

# Scientific Advisory Board Report

Jack Dekkers and Chuck Langley

*[Report submitted at Annual CTGN Project Meeting, June 22-23, 2009. CTGN Project responses to recommendations follow below in italic Times New Roman font.]*

Substantial progress has been made on each of the research objectives. Coordination on aspects of the project (sample collection, genotyping, data base development) are good, but further planning and collaboration will be needed to ensure the project will deliver on its goals of validating SNPs and providing stakeholders with tools for MAS/MAB.

## **Objective 1. Validate SNP associations discovered in prior grants in operational tree improvement populations.**

Data base system for entry, storage, and retrieval of phenotype and genotype data has been developed and is in place.

***Recommendation: to facilitate future use of the data base and joint analysis of populations, attention needs to be given to:***

- ***Further standardization of phenotypes***
- ***Entry of raw phenotypic data***
- ***Pedigrees***

*Every effort will be made to standardize phenotypes, but there are constraints imposed by the organizations (tree improvement cooperatives) doing much of the phenotyping. These cooperatives have historical and practical reasons for the measurement protocols used for specific phenotypes, resulting in protocols that may not be consistent from region to region. Protocols for new traits (e.g., wood properties) will be standardized to the extent possible (although species differ). Furthermore, raw phenotypic data and pedigree information will be added to the database.*

A comprehensive system for receipt of field samples, DNA extraction, tracking and genotyping has been developed and is in place.

Genotyping platforms are in place and operational – 1536 Golden Gate panel  
- 7,600 iSelect panel

***Recommendation: Incorporate plans for population-specific gene calling – using raw bead data and Illumina Genome Studio software.***

*Illumina generously gave a free copy of the Bead Studio software for SNP calling to each of the collaborators. This will allow each collaborator to do population-specific SNP calling.*

**Recommendation: Develop plans to facilitate future work with further scaling-up to larger panels and for subsequent re-genotyping using larger panels.**

*Yes, plans are being made. DNA samples for all trees will be saved at the CTGN stock center so they can be genotyped with higher density SNP chips when and if additional funding becomes available.*

Individual collaborator projects for validation and discovery have made good progress on developing populations and collecting phenotypes and samples for genotyping. The different validation populations and data sets are at different stages of development. Genotyping has started for some populations. The complete path that will be taken in each population to achieve the goal of validating associations detected in earlier studies is not clear.

It is unclear to what extent the emphasis in each population/project is on using LD versus within-family linkage – this has large impact on the individuals that should be genotyped and on how the results should be implemented in breeding.

**Recommendation: Evaluate the impact of alternate measures of phenotype (raw phenotypes/BLUPs) in association analyses.**

*Yes, we discussed this at the annual meeting, and plan to investigate this question (e.g., raw phenotypes, BLUPs, deregressed breeding values).*

**Recommendation: Develop a detailed time-line for each population for sample submission, genotyping, and analysis.**

*Yes, this was essentially done but not reported at the meeting. We will update these plans if needed.*

**Recommendation: Use genotyping strategies (individuals to genotype) that would enable subsequent use of genotypes for future projects (e.g. high-density – low-density genotyping strategies, as in NAMS design).**

*To the extent possible, we are genotyping and phenotyping individuals that can be used in future projects. That is, individuals that will be maintained in archives or used for future breeding.*

**Recommendation: Use genotyping data to further evaluate the extent of LD in breeding populations, how LD has changed during the cycles, and the consistency of LD across populations. Look for evidence of geographic differentiation in SNPs associated with specific traits.**

*Yes, this is a basic population genetic question that the UC Davis group will look at as the SNP genotypes become available for multiple generations and populations.*

**Recommendation:** *Develop detailed strategies and plans for data analysis, including separate analysis of each population, joint analysis of multiple populations, who will develop and conduct which analyses on which populations, etc. Consider sharing/exchange of populations and software. For associations analyses, consider Genomic Selection-type of analyses of all SNPs jointly.*

*We will develop this strategy over the next year, but some of these decisions will depend on the results of our simulations (see below). Yes, genomic selection holds a great deal of promise for tree breeding programs. Joint analyses will be conducted after each collaborator has had some success analyzing their own data, toward the end of the project.*

**Recommendation:** *Consider evaluation of SNP associations with more detailed expression, morphological and physiological phenotypes.*

*Association genetics with molecular phenotypes (transcriptome and metabolome) is being done under the NSF-funded ADEPT2 project. There is nothing in the CTGN budget for this very expensive molecular phenotyping. Furthermore, operational breeding populations may be suboptimal for initial analyses of new expensive phenotypes of uncertain value to tree breeders.*

## **Objective 2. MAS/MAB applications**

Several simulation programs have been developed to assess various partially complementary and partially overlapping aspects of the use of markers for association analysis, QTL mapping, and the application of markers in MAS/MAB. Different simulation software may be needed depending on the targeted application/audience.

*All modeling and simulation work currently undertaken in this project, with the exception of economic risk analysis software (Simetar) was proposed, or subsequently initiated under Objective 1 goals. However, given the commonality of goals and their relationships to potential applications of MAS/MAB, it is reasonable and perhaps desirable to officially move them to this objective.*

**Recommendation:** *Coordinate development and use of simulation programs across groups and integrate software where possible. Also integrate software with methods of analysis.*

*We are exploring possibilities for coordinating the development and integration of software between groups. We undertook multiple simulation activities to address alternative objectives and approaches of interest. Therefore, complete integration of the software development activities is not necessarily desirable. However, we do plan to conduct similar simulations using alternative approaches to test the simulation programs, assumptions, and robustness of the approaches. The integration of simulation programs and analytical routines is highly desirable, and we are now moving into that phase of these activities. This integration will clearly expedite the use of simulations to answer specific questions, such as those outlined in our response to the next recommendation.*

**Recommendation: Use simulation software to identify optimal population structures and SNP densities required for LD and/or linkage analysis – use LD that exists in operational tree breeding populations.**

*This is a key objective for the CTGN project. Given the nature of tree breeding populations and current status and SNP markers in conifers, we expect that combined LD and linkage analysis will become very important. In the long-run, genomic selection holds a lot of promise for tree breeders. We have already run simulations to address optimal population structures and SNP densities, and this will receive more attention in the coming year. Experimental data on the extent of LD in operational breeding populations are not yet available, but existing data from natural populations should be valuable, particularly for the first-generation breeding materials. Furthermore, we soon expect to obtain data that will allow us to incorporate realistic levels of LD into our simulations.*

**Recommendation: Use simulation software to identify importance of LD vs linkage in MAS/MAB in operational tree breeding populations.**

*This is also a key CTGN objective. We expect that both linkage within pedigrees and LD across the breeding population will be useful tools for applying markers in breeding programs, although their relative importance may vary for different traits and different applications. Our objective is to implement computationally tractable approaches for using both population-wide LD and linkage within specific pedigrees to identify marker genotypes that are useful for predicting breeding values.*

**Recommendation: Focus on development of genetic evaluation software that will allow stakeholders to incorporate marker information in BV estimation.**

*Staff in our cooperative breeding programs implement the analyses needed to predict breeding values. Therefore, genetic evaluation software that incorporates marker information to estimate breeding values will be mostly used by cooperative staff, rather than by other stakeholders. Furthermore, cooperative staff also serve as consultants to forest industry, and can help them analyze their internal data, if such help is needed. Currently, we expect to use extensions of current methods to incorporate markers, including mixed model approaches using programs such as ASReml, or Bayesian approaches using programs such as those being developed at the University of Florida (Huber and colleagues).*

Software (Simetar) has been developed as a tool for education and to allow rapid evaluation of alternate breeding programs. It is not clear to what extent these incorporate multiple-trait breeding goals (with economic values), multiple-trait selection indexes, or use of selection index methods to incorporate marker information and to model BLUP EBV and to evaluate impacts on inbreeding. Incorporating these concepts would also capitalize on ‘teaching opportunities’ to further promote use these basic breeding methods.

**Recommendation: Expand Simetar to incorporate additional concepts and aspects of comprehensive multi-trait breeding programs.**

*Simetar is a commercial product designed to work with general agricultural economics issues. It is not specifically designed for crop improvement activities, but it is highly flexible and programmable. We can model any multi-trait or BLUP-derived EBV through input variables defined as gain. This can be done using a gain distribution that would be treated stochastically in the model. In the end, this model is intended to provide insight on the economic feasibility of MAS to potential technology users. We believe this can be done by keeping the gain input variables relatively simple.*

*Our next objective is to integrate analytical routines into the simulation package and produce a set of estimates of genetic gain for a single trait under a set of plausible scenarios of tree breeding applications of genomic information. Once this objective is achieved for a single trait, the simulation program will be revised to simulate multiple phenotypes with varying degrees of genetic correlations, pleiotropic or epistatic effects, and analogous estimates of genetic gain will be produced. These can then be used as input for multi-trait economic analyses in the Simetar model.*

### **Objective 3 - Bioinformatics**

Through Treegenes and Dendromes, a comprehensive web-based system has been developed with genomic resources for conifers, which serves an important role for the US and worldwide conifer genetics and genomics community.

**Recommendation:** *It will be important for the conifer genetics and genomics community to provide feedback on utility and suggestions for improvement of the resource.*

*Yes, this has been ongoing for nearly 20 years since the Dendrome project was started.*

**Recommendation:** *A stable source of longer-term funding to ensure maintenance and further development of the resource must be found.*

*A chronic problem still awaiting a solution. CSREES and NSF have declined for many years to support Dendrome as it does support other plant genome database projects. The US Forest Service would be the most logical agency to support Dendrome but they too have declined over the years.*

### **Objective 4 – Conifer Genetic Stock Center -**

Living and DNA-based resources have been established to preserve genetic material from previous research projects. It is not clear how these resources will be used, other than that the genetic material will be maintained through these efforts.

**Recommendation:** **Need efforts to advertise existence of these resources to the research and stakeholder communities, including opportunities to collect and evaluate additional phenotypes.**

*The availability of the accumulated resources will be announced initially through CTGN research publications that report characterization and usage of the materials. Once a distribution system for the resources is in place, a broader advertising to stakeholder communities will be made. The ability to expand the acquisition, evaluation, and maintenance of individuals that are not used in the CTGN project will have to be opportunistic as funding resources permit.*

**Recommendation:** **Expand efforts to other species**

*We recognize that it will be efficient and economical to extend this to other species once a system for genetic resources maintenance and distribution is in place, and as staff and funding permit.*

**Recommendations:** **Given the long generation intervals and the rapidly changing technologies (higher-density SNP panels – complete sequencing), determine what populations would be most useful to establish now for future R&D**

*We agree with this proactive approach, and want to ensure that the continued maintenance and distribution of materials is justified. Therefore, we will seek periodic advice from users and curators of other collections about such issues.*

General consideration: the group should further discuss the future and impact of tree breeding in the US in the context of the changing global industry and market shifts and emerging science, and evaluate the implications for research.