Aim 3 – Genetics
PINEMAP Year 2 Final Progress Report
March 2013

This is the final Aim progress report for year 2 (March 2012-February 2013). Please update the outputs list and provide brief progress updates on milestones and work plan tasks as applicable. The information provided in these reports is used to track Aim-level outputs and outcomes over the course of the project and to fulfill NIFA reporting requirements via the annual continuation proposals and CRIS progress report.

Please return the updated report to Jessica Ireland no later than **March 22**.

Outputs

List ***Products*** developed/completed January 2012-current (including published, in press, or in review peer-reviewed publications; other written materials such as white papers, research summaries, fact sheets, or popular press articles; audio or video products; etc.).

A list of documented Aim 3 year 2 publications is provided below; please update as necessary and highlight in yellow any publications added to the list in the March 2013 Progress Report.

Peer reviewed publications (January 2012-current)

Aspinwall, M.J., S.E. McKeand, and J.S. King. 2012. Carbon sequestration from 40 years of planting genetically improved loblolly pine across the Southeast United States. *Forest Science* 58(5): 446-456.

Alberto, F., S. Aitken, R. Alia, S.C. González-Martínez, H. Hanninen, A. Kremer, F. Lefèvre, T. Lenormand, S. Yeaman, R.W. Whetten, and O. Savolainen. 2013. Evolutionary response to climate change - evidence from tree populations. Global Change Biology, in press.

Holliday, J.A., T.L. Wang, S.N. Aitken. 2012. Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. *G3: Genes, Genomes, Genetics* 9:1085-1093.

Zhou, L., Holliday, J.A. 2012. Targeted enrichment of the black cottonwood gene space using sequence capture. *BMC Genomics* 13:703.

Resende, M.F.R.,Jr., Muñoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F., Kirst, M., 2012. Accuracy of Genomic Selection Methods in a Standard Dataset of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190: 1503-10.

Resende, M.F.R., Jr., Muñoz, P., Acosta, J.J., Peter, G.F., Davis, J.M., Grattapaglia, D., Resende, M.D.V., Kirst, M. 2012. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193: 617-624.

Events/Activities (January 2012-current)

→ Provide a bulleted list of presentations (oral and poster) given at meetings or conferences. The format for citing presentations is as follows:
Presenter(s)/Author(s). Date. Name/title of meeting/conference, location.
*Indicate poster presentations by placing [poster] at end of citation.
Highlight in yellow presentations added to the list in the March 2013 Progress Report.

Bawa R. and J. Holliday. July 6-12, 2012. Signatures of natural selection and local adaptation in *Populus trichocarpa* and *Populus deltoides* along latitudinal clines. Joint Congress on Evolutionary Biology. Ottawa, Ontario, Canada.

Chhatre, V. May 22-23, 2012. Evolutionary and population genetics of the Lost Pines: Lessons from history. Western Gulf Forest Tree Improvement Contact Meeting, Bastrop, TX

Chhatre, V., T. Byram, D.B. Neale, J.L. Wegrzyn, and K.V. Krutovsky, January 14-18, 2012. Genome Wide Analysis of Genetic Associations with Environmental Variables in East Texas Loblolly Pine (Pinus taeda L.). Plant & Animal Genome XX . The International Conference on the Status of Plant and Animal Genome Research, San Diego, CA, USA

Holliday JA, Zhou L. "Toward whole-genome association mapping and landscape genomics in a model tree." Botany 2012 – The Next Generation (Annual Meeting of the Botanical Society of America). Columbus, OH, USA. July 7-11, 2012.

Krutovsky, K.V., January 14-18, 2012. Forest Genomics for Mitigating Climate Change and Breeding Resilient Trees. Plant & Animal Genome XX . The International Conference on the Status of Plant and Animal Genome Research, San Diego, CA, USA

Farjat, A., May 10, 2012. Climate variables and height growth in the Plantation Selection Seed Source Study field trials. NC State University Cooperative Tree Improvement Program Advisory Board Meeting, Raleigh, NC.

Holliday, J.A. and L. Zhou. July 7-11, 2012. Toward whole-genome association mapping and landscape genomics in trees. Botany 2012 – The Next Generation (Annual Meeting of the Botanical Society of America), Columbus, OH, USA.

Holliday JA, Zhou L. October 31, 2012. "Toward genome-wide association mapping in a model tree." Texas Tech University Department of Biology seminar. Lubbock, TX.

Holliday JA, Zhou L**.** October 15, 2012. "Toward genome-wide association mapping in a model tree." University of Maryland Department of Plant Science seminar. College Park, MD.

Isik, F. Oct 25, 2012. Universal Response Function development from the Plantation Selection Seed Source Study field trial data. NC State University Cooperative Tree Improvement Program Contact Meeting, Asheville, NC

Krutovsky, K. May 22-23, 2012. How genomics can help restore the Lost Pines. Western Gulf Forest Tree Improvement Contact Meeting, Bastrop, TX

Krutovsky, K.V., Oct 3 – 5, 2012. Forest Genomics for Mitigating Climate Change and Breeding Resilient Trees. Keynote presentation at the AdapCAR and IUFRO (WP 2.02.00) meeting Genetic aspects of adaptation and Mitigation: forest health, wood quality and biomass production, Riga, Latvia.

Krutovsky, K.V., Oct 16 – 17, 2012. Forest ecogenomics and genomic selection for mitigating climate change and breeding resilient trees. Final Conference and Workshops of Noveltree Project. Tree Breeding, Genomics and Evolutionary Biology: New synergies to tackle the impact of climate change in the 21 st century, Helsinki, Finland.

Whetten, R. May 10, 2012. Preliminary results from Genotyping-By-Sequencing experiments in loblolly pine. NC State University Cooperative Tree Improvement Program Advisory Board Meeting, Raleigh, NC.

Whetten, R. Oct 24, 2012. Rust resistance screening and new molecular marker methods for pine. NC State University Cooperative Tree Improvement Program Contact Meeting, Asheville, NC

Whetten, R. Jan 12, 2013. The Pine Reference Genome Sequence and Applied Tree Breeding Programs. Pine Reference Genome Workshop at the 21st International Plant and Animal Genome Conference, San Diego CA USA.

Koralewski, T.E, T.D. Byram, E.M. Raley. Feb. 4 – 7, 2013. A climate change response function for loblolly pine (*Pinus taeda*) from the Western Gulf region of the United States. Breeding for Value in a Changing World, Jacksonville, FL, USA

Lu, M., Loopstra, C., and Krutovsky, K. 2013. Preliminary SNP Discovery in Loblolly Pine (*Pinus taeda* L.) Exome. P0459. Plant and Animal Genome XXI. January 12-16, 2013. San Diego, CA.

Zhou L, Holliday JA. “Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture.” Plant and Animal Genome Conference. San Diego, CA. January 12-16, 2013.

→ Provide a short narrative describing any workshops, courses, and/or trainings conducted.
Highlight in yellow items added to the list in the March 2013 Progress Report.

Fikret Isik of the NC State University Cooperative Tree Improvement Program co-taught a week-long workshop entitled “Genetic Data Analysis: Applications in Plant and Animal Breeding” at the Czech University of Life Sciences in Prague, Czech Republic, June 11 – 15, 2012. The course covered statistical methods for analysis of field trials, including methods appropriate for trials planted at multiple sites to estimate impacts of climate variation on forest productivity. The course was jointly organized by NC State University Forestry Education and Outreach Program and forestry faculty from the Czech University of Life Sciences.

Konstantin Krutovsky of Texas A&M University chaired and co-taught a week-long workshop entitled “ Population genetic and genomic approaches to mitigate global climate change impacts on forest genetic resources and to breed more resilient trees” organized by Büsgen Institute, Department of Forest Genetics and Forest Tree Breeding, University of Göttingen and funded by EvolTree (Evolution of Trees as Drivers of Terrestrial Biodiversity, EU-funded Network of Excellence), September 24-28, 2012, Göttingen, Germany. The workshop taught how the application of population genetics and genomics can help to tackle global climate change challenge and guide forest protection and conservation programs aimed to mitigate its undesirable effects. It also demonstrated how modern genomic selection can be used for breeding more resilient trees that will be productive in harsh and variable environments. The workshop also covered population genetics and genomics tools and methods that can be used to better characterize available forest tree genetic resources. (http://www.evoltree.eu/index.php/component/jevents/icalrepeat.detail/2012/09/24/202/-/- ,http://www.uni-goettingen.de/en/evoltree-summer-school-in-goettingen/361359.html)

→ Provide a short narrative describing experiments or surveys conducted and/or analyzed.

Progress Updates: Milestones and Work Plan Tasks

Provide a short narrative describing progress and accomplishments on the year 2 milestones and work plan tasks listed below.

Progress updates are carried over from the November 2012 Interim Report. Please provide additional progress updates as applicable for each milestone and work plan task under the March 2013 Progress Report heading.

Year 2 Milestones

→ Version 1 of genetic deployment tool with provenance information (October 2012).

November 2012 Interim Report:

1) Texas A&M: Analysis of 73 tests of the WGFTIP Seed Source Study of 43 selected families (divided into two series) has been conducted with focus on using 1) mean minimum temperature of the coldest month (MMIN) and 2) summer heat/moisture index (SHM) as dependent variables and 1) height growth, 2) diameter and 3) volume as independent variables. MMIN corresponds to the most extreme event, and strongly correlates with the other minimum temperature variables investigated

($P < 0.0001$). The trend for MMIN was clearly linear and closely related to latitude. SHM showed a relationship with longitude that could be interpreted as curvilinear or perhaps indicating a threshold between two different distributions occurring at approximately 93.5° W. Highly significant response surface models have been developed; however, they pose further questions regarding the large portion of variation that is not explained by the weather explanatory variables alone. All raw progeny test data from this study has been supplied to University of Florida for the Tier 1 database.

2) NC State University: Growth (height at age 8) and climate data were analyzed for 19 field test sites of the Plantation Selection Seed Source Study using a multiple regression approach that tested a total of 68 climate variables and interaction terms for significance, to create a Uniform Response Function for the NCSU pine breeding populations. After refinement the model included 19 variables and interaction terms, and explained almost 68% of the variation in height growth in the dataset. Cross-validation to test the predictive accuracy of the model found predictive power was low, indicating that the problem is ill-conditioned due to co-linearity in the climate variables. Further refinement of the statistical model will use regularization or dimensional reduction methods to deal with the co-linearity in explanatory variables in an attempt to increase the predictive power of the model. Including additional independent variables related to site quality in the modeling effort is likely to increase the predictive accuracy of the model, provided that suitable data can be obtained for the test sites. The progeny test data from the 19 field sites used in this analysis have been provided to University of Florida for the Tier 1 database.

March 2013 Progress Report:

1) Texas A&M: Analysis of the data has been conducted with a focus on age 15 height and diameter measurements and planted tree volume. Mean minimum temperature of the coldest month (MMIN) and aridity index (AI) were used as independent variables. The modeling efforts provided robust support for current and future seed deployment guidelines at the provenance level. Moreover, with focus primarily upon the climate at the seed source as a common evolutionary background, additional gains could be made from taking advantage of the considerable tree-to-tree genetic variation. Change in climate between provenance and the test site as the sole controlled parameter with other environmental factors assumed constant (e.g. site preparation, fertility, drainage) leaves room for further improvement of future loblolly pine productivity through application of better silvicultural methods, targeted seedling deployment supported by continued breeding and progeny testing and the integration of the two efforts.
2) NC State University: Growth (height at age 8) and climate data were analyzed for 19 field test sites of the Plantation Selection Seed Source Study using multiple regression approaches. Main effects and higher order terms were fit to growth data to develop prediction models under climate change scenarios for the Southeast. A model with 19 predictors explained about 68% of the variation in height growth. Cross-validation showed that the predictive power of models was low, indicating that the problem is ill-conditioned due to co-linearity in the climate variables. In order to find parsimonious models that are stable, different approaches for model

selection and parameter estimation were explored, namely the stepwise method combined with ordinary least squares (OLS), ridge regression, and LASSO regression. The predictive accuracies of these approaches were compared through cross-validation. Although a formal hypothesis test reveled significant differences between OLS and LASSO, and between ridge and LASSO; the three models perform similarly and explained 22% of variation in height. A hypothetical climate scenario was created from historical data, assuming a 5% decrease in precipitation, 2% increase in maximum temperatures, and an increase of 2° C in minimum temperatures. The results suggested that local seed sources perform very well in the Georgia-Florida coastal plains, and their estimated growth decreases as the seed sources are moved to the north. On the other hand, South Carolina seed sources exhibited a more robust performance relative to local seeds, specifically in North Carolina. We observed a rapid decline in performance of Virginia seed sources under the climate change scenarios when they were moved to southern regions. The models developed can be used as quantitative tools to predict the norm of reaction of seed sources under climatic (minimum temperatures) change scenarios. Larger data from older genetic field trials that represent a wider geographic area should be used to further refine the models.

3) University of Florida: Analysis of provenance progeny trials with 15 year height has been conducted with two sets of climate data, PRISM and SECC, using a multiple regression approach. The multicolinearity of the climate data was dealt with by first using LASSO, a Bayesian variable selector, to identify the most significant climate parameters. With the PRISM data, the most significant parameters are site mean winter monthly temperature, site mean coldest monthly temperature, site mean annual precipitation, provenance mean coldest monthly temperature, and provenance longitude. These explain 52% of the variation in height. With the SECC data, the same 5 climate variables were significant and an additional one, site mean annual radiation. These six variables also explained 52% of the variation in height. To test the stability of the models, the height for Atlantic Coastal and Florida provenances was predicted based on climate variables across the SE. Overall, the ACP shows increased growth in the north with the tallest trees in central GA, AL and AR. The predictions in SC are more variable. Somewhat unexpectedly, the Florida provenance shows a similar pattern with better growth north but extends further south into FL. These preliminary results need validating.

→ Determine appropriate genome reduction methods for genotyping by sequencing (October 2012).

November 2012 Interim Report:

1) Texas A&M: Agilent's SureSelect Target Enrichment System was used to select exon regions in loblolly pine. 647,634 baits were designed using 35,550 unigenes and the Agilent's eArray software. The total capture library size was approximately 78MB. Subsequently, genomic DNA of a haploid megagametophyte and a diploid embryo from the same seed from the reference tree 20-1010 were isolated. After fragmenting the total genomic DNA into 150 to 200 bp long fragments, two

sequencing libraries were built, one from each source of genomic DNA, and then hybridized with the capture library. After hybridization, the captured DNA was isolated using magnetic beads and amplified with index primers. The amplified selected sequences were sequenced using a Hiseq 2000 sequencer. After trimming and discarding low quality reads, sequences of 6.4 Gbp and 5.8 Gbp were obtained for the megagametophyte and embryo libraries respectively. These were *de novo* assembled into 184,907 contigs representing 58Mb. Blastn was used to check the *de novo* assembly against the draft 18Gb loblolly pine genome with the result that only 5 contigs failed to give hits. Attempts to map the sequence to the draft genome using CLC are currently underway. After filtering, this method has identified 51,623 SNPs or approximately one SNP for every 1,119 bp. Less than 8% of the initial unigene set is not represented. Foliage samples were collected from 397 individuals represented in the ADEPT II Association mapping population for the next part of the project.

2) NC State University: A published procedure for Genotyping-By-Sequencing using two restriction enzymes to fragment genomic DNA (Poland et al, PLoS One 7: e32253, 2012) was used to prepare DNA samples from two parents and 90 progeny of loblolly pine. Results were good - each of two Hiseq lanes yielded 175 million paired-end 100-nt sequencing reads, and analysis using the Java-based GBS tools in the TASSEL software package yielded from 47,131 to 103,669 candidate marker loci depending on software parameters. These candidate loci have been compared to the draft loblolly pine reference genome assembly produced by the PineRefSeq genome sequencing project to estimate the proportion of single-copy sequences. About 70% of the candidate loci mapped to the draft genome assembly, and of those, about half mapped to a single location, yielding over 16,000 candidate single-copy genetic markers. A linkage map with 12 linkage groups, containing 529 markers distributed over a total map distance of 1281 cM, was constructed to confirm the Mendelian segregation of a sample of these candidate markers. An additional cluster of markers is found in a small linkage group that is not yet connected to one of the 12; addition of more marker loci to the mapping dataset is expected to allow this fragment to converge with one of the 12 larger groups.
3) Virginia Tech: Genome complexity reduction by the ‘RAD-seq’ approach was tested for 10 haploid megagametophyte and two diploid needle samples. An initial test was completed in an attempt to optimize the stoichiometry between adapter sequences and available genomic DNA fragments from the restriction digest. Specifically, the P1 adapter was titrated between 0.01 and 1x relative to the protocol modified from Miller et al (2007) (https://www.wiki.ed.ac.uk/display/RADSequencing/Home). Once we were satisfied based on the bioanalyzer results of these titrations, we completed library preparation of the indexed samples were sequenced on a single lane of an Illumina MiSeq instrument (2x100 PE format). Results from this sequencing run were not encouraging – much of the data generated was from adapter sequences, indicating an excess of concatenated adapters with no genomic DNA inserts. Discussions with a variety of colleagues revealed that this is a common issue with RAD-seq, and due to the success of the two-enzyme GBS approach being employed at NC State University, we decided to discontinue efforts to optimize RAD-seq and re-direct our efforts to the two-enzyme method described in (2) above.

March 2013 Progress Report:

1) The University of Florida is in the process of GBS of the CCLONES population using Agilent's SureSelect target enrichment in collaboration with a USDA funded planting breeding training grant. In all 850 individuals are being sequenced at 14,729 target loci. Libraries and selection is 80% complete.
2) NC State University has collected 1600 tissue samples from the PSSSS test site in Screven County, GA, and completed DNA extraction and quantitation of those samples. Library construction for sequencing of PstI-MspI double-digest restriction fragments is underway. In parallel, software development efforts are underway to create a pipeline for expedient analysis of the sequencing reads, including calling SNP haplotypes based on comparison to the loblolly pine reference genome sequence and haplotype-based analysis of the genomic basis of phenotypic variation measured at the Screven County GA test site.
3) At Texas A&M University, we applied the Agilent SureSelect Target Enrichment method to capture unigene-based targeted genomic sequences in loblolly pine (*Pinus taeda* L.). We used 35,386 out of 35,550 unigenes that were assembled by Dr. Chun Liang (Miami University, Oxford, Ohio) and available on http://bioinfolab.muohio.edu/txid3352v1 to design 647,634 oligonucleotide hybridization probes (baits). To make this approach more affordable for population studies, it is important to be able to use multiple barcoded individuals in a single hybridization reaction. Two single (A and B) and two multiplexed (C and D) DNA libraries were constructed and sequenced to test two multiplexed strategies: A and B were non-multiplexed samples representing DNA of the haploid megagametophyte and embryo from a single seed, respectively, while C and D were multiplexed pools composed of four and eight indexed individual DNA samples of megagametophytes, embryos or needles from two and seven individual trees, respectively. Each library was hybridized to the same number of probes. After capturing the targeted sequences, all samples were sequenced on an Illumina Hiseq2000 using paired-end sequencing (2×100bp). A and B were pooled and sequenced in lane 1, while C and D were sequenced in lanes 2 and 3, respectively.

   We obtained 70M, 275M and 234M reads (one direction) from lanes 1, 2 and 3, respectively. After filtering, high quality reads were mapped to the original unigenes and to the draft loblolly pine reference genome assembly (v0.9, provided by the PineRefSeq project; http://pinegenome.org/pinerefseq) using BWA and SAMtools. With the same mapping parameters, 56% of the reads obtained for non-multiplexed samples were mapped to unigenes, while 40% and 44% for C and D pools, respectively. 0.97% (including 762 unigenes) of the total unigene length were uncovered for non-multiplexed A and B samples, while 0.703% (including 534 unigenes) and 0.611% (including 438unigenes) for four-multiplexed C pool and eight-multiplexed D pool, respectively.

   97% and 93% of the reads in non-multiplexed samples and multiplexed pools were mapped to the draft loblolly pine reference genome assembly, respectively. SNP detection was done with each pool using SAMtools. 568,422; 2,370,705 and 2,810,893 SNPs were detected in the non-multiplexed, four multiplexed and eight multiplexed pools, respectively, with minimum read depth of 10 and 30% cutoff threshold. The SNP densities are 0.16337 SNPs/kb, 0.328SNPs/kb, 0.396SNPs/kb. The numbers of genotyped SNPs decrease in four-multiplexed and eight-multiplexed pools to 169,769 and 34,627 SNPs with read depth of at least 8 reads per each individual tree in the pools.

The conclusions are: 1) Multiplexing strategies worked well for capturing targeted sequences and SNP discovery. 2) Higher multiplexing level reduces the coverage of each sample, but still provides a high number of SNPs for efficient genotyping. 3) Sequencing depth for each sample can be increased by decreasing the number of targeted genes.

→ Comparison of methods:

November 2012 Interim Report:

1. As noted above, the RAD-Seq method was dropped from further consideration. No data have been obtained for DArT-seq, but that procedure differs only minor respects from the two-enzyme GBS protocol tested, so the key comparison to be made is between SureSelect and double-digest GBS.
2. Cost per sample for the SureSelect method as implemented was about ten-fold more than for GBS
3. Recovery of polymorphic SNPs was comparable; SureSelect identified about 8% more total SNPs than GBS identified candidate markers (51,623 total vs 47,131 candidates)
4. Ability to distinguish paralogous variants from allelic variants has not yet been tested for SureSelect; 35% of GBS candidate marker sequences map to single locations on the draft genome assembly
5. Suitability for high-throughput sequencing is similar for both methods, although SureSelect is more technically complex
6. The degree of focus on transcribed regions is greater for SureSelect; virtually all candidate SNPs are in or very near transcribed regions or sequences similar to transcribed regions, while 17.4% (8199) of the GBS candidate markers align to sequences in the reference pine transcriptome assembly.

March 2013 Progress Report:

1. No additional comparative analysis was carried out in this quarter – additional analyses and comparisons will be conducted after collection of the first complete datasets are available.

Year 2 Work Plan Tasks

→ Agree to standardized traits for analysis and a methodology for stratifying locations (April 2013).

November 2012 Interim Report

Analyses to date for development of Uniform Response Functions have been carried out independently by the three tree breeding programs, to avoid infringement on the rights of cooperators to the proprietary datasets held by those breeding programs. Discussions are needed among the project investigators to decide if a joint analysis of all data together would be beneficial. If so, then consultation with cooperative directors and advisory board representatives will be essential to develop a mechanism for merging data across cooperatives for a region-wide analysis while maintaining the intellectual property rights of cooperative members.

March 2013 Progress Report:

Potential merging of the data and standardization of traits and methodology will be discussed during the Aim Breakout session at the PINEMAP 2013 Annual Meeting.

→ Develop mixed model analysis for URF (October 2012).

March 2013 Progress Report:

In order to handle colinearity, and develop parsimonious, stable prediction models, NC State University employed variable reduction methods, such as LASSO and Ridge Regression and compared the subset of models with OLS regression. The predictive power of the best models is much lower than the proportion of variation explained by the OLS regression models, indicating that current models have relative low power to predict plantation yields forward into the future.

→ Develop outline for design of the PINEMAP DSS Deployment Tool

March 2013 Progress Report:

There will be two conference calls prior to the PINEMAP 2013 Annual Meeting to discuss progress, decision process, steps of DSS, intended audience and data restrictions. A DSS activity is scheduled for the PINEMAP 2013 Annual Meeting.

→ Continue Phenotyping for: insect and disease resistance; response to fertilizer; insect and disease resistance (November 2012).

November 2012 Interim Report:

Measurements of growth in field tests are typically taken in the fall and winter, during the dormant season.

March 2013 Progress Report:

NC State University had originally planned to measure a set of clonally-replicated progeny tests derived from advanced-generation parents to obtain data relevant to these objectives, but the drought of the past few years slowed growth, and the tests were not considered to have grown sufficiently to yield reliable phenotypes.

→ Reduction of Genome Complexity for generating molecular markers (April 2013).

November 2012 Interim Report:

1. Texas A&M: Attempts to map the sequence reads to the 18 Gb draft loblolly pine genome will continue. Protocols to maximize efficiencies will continue to be refined through 1) redesigning baits and 2) improved multiplexing of samples. DNA will be extracted from the 397 individuals in the ADEPT Association mapping population in preparation for genotyping. This is to support gene discovery through the analysis of the geographical distribution of alleles.
2. NC State University: Additional refinement of the double-digest GBS method will include testing an automated size-selection step to increase the reproducibility of recovery of genotypes at specific loci across multiple independent library preparations. DNA extraction and genotyping of tissue samples from progeny tests will be initiated. Customization of the open-source software used to analyze GBS datasets will be undertaken to improve the efficiency of data analysis and provide better integration with breeding program data management methods.
3. Virginia Tech: As noted above, the results from the RAD-seq approach were not encouraging. With colleagues at Texas A&M and NC State, we have therefore decided to focus on the two-enzyme approach and sequence capture. To improve the genomic coverage of data obtained from GBS, we plan to complete an additional test using non-methylation sensitive enzymes. We expect this method to provide data in non-repetitive regions, without a bias with regard to gene content. This will provide complementary data to that being generated by NC State using methylation sensitive enzymes (i.e., enriching for gene regions) for the purpose of genomic selection in loblolly pine pedigrees.

March 2013 Progress Report:

1. NC State University – preliminary experiments have been conducted to test the automated size-selection procedure and evaluate its effects on the library preparation workflow and on the quality of the resulting data. Occasional contamination of the size-selected fraction with high-molecular-weight DNA has been observed; it is not yet clear if this is due to a problem with the size-selection procedure or if it is an artifact of the nature of the samples submitted for size selection. The first set of libraries made using this procedure have been sequenced, but the data have not yet been analyzed, so firm conclusions regarding the benefit of the procedure cannot yet be drawn.

→ Continue Phenotyping for response to fertilizer (April 2013).

March 2013 Progress Report:

Aim 3 Article for *PINEMAP Press*

The purpose of PINEMAP's quarterly newsletter, the *PINEMAP Press* is to inform and educate on research results, programs, and outcomes of the PINEMAP project. The newsletter is distributed to internal project collaborators as well as stakeholders, NIFA, and university administrators.

Issue #3 will be published in May/June 2013. Please identify a topic for a brief (600 words or less) article for the May 2013 issue. Provide a preliminary title, list of authors, and a 3 -5 sentence summary.

Title:

Author(s):

Summary: