

PINEMAP Year 3 Progress Report 2

April 2014

Aim 3 (Genetics)

OUTCOMES/IMPACTS

Aim 3 activities will contribute to project-level outcomes and impacts by providing a genetic deployment tool necessary for stakeholders to make sound genetic deployment decisions under changing climatic conditions and by discovering genes and markers associated with important mitigation and adaptation traits that can be incorporated into future breeding programs. Aim 3 will work closely with Aims 1 and 2 to provide input and benefit from insights gleaned from the stand-level studies and modeling to make tree improvement more efficient. Simultaneously, knowledge of traits varying at the gene level should contribute to the ecophysiology studies.

OUTPUTS

Products

Peer-reviewed publications

Alberto, F.J., S.N. Aitken, R. Alia, S.C. González-Martínez, H. Hanninen, A. Kremer, F. Lefèvre, T. Lenormand, S. Yeaman, R.W. Whetten, and O. Savolainen. 2013. Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology* 19(6): 1645-1661. doi: <http://dx.doi.org/10.1111/gcb.12181>

Chhatre, V., T. Byram, D.B. Neale, J.L. Wegrzyn, and K.V. Krutovsky, 2013 Genetic structure and association mapping of adaptive and selective traits in the East Texas loblolly pine (*Pinus taeda* L.) breeding populations. *Tree Genetics and Genomes* 9(5): 1161-1178.

Eckert, A. J., J. L. Wegrzyn, J. D. Liechty, J. M. Lee, W. P. Cumbie, J. M. Davis, B. Goldfarb, C. A. Loopstra, S. R. Palle, T. Quesada, C. H. Langley, and D. B. Neale. 2013. The evolutionary genetics of the genes underlying phenotypic associations for loblolly pine (*Pinus taeda*, *Pinaceae*). *Genetics* 195: 1353-1372.

Koralewski, T.E., Brooks, J.E. and K.V. Krutovsky. 2014 Molecular evolution of drought tolerance and wood strength related candidate genes in loblolly pine (*Pinus taeda* L.). *Silvae Genetica* (in press)

- Neale, D.B., J. L. Wegrzyn, K.A. Stevens, A.V. Zimin, D. Puiu, M.W. Crepeau, C. Cadeno, M. Koriabine, A.E. Hotz-Morris, J.D. Liechty, P.J. Martinez-Garcia, H.A. Vasquez-gross, B.Y. Lin J.J. Qieve, W.M. Dougherty, S. Furentes-Soriano, L.-S. Wu, D. Gilbert, G. Marcais, M. Rogerts, C. Holt, M. Yandell, J.M. Davis, K.E. Smith, J. FD Dean, W.W. Lorenz, R.W. Whetten, R. Sederoff, N. Wheeler, P.E. McGuire, D. Main, C.A. Loopstra, K. Mockett, P.J. deJong, J.A. Yorke, S.L. Salzberg, C.H. Langley. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology*. <http://genomebiology.com/2014/15/3/R59>.
- Palle, S. R., C. M. Seeve, A. J. Eckert, J. L. Wegrzyn, D. B. Neale, C. A. Loopstra. 2013. Association of loblolly pine xylem development gene expression with single nucleotide polymorphisms. *Tree Physiology* 33: 763-774.
- Resende, M.F.R., Jr., P. Muñoz, J.J. Acosta, G.F. Peter, J.M. Davis, D. Grattapaglia, M.D.V. Resende, and M. Kirst. 2012. Accelerating the domestication of trees using genomic selection: Accuracy of prediction models across ages and environments. *New Phytologist* 193: 617-624. doi: <http://dx.doi.org/10.1111/j.1469-8137.2011.03895.x>
- Resende, M.F.R., Jr., P. Muñoz, M.D.V. Resende, D.J. Garrick, R.L. Fernando, J.M. Davis, E.J. Jokela, T.A. Martin, G.F. Peter, and M. Kirst. 2012. Accuracy of genomic selection methods in a standard dataset of loblolly pine (*Pinus taeda* L.). *Genetics* 190: 1503-1510. doi: <http://dx.doi.org/10.1534/genetics.111.137026>
- Wegrzyn JL, Liechty JD, Stevens KA, Wu L-S, Loopstra CA, Vasquez-Gross HA, Dougherty WM, Lin BY, Zieve JJ, Martinez-Garcia PJ, Holt C, Yandell M, Zimin AV, Yorke JA, Crepeau MW, Puiu D, Salzberg SL, de Jong PJ, Mockett K, Main D, Langley CH, Neale DB: Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 2014, 196:891-909
- Yeaman S, Hodgins KA, Suren H, Nurkowski KA, Rieseberg LH, Holliday, JA, Aitken SN (2014) Conservation and divergence of gene expression plasticity following ~140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (natural hybrid populations of *Picea glauca* and *Picea engelmannii*). *New Phytologist* (In press).

Other publications

- Byram, T.D., T.E. Koralewski and E.M. Raley. 2013. A climate change response function for loblolly pine (*Pinus taeda* L.) from the Western Gulf region of the United States. In: Proceedings of the 32nd Southern Forest Tree Improvement Conference. Advancing the Value of Forest Plantations. June 10-13, 2013, Clemson, SC, USA. <http://www.sftic.org>, p 36
- Byram, T., T. Koralewski, G. Peter, J. Zhang, F. Isik, A. Farjat. 2013. Using Historical Progeny Tests to Optimize Pine Breeding and Deployment Strategies. PINEMAP (Pine Integrated Network: Education, Mitigation, and Adaptation Project) Year 2 Annual Report | March 2012-February 2013 "Mapping the future of southern pine management in a changing world". Pp. 28-29 <http://www.pinemap.org/reports/annual-reports/PINEMAP Year 2 Annual Report FINAL.pdf>

- Chhatre, V.E., M.F.R. Resende Jr., P.R. Munoz, G.F. Peter, J.M Davis, M.M. Kirst, C.S. Echt, K.V. Krutovsky, C. Dana Nelson, 2013. Multiple pedigrees allow construction of a densely populated reference linkage map in loblolly pine (*Pinus taeda* L.). In: Proceedings 32nd Southern Forest Tree Improvement Conference, Edited by Michael W. Cunningham, p. 53, June 10-13, 2013. Clemson, SC, USA
(<http://www.sftic.org/f/2013%2032nd%20SFTIC%20Proceedings.pdf#page=67>)
- Farjat, A., F. Isik, R. Whetten, S. McKeand, 2013. Modeling climate change effects on the growth of loblolly pine seed sources in the southeastern United States. In: Proceedings of the 32nd Southern Forest Tree Improvement Conference; 10-13 June 2013, Clemson South Carolina, USA. <http://www.sftic.org>, pp 24-26.
- Krutovsky K., Byram T., Whetten R., Wheeler N., Neale D., Lu M., Koralewski T., Loopstra C. PINEMAP + PineRefSeq = Future Forests. PINEMAP (Pine Integrated Network: Education, Mitigation, and Adaptation Project) Year 2 Annual Report | March 2012-February 2013 “Mapping the future of southern pine management in a changing world”. 2013. P. 26-27
(<http://www.pinemap.org/reports/annual-reports/PINEMAP Year 2 Annual Report FINAL.pdf>)
- Steiger, J.J., B.P. Bullock, R.C. Abt, R.W. Whetten, 2013. Impacts of different levels of genetic homogeneity on juvenile stem characteristics and potential stand-level value in loblolly pine. In: Proceedings of the 32nd Southern Forest Tree Improvement Conference; 10-13 June 2013, Clemson South Carolina, USA. <http://www.sftic.org> , pp 27-29.
- Wheeler, N.C., R. Whetten, 2013. The pine reference genome sequence and applied tree breeding. In: Proceedings of the 32nd Southern Forest Tree Improvement Conference; 10-13 June 2013, Clemson South Carolina, USA. <http://www.sftic.org>, pp 80-84.
- Whetten R., Krutovsky K., Holliday J. Comparing Genotyping Technologies for Efficiency and Cost-effectiveness. PINEMAP (Pine Integrated Network: Education, Mitigation, and Adaptation Project) Year 2 Annual Report | March 2012-February 2013 “Mapping the future of southern pine management in a changing world”. 2013. P. 30-31
(<http://www.pinemap.org/reports/annual-reports/PINEMAP Year 2 Annual Report FINAL.pdf>)
- Whetten R., Krutovsky K., Holliday J. Comparing Genotyping Technologies for Efficiency and Cost-effectiveness. PINEMAP Press (the Pine Integrated Network: Education, Mitigation, and Adaptation Project newsletter). 2012. №1: P. 6
(<http://www.pinemap.org/publications/quarterly-newsletter-pinemap-press/Issue 1 Fall 2012.pdf>)

Events/Activities

Presentations

Author(s)/Presenter(s)	Title	Type	Date	Venue/Location
Byram, T.	PINEMAP Overview	Presentation (Meeting)	May 2011	Western Gulf Forest Tree Improvement Program Contact Meeting
Byram, T.	PINEMAP Overview	Presentation (Meeting)	February 8, 2012	Drought Impacts to Forestlands and Where are We Going? Symposium, Diboll, TX
Byram, T., T. Koralewski, E. Raley.	Strengths and weaknesses of universal response function approaches for support of deployment decisions	Presentation (Meeting)	April 24, 2013	PINEMAP 2013 Annual Meeting, Athens, GA
Byram, T., T. Koralewski, E. Raley.	WGFTIP Seed Source Study and Seed Movement	Presentation (Meeting)	May 16, 2013	WGFTIP Contact Representatives Meeting
Byram, T., T. Koralewski	A climate change response function for loblolly pine (<i>Pinus taeda</i> L.) from the Western Gulf region of the United States	Presentation (Conference)	September 11, 2013	Southern Forest Tree Improvement Conference, Clemson, SC
Chhatre, V., T. Byram, D.B. Neale, J.L. Wegrzyn, and K.V. Krutovsky	Genome wide analysis of genetic associations with environmental variables in East Texas loblolly pine (<i>Pinus taeda</i> L)	Presentation (Conference)	January 14-18, 2012	Plant & Animal Genome XX. The International Conference on the Status of Plant and Animal Genome Research, San Diego, CA
Dasgupta, M. G., V. Dharanishanthi, and K.V. Krutovsky	Development of Genetic Markers in Eucalypt Species by Target Enrichment and Exome Sequencing	Poster Presentation	January 11-15, 2014	Plant and Animal Genome XXII. The International Conference on the Status of Plant and Animal Genome Research, San Diego, CA
Farjat, A., F. Isik, R. Whetten, and S. McKeand	Modeling the Growth of Loblolly Pine in the Southeastern United States	Poster Presentation	April 24-26, 2013	PINEMAP Annual Meeting, Athens, GA
Farjat, A., F. Isik, R. Whetten, and S. McKeand	Modeling Climate Change Effects on the Growth of Loblolly Pine Seed Sources in the Southeastern United States	Presentation (Meeting)	May 22 – 23, 2013	NCSU Cooperative Tree Improvement Program Advisory Board Meeting, Raleigh NC

Author(s)/Presenter(s)	Title	Type	Date	Venue/Location
Farjat, A., F. Isik, R. Whetten, S. McKeand	Modeling climate change effects on the growth of loblolly pine seed sources in the southeastern United States	Presentation (Meeting)	June 10 – 13, 2013	32nd Southern Forest Tree Improvement Conference, Clemson SC.
Holliday, J.A. and L. Zhou	Toward genome-wide association mapping in a model tree	Presentation (Meeting)	October 31, 2012	Texas Tech University Department of Biology Seminar, Lubbock, TX
Holliday, J.A. and L. Zhou	Toward genome-wide association mapping in a model tree	Presentation (Meeting)	October 15, 2012	University of Maryland Department of Plant Science Seminar, College Park, MD
Isik, F.	Universal Response Function development from the Plantation Seed Selection Source Study field trial data	Presentation (Meeting)	October 25, 2012	North Carolina State Cooperative Tree Improvement Program Contact Meeting, Asheville, NC
Koralewski, T.E, T.D. Byram, E.M. Raley	A climate change response function for loblolly pine (<i>Pinus taeda</i>) from the Western Gulf region of the United States	Presentation (Conference)	February 4-7, 2013	Breeding for Value in a Changing World, Jacksonville, FL
Koralewski, T.E, T.D. Byram, E.M. Raley	A climate change response function for loblolly pine (<i>Pinus taeda</i> L.) from the Western Gulf region of the United States	Poster Presentation	April 24-26, 2013	PINEMAP Annual Meeting, Athens, GA
Krutovsky, K.V.	Forest genomics for mitigating climate change and breeding resilient trees.	Presentation (Conference)	January 14-18, 2012	Plant and Animal Genome XX. The International Conference on the Status of Plant and Animal Genome Research, San Diego, CA
Krutovsky, K.V.	Forest genomics for mitigating climate change and breeding resilient trees, keynote presentation	Presentation (Conference)	October 3-5, 2012	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.	Forest ecogenomics and genomic selection for mitigating climate change and breeding resilient trees	Presentation (Conference)	October 16-17, 2012	Final Conference and Workshops of Noveltree Project, Tree Breeding and Evolutionary Biology: New Synergies to Tackle the Impact of Climate Change in the 21st Century, Helsinki, Finland

Author(s)/Presenter(s)	Title	Type	Date	Venue/Location
Krutovsky, K.	Population Genetic and Genomic Approaches to Mitigate Global Climate Change Impacts on Forest Genetic Resources and to Breed More Resilient Trees	Training/Workshop	Sept. 24-28, 2012	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), Göttingen, Germany
Krutovsky, K.V.	High-throughput genome-wide genotyping and targeted sequencing in loblolly pine (<i>Pinus taeda</i> L.)	Presentation (Meeting)	June 7-9, 2013	Second meeting of the Alpine Forest Genomics Network (AForGen), Neukirchen am Großvenediger, Salzburg, Austria
Krutovsky, K.V.	The <i>Pinus sibirica</i> and <i>Larix sibirica</i> genome projects	Presentation (Summit)	June 14-17, 2013	The 2013 Conifer Genome Sequencing Summit in Björkliden, Lapland, Sweden
Krutovsky, K.V.	Forest tree genomics for understanding local adaptation and population responses to changing climate	Invited presentation (Colloquium)	December 10, 2013	Ecology, Conservation and Biodiversity Colloquium, University of Marburg, Germany
Krutovsky, K.V.	Innovative approaches for <i>de novo</i> sequencing of large complex genomes.	Presentation (training workshop)	July 19-21, 2013	Scientific Training School "Genomic Sequencing and Data Analysis", Novosibirsk, Russia (http://conf.nsc.ru/seq2013)
Krutovsky, K.V.	Targeted and complete genome <i>de novo</i> sequencing in conifer trees with giant and complex genomes	Presentation (Conference)	July 21-25, 2013	"High-Throughput Sequencing in Genomics" (HSG-2013), Novosibirsk, Russia (http://conf.nsc.ru/HSG/scientific_program ; http://conf.nsc.ru/files/conferences/HSG/183699/%D1%82%D0%B5%D0%B7%D0%B8%D1%81%D1%8B%20HSG_end.pdf)
Krutovsky, K.V., V.E. Chhatre, M. Lu, T.D. Byram, J.L. Wegrzyn, D.B. Neale, and C. Loopstra	High-Throughput Genome-Wide Genotyping, Targeted Sequencing and Association Mapping of Adaptive and Breeding Traits in Loblolly Pine (<i>Pinus taeda</i> L.) Populations	Presentation (Conference)	January 11-15, 2014	Plant and Animal Genome XXII. The International Conference on the Status of Plant and Animal Genome Research, San Diego, CA
Lu, M., C. Loopstra, and K. Krutovsky	Preliminary SNP Discovery in Loblolly Pine (<i>Pinus taeda</i> L.) Exome	Presentation (Conference)	January 12-16, 2013	P0459. Plant and Animal Genome XXI
Lu, M., C. Loopstra, and K. Krutovsky	Genome Target Sequencing in Loblolly Pine (<i>Pinus taeda</i> L.) using Different Multiplexing Strategies	Poster Presentation	April 24-26, 2013	PINEMAP Annual Meeting, Athens, GA

Author(s)/Presenter(s)	Title	Type	Date	Venue/Location
Lu, M., C. Loopstra, and K.V. Krutovsky	Genome Target Sequencing in Loblolly Pine (<i>Pinus taeda</i> L.) using Different Multiplexing Strategies	Presentation (Symposium)	May 15-16, 2013	The Molecular and Environmental Plant Sciences (MEPS) symposium "Plant signaling systems – from cells to environment", Texas A&M University, College Station, TX
Martin,T.A.	PINEMAP Overview	Presentation (Meeting)	April 2011	Cooperative Forest Genetics Research Program Annual Meeting
Martin, T.A., G.F. Peter, T. Byram, and R. Whetten	Genetics research and outreach in support of enhanced climate adaptation and mitigation in southern pine	Presentation (Meeting)	October 22-24, 2012	NIFA Project Directors' meeting, Cincinnati, OH
McKeand, S., J. Steiger, J. Andrews, D. Barker, R. Whetten, T. Brooks, F. Isik.	And they're off! - Fourth-cycle of loblolly pine breeding has begun in the NCSU Cooperative Tree Improvement Program	Presentation (Meeting)	June 10 -13, 2013	32nd Southern Forest Tree Improvement Conference, Clemson SC.
Townsend, L. and R. Whetten	Identifying Genetic Variation in Site Adaptability in Loblolly Pine	Poster Presentation	April 24-26, 2013	PINEMAP Annual Meeting, Athens, GA
Townsend, L., R. Whetten.	Identifying Genetic Variation in Site Adaptability in Loblolly Pine	Poster Presentation (best poster award)	June 10 – 13, 2013	32nd Southern Forest Tree Improvement Conference, Clemson SC.
Townsend, L., R. Whetten	High-throughput sample collection and DNA extraction methods for tree breeding	Presentation (Meeting)	May 22 – 23, 2013	NCSU Cooperative Tree Improvement Program Advisory Board Meeting, Raleigh NC
Wheeler,N.C., R. Whetten	The pine reference genome sequence and applied tree breeding	Presentation (Meeting)	June 10 – 13, 2013	32nd Southern Forest Tree Improvement Conference, Clemson SC.
Whetten, R.	PINEMAP Overview	Presentation (Meeting)	November 2011	North Carolina Cooperative Tree Improvement Program meeting
Whetten, R.	Genetic Approaches to Understanding and Increasing the Resilience of Pine Plantations to Climate Change	Webinar	November 16, 2012	PINEMAP Internal Webinar Series
Whetten, R	Rust resistance screening and new molecular marker methods for pine	Presentation (Meeting)	October 24, 2012	North Carolina State Cooperative Tree Improvement Program Contact Meeting, Asheville, NC

Author(s)/Presenter(s)	Title	Type	Date	Venue/Location
Whetten, R.	The Pine Reference Genome Sequence and Applied Tree Breeding Programs	Presentation (Conference)	January 12, 2013	Pine Reference Genome Workshop at the 21st International Plant and Animal Genome Conference, San Diego CA USA
Whetten, R. and L. Townsend	Identifying genetic variation in site adaptability in loblolly pine [poster]	Poster Presentation	May 15-16, 2012	PINEMAP Annual Meeting, Atlanta, GA
Whetten, R., W. Kohlway, L. Townsend	Impacts of cost-effective high-throughput genotyping of loblolly pine on applied tree breeding programs	Presentation (Meeting) (best oral presentation award)	June 10 – 13, 2013	32nd Southern Forest Tree Improvement Conference, Clemson SC.
Whetten, R.W.	PINEMAP research and applied tree breeding	Presentation (Meeting)	May 22 – 23, 2013	NCSU Cooperative Tree Improvement Program Advisory Board Meeting, Raleigh NC
Zhang, J., S. Gezan, and G.F. Peter	Integrating Climate and Genetic Effects of Loblolly Pine by Universal Response Functions	Poster Presentation	April 24-26, 2013	PINEMAP Annual Meeting, Athens, GA
Zhao, L. and J.A. Holliday	Targeted enrichment of the black cottonwood (<i>Populus trichocarpa</i>) gene space using sequence capture	Presentation (Conference)	January 12-16, 2013	Plant and Animal Genome Conference. San Diego, CA

Trainings, workshops, and courses

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 challenges posed by global climate change, 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>)

The 2013 Western Gulf Forest Tree Improvement Contact Representatives Meeting, May 15-16, 2013 in Idabel OK, focused on the PINEMAP program. Tree improvement practitioners heard

talks from Rod Will, 'Advancing silviculture in the era of climate change: An overview of PINEMAP', and Tom Byram, 'WGFTIP Seed Source Study and seed movement'. In addition they visited the Oklahoma Tier III site and were introduced to the breadth of the research data being collected.

Krutovsky, K.V. served as Instructor and Lecturer at the School of Young Scientists "Genome Sequencing and Data Analysis", organized by the Institute of Chemical Biology and Fundamental Medicine, Siberian Branch, Russian Academy of Science, July 19-21, 2013, Novosibirsk, Russia; <http://conf.nsc.ru/seq2013/en>); presented a lecture on "Innovative approaches for *de novo* sequencing of large complex genomes", <http://conf.nsc.ru/seq2013>)

The NCSU Cooperative Tree Improvement Program conducted a Tree Improvement Short Course in Birmingham AL, Nov 5 – 7, 2013. S. McKeand, F. Isik, and R. Whetten presented lectures on strategies for tree breeding, including sections on geographic variation in loblolly pine, breeding for site adaptability and resilience, and analysis of performance data from multi-site progeny trials.

Experiments, surveys, and data collection

Genome complexity reduction methods for genotyping by sequencing

The University of Florida: Needle samples from 832 individuals representing 61 families of loblolly pine were collected at Nassau (Florida, USA). DNA was extracted from all freeze dried needle samples. Previously, a total of 54,773 oligonucleotide probes were designed to capture 14,729 genes, with a target region size of 6.57 Mbp (Neves et al. 2013). Following a pilot study, probes that did not perform well due to low capture efficiency and probes that captured repetitive regions were excluded. For this project, a subset of 38,507 probes will be used, representing 11,316 genes. Multiplexed exome capture will be performed following Agilent's Sure Select Target Enrichment protocol (Gnirke et al. 2009). Sequencing libraries will be pooled together, resulting in 104 pools containing 8 samples. The captured DNA will be eluted from the beads and the 104 captured pools will be combined in pairs, generating 52 pools containing 16 genotypes each. The 52 pools will be sequenced in separate lanes of the Illumina HiSeq 2000 platform. Library preparation and capture has been completed for 96 capture pools. We expect all capture pools will be completed by the end of September, 2013. Sequencing will commence once pools are completed.

North Carolina State University: Tissue samples were collected from 1600 trees in a Plantation Selection Seed Source Study field trial growing in Screven County, Georgia in October 2012. DNA extraction and preparation of Illumina sequencing libraries were successful for 1470 samples, and sequence data files are in hand of paired-end 100-nt reads from 16 lanes of HiSeq runs. The sequencing was carried out on 16 libraries, each consisting of 90 to 96 samples in a microtiter plate, which were indexed and pooled in groups of four; each pool of four libraries was sequenced in four Illumina lanes. This experimental design means that variation in library quality is orthogonal to variation in sequencing lane quality, so genotyping data analysis can resolve these sources of variation. Preliminary quality assurance analysis of the sequencing data

shows that the data are generally of good quality, with a total yield of 3.3 billion paired-end reads and 660 gigabases of raw DNA sequence at a cost (reagents, consumable supplies, and service fees) of less than \$25 per sample, including DNA extraction, library preparation, and sequencing. Further analysis of these data using the STACKS software package suggest that many of the sequenced regions of the genome contain repetitive DNA sequences, so the yield of single-copy regions useful as genetic markers is lower than expected based on the yield of reads. Individual samples typically have data for a few hundred to a few thousand loci, each of which may be polymorphic at one to several nucleotide positions. Analysis of the sequence data using alternative approaches is underway to test whether the yield of useful marker loci is a function of STACKS analytical methods rather than intrinsic properties of the data themselves.

Virginia Tech: The Virginia Tech group is conducting genotyping-by-sequencing on the Plantation Selection Seed Source Study (PSSSS) in Appomattox County, Virginia. Genotyping of this study is of interest due to its location at the northern edge of the current loblolly pine range, and will complement co-PI Whetten's sampling/genotyping many of the same families planted in Georgia. This will allow estimation of genotype-by-environment effects and allow prediction of performance in northern climates that are expected to become available for pine plantations with climate change. The Appomattox trial was planted in 1996, and is comprised of 102 pollen mix families, of which 70 are in common with the Georgia site. Available phenotypes include height and diameter, measured in years four, eight, and twelve, and cold injury measured following significant freeze events. We have sampled approximately 500 individuals and will complete an additional 500 in early spring 2014. Library construction is underway using the restriction site-associated DNA sequencing (RAD-seq) approach. The restriction enzyme PstI was chosen for maximal overlap with data generated by the group led by Co-PI Whetten at NC State. A preliminary run of 48 samples on the Illumina HiSeq at the Virginia Bioinformatics institute yielded approximately one million paired-end reads per sample. Completed libraries will be multiplexed in ~10 lanes on an Illumina HiSeq instrument at Virginia Tech following library preparation.

Texas A&M University: Agilent SureSelect Target Enrichment method was used 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). 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 in Illumina HiSeq2000 using paired-end sequencing (2×100 bp). 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 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, 92% to 97% of the reads in each

sample were mapped to the draft loblolly pine reference genome assembly, respectively. 51% to 59% of the mapped reads were uniquely mapped reads.

Among the uniquely mapped reads, the sequencing depth of four-multiplexed pool is almost twice than that of eight-multiplexed pool. For checking the efficiency of the baits in different multiplexed strategies, 41% to 56% of the total filtered reads can be mapped to the unigenes. 625,962 and 608,406 baits were found can be mapped by uniquely mapped reads in four-multiplexed and eight-multiplexed pools, respectively, indicating most of the baits contribute to the capture.

SNP detection was done with each library using SAMtools and Freebayes. 2,370,705 (0.328 SNPs/kb) and 2,810,893 SNPs (0.396 SNPs/kb) were detected in the four-multiplexed and eight multiplexed pools, respectively, with minimum read depth of 10 and 30% cutoff threshold by using SAMtools. 1,870,997 SNPs in four-multiplexed and 1,816,724 SNPs in eight-multiplexed pools can be approved by Freebayes. The number of genotyped SNPs decreases in four-multiplexed and eight-multiplexed pools to 168,256 and 34,342 SNPs with read depth of at least 8 reads per each individual tree in the pools.

Preliminary loblolly pine exome sequencing and SNP genotyping using the Agilent SureSelect Target Enrichment system and hybridization probes based on 35,550 unigenes from PineDB (<http://bioinfolab.muohio.edu/txid3352v1>) demonstrated the great potential of the exome capturing technique for SNP discovery and genotyping. The newly published loblolly pine gene annotation (Wegrzyn et al 2014; "Pita_Annotation_v2" PINEREFSEQ, http://loblolly.ucdavis.edu/bipod/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01/Pita_Annotation_v2/) revealed a more complete and less redundant set of exons which will improve the completeness and specificity of our exome capture. We will take advantage of this newly published exon annotation and a less expensive sequence capture system from Nimblegen (SeqCap EZ System) to decrease the cost and to increase efficiency of the SNP genotyping in our experiment. A total of 199,741 exons were extracted from the loblolly pine v.1.01 reference genome (PINEREFSEQ project, http://loblolly.ucdavis.edu/bipod/ftp/Genome_Data/genome/pinerefseq/Pita/v1.01/), including 195,404 coding regions and 4,337 untranslated regions. After length filtering and duplication removal, 43 Mb of exon sequences were sent to Nimblegen for probe design. The DNA extractions for the ADEPT II population (394 trees) have been done. Ten samples will be multiplexed and sequenced in one lane. The SNP detection pipeline using SAMtools was developed during the previous analysis. Preparations for our phenotyping work are in progress. We were awarded an intern from the 2014 PINEMAP undergraduate fellowship program to help us phenotype traits of importance including specific leaf area, stomata density, wood density and growth, branch angle and carbon isotope discrimination. We will travel to Gulfport, Mississippi in May to collaborate with Dana Nelson (PINEMAP, US Forest Service) on the collection of the phenotypic data this summer.

Genetic deployment tool

University of Florida: Climate data was obtained from the PRISM data and SECC. In the Lower Gulf Provenance tests, tree height and DBH were analyzed at age 10, and from the Florida wild

seed trials, age 6 tree size was used. A number of analytical methods were tested to relate provenance and site climate variables with performance. Due to the high extent of multicollinearity and autocorrelation present in the climate data, the best models were obtained using the penalized regression method, least absolute shrinkage and selection operator (LASSO).

With the PRISM data and provenance-progeny tests, the best models accounted for ~44% of the variation in tree size. The site and provenance mean coldest monthly temperature (MCMT) had the strongest effect and together accounted for 35% of the variation. In addition to MCMT, provenance longitude and site mean annual precipitation (MAP) and site mean warmest month temperature (MWMT) were significant. Analysis of the provenance-progeny test data with the SECC climate data showed the same variables site and provenance MCMT, provenance longitude, and site MWMT and MAP. In addition, site mean annual radiation was also significant. The PRISM data does not contain this variable. Together, these variables accounted for ~53% of the variation. The conclusion is that MCMT is the most significant climate predictor and agrees well with previously published results by Schmidting.

With the age 6 data from the Florida wild seed trial and SECC or PRISM data, models accounted for substantially lower variation, ~0.1 or 10% of the variation. This may be due to the early age data available.

We are currently using a Bayesian model selection approach to identify the best models and terms for predicting performance with climate variables.

North Carolina State University: Geo-spatial statistical models were developed to estimate the effect of yearly average minimum temperatures on loblolly pine performance throughout the Southeastern United States. A poster about this work was presented in the 2012 PINEMAP annual meeting. Statistical models were developed to predict the responses of different seed sources to climate change using climate variables as predictors. The models can be used as a quantitative tool to predict the effects of climatic variables and genetic effects (pine seed sources) on the performance of pine plantations. Also, the model could also be used to identify superior performing families through simulation studies, and to estimate their growth potential for a given planting site under a given future climate scenarios. A poster of this work was presented in the 2013 PINEMAP annual meeting and a draft manuscript has been completed. The manuscript will be submitted to a journal in spring 2014.

Texas A&M University: To account for soil conditions across the region, additional analyses of planted tree volume at age 15 were performed using WaSSI-derived data for watersheds in the loblolly pine range. The investigated variables included PPT (monthly total precipitation), AET (monthly total actual evapotranspiration rate), PET (monthly total potential evapotranspiration rate), GEP (monthly total gross ecosystem productivity), SOILPSAT (monthly mean soil water storage expressed as a percent of saturation), and AET/PET ratio as a drought severity index, all averaged over summer months (June - August). Coefficient of variation (CV) was calculated for each of the variables. The means and their corresponding CVs along with T_M (minimum value of monthly mean air temperature) were used as independent variables. We observed very moderate improvement over the previously developed models that were based on the PRISM weather data when predicted East-West seed movement but no improvement in predicting North-South seed movement.

MILESTONES

Version 1 of genetic deployment tool with provenance information

Texas A&M University: 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. Incorporating numerous soil moisture projections from the WaSSI models resulted in very small improvements in the explanatory power of the equations over that achieved with the values from the PRISM weather dataset. It is likely that the Aridity Index (AI) calculated from the weather data captured much of the same information supplied by the WaSSI models.

North Carolina 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 revealed 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. Additional analysis of variation among families within seed sources, using maximum and minimum annual temperature and total precipitation as climate variables, shows that 40% to 60% of the variation in family performance relative to locally-

adapted checklot families can be explained by three climate factors. The predictive power of these models will be tested by cross-validation to evaluate the stability of the models, and additional climate factors will be tested to assess which models have the best predictive power. If this method is successful with the PSSSS data, it can be extended to include other progeny test data to evaluate the approach across a more diverse set of germplasm from the current breeding population.

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 multicollinearity 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.

Version 2 of genetic deployment tool with progeny information

All three institutions identified some measurement of winter temperature and water availability as key elements in the development of explanatory models. The next steps needed to develop Version 2 of the genetic deployment tool are to convert these to predictive equations and to produce regionally appropriate representation software. The AIM 3 Group will be working with the Deployment Decision Support Group to produce an interactive web application to produce 'heat-maps' showing historic and projected values for climate variables identified as important in our analysis.

Determine appropriate genome reduction methods for genotyping by sequencing

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 plant breeding training grant. In all 850 individuals are being sequenced at 14,729 target loci. Libraries and selection is 80% complete.

North Carolina 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 complete, and data are in hand. 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. SNP

genotype calling is underway using the 660 billion basepairs of DNA sequence obtained from the Georgia PSSSS site, using both existing and novel software tools, and development of a haplotype-based method of analysis will follow in turn. The relatively low yield of marker loci from the double-digest GBS method should still be useful for construction of kinship matrices based on Identity-By-Descent using reference-genome-based haplotypes, and these matrices will be incorporated into genetic analysis of the PSSSS data to assess the added value obtained from more precise measures of relationships. We are working in collaboration with Trudy Mackay's research group in the Genetics Program at NC State to re-assess the potential value of RNA-seq measurements of gene expression as predictors of phenotypic variation, based on unpublished results they have recently obtained in *Drosophila* lines derived from wild populations. Independent analysis of public data on gene expression levels and phenotypic variation in the mouse Collaborative Cross has confirmed the observation that genetic variation in levels of gene expression can be used to train a statistical model with some power to predict phenotypic variation, so a pilot experiment to determine the scale and power of an RNA-seq experiment on the PSSSS population is being planned.

At Texas A&M University, application of the Agilent SureSelect Target Enrichment method led to the following conclusions: 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 high number of SNPs for efficient genotyping. 3) Sequencing depth for each sample can be increased by decreasing number of targeted genes.

Genotyping using Nimblegen SeqCap EZ System with the exon annotation based probes is underway.

BROAD IMPACTS

The NCSU Cooperative Tree Improvement Program is also participating in another NIFA Coordinated Agricultural Project, the Integrated Biomass Supply System (IBSS) project funded at University of Tennessee-Knoxville under the bioenergy component of the CAP program. The IBSS research conducted by NCSU CTIP is focused on chemical and physical properties of loblolly pine biomass, with a strong focus on wood. These properties are relevant to PINEMAP objectives because they affect the market for southern pine forest products, including solid wood products, pulp and paper as well as potential bioenergy markets that may develop. In addition, close ties are maintained with the PineRefSeq CAP, also funded by the bioenergy component of the CAP program, to fully exploit the value of the draft loblolly pine genome sequence for the breeding and genetic analysis objectives of the IBSS and PINEMAP research efforts.

TRAINING

Last name	First name	Position	University	Role
Bawa	Rajesh	Ph.D. Student	VT	Research focus: Conducting Phenotyping of adaptive traits (i.e., cold hardiness and bud phenology) in the loblolly pine progeny trials and working with universities across region to conduct genome-wide association mapping between these traits and the re-sequencing data.
Farjat	Alfredo	Ph.D. Student	NCSU	Research focus: Carrying out joint analysis of progeny test data and climate data for the Cooperative Tree Improvement Program.
Koralewski	Tomasz	Postdoc	TAMU	Building a location database for the progeny tests and the parents represented in those progeny tests that can be linked to climatic data, and performing URF analysis to optimize seedling deployment.
Lu	Mengmeng	Ph.D. Student	TAMU	Research focus: Evaluating efficiency of the Agilent SureSelect target gene enrichment system for Illumina GAI and HiSeq high-throughput sequencing instruments, based on capture probes designed using the Agilent eArray tool and the most recent and comprehensive unigene set consisting of 35,550 sequences representing individual genes expressed in loblolly pine.

Kohlway	Will	Undergraduate Research Assistant	NCSU	Working on pilot scale experiments to test genotyping-by-sequencing as a cost-effective means of genotyping the experimental populations.
Rusche	Ben	Undergraduate Research Assistant	NCSU	Working on pilot scale experiments to test genotyping-by-sequencing as a cost-effective means of genotyping the experimental populations.
Townsend	Laura	MS	NCSU	Research focus: Genetic variation in loblolly pine for adaptability to climate variables.
Zhang	Jianxing	Ph.D. Student	UF	Research focus: conducting analyses linking progeny test performance with climatic data and developing a database for the URF analysis to optimize seedling deployment.