	250
1991	Genetic mapping of genes for wood specific gravity in loblolly pine	Neale	USFS, UC Davis	USFS	225
1992	Genetic mapping of complex traits in loblolly pine	Neale	USFS, UC Davis	WeyCo	10
1993	Codominant PCR-based markers for pines and other conifers	Harry / Neale	USFS, UC Davis	USDA	200
1993	Molecular marker and quantitative trait mapping in loblolly pine	Neale / Wheeler	USFS, UC Davis	USDA	250
1993	Genetic mapping of complex traits in loblolly pine	Neale	USFS, UC Davis	WeyCo	30
1994	Mapping quantitative trait loci coding for wood specific gravity in loblolly pine	Neale	USFS, UC Davis	WeyCo	30
1996	Development and validation of MAS methods for wood property traits in loblolly pine and hybrid poplar	Tuskan, Neale, Bradshaw et	DOE	Agenda 2020	750
1996	Comparative mapping in the genus Pinus	Neale	USFS, UC Davis	USDA	250
1996	Fine structure mapping and verification of wood quality QTLs in loblolly pine	Neale	USFS, UC Davis	USDA	204
1997	Quinate, focal point of metabolism leading to lignin	Jensen / Bonner	U. Florida	?	109
1997	Wood properties of pine with genetically modified lignin	Sederoff, Mackay	NCSU	?	111
1998	A molecular analysis of carbohydrate regulation in loblolly pine	Ellsworth	Brookhaven	DOE	145
1998	Search for major genes using progeny test data to accelerate the development of genetically superior loblolly pines	Li et al.	NCSU	DOE	?
1998	Pine gene discovery project	Whetten, Sederoff et al	NCSU	DOE	?
1999	Quinate, focal point of metabolism leading to lignin	Jensen / Bonner	U. Florida	?	234
1999	Genomics of wood formation in loblolly pine	Sederoff/ Neale/ O'Malley	NCSU etc	NSF	4450
2000	QTL and candidate genes for growth traits in loblolly pine	Williams	Texas A&M	DOE	?
2001	Allele discovery for genes controlling economic traits in loblolly pines	Neale, Davis, Dean, Covert	UC Davis, UF, UGA	IFAFS	1,400
2001	Wood and fiber quality of juvenile pine: characterization and utilization	Chang, Kadla, Sederoff, O'Malley	NCSU	?	3,000

[2] Much of this information courtesy of Glenn Howe, Oregon State University

2002	Genomics of loblolly pine embryogenesis	Cairney et al	IPST	NSF	1,380
2002	Transcriptome responses to environmental conditions in loblolly pine roots	Dean et al.	UGA	NSF	1,651
2004	Accelerating pine genomics through development and utilization of molecular and cytogenetic resources	Peterson, Faridi, Nelson	MSU, SIFG	NSF	1,600
2000	Host:Pathogen Signaling in Southern Pine Pathosystems	Davis, Morse	UFL	SRS--SIFG	100
2002	Development of a Loblolly Pine FISH-based Karyotype	Faridi, Stelly	Texas A&M	SRS--SIFG	33
2002	Performance and Value of CAD-Deficient Pine	Mullin, et al.	NCSU, SIFG	DOE--Agenda 2020	1,125
2003	Development and Comparison of FISH-based Karyotypes of Loblolly, Shortleaf, Longleaf and Slash Pines	Faridi	Texas A&M	SRS--SIFG	58
2003	Development of SSR markers from EST sequences in loblolly pine	Nelson, Echt, Peterson.	SIFG, MSU	SRS--Agenda 2020	83
2004	Differential gene expression in loblolly pine (*Pinus taeda* L.) challenged by the the fusiform rust fungus, *Cronartium quercuum* f.sp. *fusiforme*	van Zyl, Amerson	NCSU	SRS--SIFG	44
2004	Genetic Screening for Resin Traits Linked to SPB Resistance	Davis, Morse, Nelson, Strom	UFL, SIFG	SRS--SPB	189

Appendix 2
2004 LPGP Workshop Attendees

Participant	Title	Institution
Amerson, Henry	Assoc.Professor & Director, Pine Rust Consortium	North Carolina State University
Bertrand, Suzanne	Director, Project Management	ArrayExpress
Byram, Tom	Professor / Director Western Gulf Tree Improve. Coop	Texas A&M
Cairney, John	Assoc. Professor	Georgia Tech / IPST
Cannon, Phil	Research Manager	Boise Cascade
Cordonnier-Pratt, Marie-Michele	Research Professor	University of Georgia
Covert, Sarah	Assoc. Professor	University of Georgia
Davis, John	Assoc. Professor	University of Florida
Dean, Jeffrey	Assoc. Professor	University of Georgia
Egertsdotter, Ulrika	Assist. Professor	Virginia Tech
Goldfarb, Barry	Professor & Dept. Head	North Carolina State University
Hinchee, Maud	CTO	ArborGen
Kirst, Matias	Assist. Professor	University of Florida
Li, Bailian	Research Professor	North Carolina State University
McKeand, Steve	Professor	North Carolina State University
Miller, Larry	WGTIC	Texas Forest Service
Mullin, Tim	Research Professor & Director NCSU Tree Improve. Coop	North Carolina State University
Nairn, Joe	Assist. Professor	University of Georgia
Neale, David	Research Geneticist& Adj. Professor	USDA Forest Service, U.C. Davis
Nelson, Dana`	Research Geneticist& Project Leader	USDA Forest Service, Saucier, MS
Peter, Gary	Assoc. Professor	University of Florida
Peterson, Dan	Assist. Professor	Mississippi State University
Pratt, Lee	Research Professor	University of Georgia
Purnell, Bob	Research Manager	International Paper
Raley, Fred	WGTIC	Texas Forest Service
Rakestraw, Jim	Research Manager	International Paper
Rottman, Will	Senior Scientist	ArborGen
Sederoff, Ron	Professor	North Carolina State University
Sheppard,		
Van Zyl, Len	Res. Assist Prof. & CSO	North Carolina State University/Array Express
Weber, Joe	Research Manager	Boise Cascade
Wegrzyn, Jill	Database Admin	U.C. Davis
Wheeler, Nicholas	Consultant	Molecular Tree Breeding Services

Appendix 3: Genomics Research Priorities

Topic	Priority Need	Specific Tasks	By Date
Gene Discovery and Sequencing	Obtain 3' and 5' sequences from an additional 500K ESTs from a broad array of tissues (e.g. 100 libraries 5000 clones deep. [Estimated cost ca. $3-4 million])	Secure long-term support Identify appropriate tissues, extract mRNA. Obtain sequence. Coordinate bioinformatics support	2007
	Complete a collection of full-length cDNAs representing 90% of all Superscripts (ca. 35K sequences [ca. $3-4million])	(most likely transposon insertion sequencing)	2008
	Increased sequencing of high(5000), moderate and low-Cot libraries (50,000 each) (ca. $250K). Target complete sequencing in 10 years.		2007
	Limited genome sequencing (meaning sequencing a substantial but limited number of targeted BACs), to contribute to the question of local organization for ESTs in the genome (ca. $8K/BAC)		2012
Functional Genomics	Improvements in detection, measurement and analysis of phenotypic data for heritable traits, collected at all scales (i.e. molecular, cellular, tissue, plant, stand)	Directed efforts to develop specific, high-throughput phenotyping capabilities.	Continuous
	Improved computational approaches for integrating genotypic and phenotypic data.	Post-doctoral support for statistical geneticists	2005 – 2008
	Increased understanding of pine gene regulation, including promoter sequences and the identification of genes involved in signal transduction.	Target upstream sequencing efforts	2009
	Access to and financial support for tractable pine transformation and association genetics reference populations to conduct forward and reverse genetics	Creation of public association genetic reference populations Development of service capability in pine transformation. Secure long-term funding support for all aspects	2007

Topic	Priority Need	Specific Tasks	By Date
Genetic Stocks & Germplasm	Establishment of a genetic stock center for pine	Secure long term support for a stock center Identify location and curator	2005
	Establish permanent clonal archive for one or both of the Weyerhaeuser reference mapping populations	Obtain access and material transfer agreements from WeyCo. Graft populations to one or more new plantations	2005
	Create clonally replicated association genetics reference populations and locate on multiple field test sites	Clonally propagate > 500 unrelated individuals from NCSU and >1000 clones from UF. Establish field sites	2006
	Identify a reference genotype for physical mapping	Same	2004
	Develop collaborations with private companies for somatic embryogenesis, cryopreservation and high-throughput transformation services	Seek partners and industrial support for the community	2005
Maps and Markers	Place 300 to 1000 SSR on loblolly pine reference maps; 100 genomic SSR, 200 to 900 gene based SSR	Map 100 genomic SSR currently available Identify 200 gene based SSR and map based on current data Identify additional 700 gene based SSR from BAC library	2004 2006 2010
	Identify and optimize 12-16 SSr markers for fingerprinting and paternity analysis applications		2005
	Identify and map 30,000 SNPs representing a loblolly pine unigene set, to reference population		2010
	The completion of a physical map of loblolly pine and integrate it with the genetic map.	Create 10x coverage BAC library BAC fingerprinting and contig assembly Placement of ESTs/QTL/SSRs/SNPs on BAC grids; FISH	2014

Topic	Priority Need	Specific Tasks	By Date
Bioinformatics	Develop data mining tools targeting smooth integration of different data types such as sequence data, expression data (whether from microarray or digital expression), map data, genotyping and phenotyping. Interfaces should provide easy visual interpretation of data types integrated and complex querying capabilities	Specialized clustering of ESTs and other expression tags. Multiple annotations (GO, IProClass, COG, Pfam, KEGG, etc…). Mapping to genome/BAC/cot sequences	Continuous
	Increased/continued integration across datasets More accurate clustering of all datasets; integration of data derived from EST, SAGE, MPSS and other expression tags as well as phenotypic data sets.		
	Deeper data mining of sequences for domains to improve clustering and use in discovery of gene expression control		
Private Sector Input	Development of a strategic and tactical plan to explain in layman's terms how genomics research can directly impact a company's bottom line	Write and publish a genomics lay article Provide workshop/field tours for corporate managers	2005
	Creation of a task force to develop a proposal to Agenda 2020, to include industry and academia, presenting a compelling proposal for a large-scale loblolly pine genomics project		2005
	Industrial members of the breeding cooperatives should consider and address issues associated with germplasm donations for genomic based research that is considered pre-competitive and non-exclusive.		On-going
	Universities conducting research on loblolly genomics that is considered pre-competitive by the industry should align their granting of intellectual property rights to meet the pre-competitive objectives of the funding program		On-going