Genetic Approaches to Understanding and Increasing the Resilience of Pine Plantations to Climate Change

Ross Whetten, presenter

on behalf of Aim 3, the Genetics Group

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Members of the Genetics Group

- University of Florida – Gary Peter (investigator), Jianxing Zhang (Ph.D student)
- North Carolina State University – Ross Whetten, Fikret Isik, Steve McKeand (investigators), Alfredo Farjat (Ph.D student), Laura Townsend (M.S. student), Will Kohlway and Ben Rusche (undergrads)
- Texas A&M University – Tom Byram, Carol Loopstra, Kostya Krutovsky (investigators), Tomasz Koralewski (postdoc), Mengmeng Lu (Ph.D student)
- Virginia Tech – Jason Holliday (investigator), Rajesh Bawa (Ph.D student)
- Dana Nelson – US Forest Service (investigator)

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Outline of presentation

- An overview of project objectives , the genetics component, and the focus of this presentation
- Some background in quantitative genetics
- Understanding mechanism vs guiding breeding
  - Are these different objectives or one and the same?
- Experimental methods, materials & results
- Plans for year 3
- Challenges to be overcome

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Pine Forests of the Southeastern US

- Forests occupy 60% of the land area, with a large fraction dominated by pines
  - 10 species total; two (loblolly and slash) are economically important
- ~85% of all forestlands are privately owned
- About half the pine forest is naturally regenerated, **half planted with genetically improved seedlings**
  - About 10 million ha / **25 million ac** each
- Contains 12 Pg of C, 36% of the sequestered forest C in the contiguous United States
- Annually sequester 76 Tg C, equivalent to 13% of regional greenhouse gas emissions
- Produce about 16% of global industrial wood, more than any other country

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

University – forest industry research cooperatives participating in this proposal

Research Cooperative	Host University (year founded)	# Members
Cooperative Forest Genetics Research Program	**University of Florida** (1953)	8
Cooperative Tree Improvement Program	**North Carolina State University** (1955)	25
Forest Biology Research Cooperative	University of Florida (1996)	8
Forest Modeling Research Cooperative	Virginia Polytechnic Institute and State Univ. (1979)	21
Forest Nutrition Cooperative	Virginia Polytechnic Institute and State Univ. / NC State Univ. (1969)	43
Plantation Management Research Cooperative	University of Georgia (1975)	17
Southern Forest Resource Assessment Consortium	North Carolina State University (1994)	22
Western Gulf Forest Tree Improvement Program	**Texas A&M Univ. / Texas Forest Service** (1969)	13

Aim 3 – Genetics / Breeding

- Analyze genetics of breeding and natural populations
  - discover alleles in genes controlling important adaptation and mitigation traits
  - enable future tree breeding strategies
- Deliver deployment guidelines for genotypes suited for varied climatic conditions
  - maximize resiliency and reduce adverse impacts of climate change on productivity

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Aim 3 – Genetics / Breeding

- Analyze genetics of breeding and natural populations
  - discover alleles in genes controlling important adaptation and mitigation traits
  - enable future tree breeding strategies

- Deliver deployment guidelines for genotypes suited for varied climatic conditions
  - maximize resiliency and reduce adverse impacts of climate change on productivity

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

How do these relate to project goals?

- Understanding the genetic mechanisms that allow pines to adapt to different climate conditions will help accurately model the changes in productivity that can be expected under low-intensity management regimes
- Developing methods to account for genetic variation in productivity due to climate will guide breeding programs, leading to more robust planting stock for future generations of plantations
- Resilient and productive plantations will contribute to achieving USDA long-term goals for mitigation of and adaptation to climate change

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Outline of presentation

- An overview of project objectives , the genetics component, and the focus of this presentation
- Some background in quantitative genetics
- Understanding mechanism vs guiding breeding
  - Are these different objectives or one and the same?
- Experimental methods, materials & results
- Plans for year 3
- Challenges to be overcome

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

Quantitative genetics in 5 slides

- Phenotypic variation is the sum of genetic variation, environmental variation, and interactions:
  - P = G + E + GxE in shorthand form
- Genetic variation can be modeled as additive (each allele has a consistent detectable effect) or non-additive (the effect of an allele is not consistent)
- Traditional statistical models used in tree breeding focus on the additive genetic variation, because that is easiest to work with
- "All models are wrong, but some models are useful"
  - George Box

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Models of genetic architecture

- Across many species and traits, additive models work well to describe variation at the population level
  - Hill et al, Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 4:e1000008 (2008)
- Several studies of humans and model organisms indicate that non-additive effects are more important at the individual level
  - Huang et al, Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. PNAS 109:15553-9 (2012)
- The number of genes that affect a trait, and the size of effect each gene has, also make a difference in how we work toward our objectives

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Oligogenic vs Polygenic Models

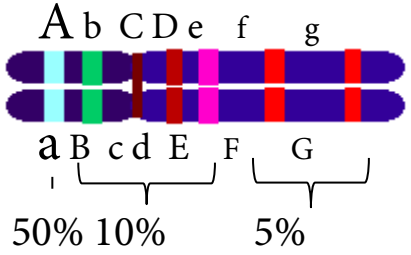
- "Oligogenic" – a relatively small number of genes each have fairly large effects on phenotypic variation, although there may be many more genes with small effects that cumulatively explain only a small proportion of the total variation
- "Polygenic" – hundreds of genes have effects on phenotypic variation, and all genes have roughly equal (and therefore very small) individual effects
- The classic model of quantitative inheritance (proposed by Fisher in 1918) is the infinitesimal model – infinitely many genes, each with an infinitely small effect. This is useful, but not biologically realistic.
  - In practice, the polygenic model is functionally equivalent to the infinitesimal model

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

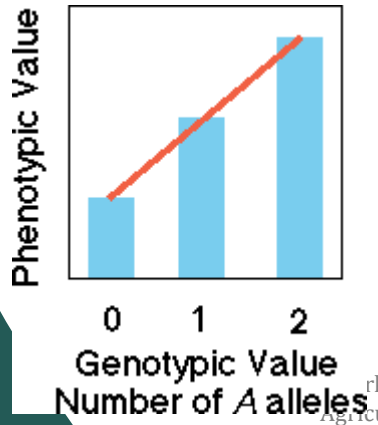
USDA United States Department of Agriculture National Institute of Food and Agriculture

Detecting the effect of a single gene is easier if effects are unequal

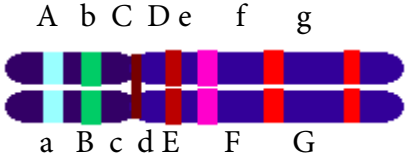
Oligogenic trait
Genes have unequal effects



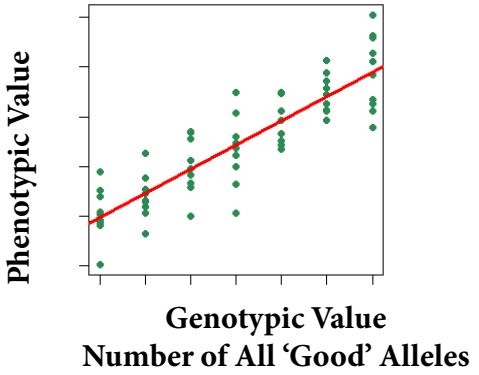
Individuals with a 'good' allele that accounts for 50% of genetic variation are easily detected in experimental populations



Polygenic trait
All genes have equal effects



If allelic effects are all equal, the best estimator of genetic value is the total number of 'good' alleles, rather than the presence of a specific allele



rk: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

How do we deal with polygenic traits?

- "Counting good alleles" is fine as an abstract concept, but it is not easy in practice
- Instead, we analyze genetic variation in populations of related individuals, and calculate the proportion of genes shared in common between pairs of individuals
- Individuals that share more genes in common are typically more similar in phenotype, if the phenotype is genetically determined
- Related individuals that share many alleles in common, and also share desirable phenotypes, are considered to have "many good alleles"
- "Best Linear Unbiased Prediction", or BLUP analysis

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Outline of presentation

- An overview of project objectives , the genetics component, and the focus of this presentation
- Some background in quantitative genetics
- Understanding mechanism vs guiding breeding
  - Are these different objectives or one and the same?
- Experimental methods, materials & results
- Plans for year 3
- Challenges to be overcome

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Two different objectives, or one?

Understanding mechanism

- Requires identifying the genes and regulatory elements that control the process of interest
- Can draw on data from model plants and crop species
- Goal: to integrate many levels of information into a comprehensive model

Guiding breeding

- Requires developing a model with predictive power
- Draws on data from ancestors and related individuals in breeding program
- Goal: to accurately identify individuals that will be good parents for breeding or deployment

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Long-term vs short-term outlook

- In the long term, a comprehensive model of how trees grow and develop in response to environmental signals would be a superb foundation for a breeding program
- In the short term, identifying genes that are significantly associated with phenotypic variation will not provide much guidance to breeding programs, unless those genes account for a significant proportion (>50%) of the observed genetic variation within current breeding populations
- For the present and near-term future, then, guidance for breeding programs and understanding of fundamental mechanisms should be considered different objectives

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Outline of presentation

- An overview of project objectives , the genetics component, and the focus of this presentation
- Some background in quantitative genetics
- Understanding mechanism vs guiding breeding
  - Are these different objectives or one and the same?
- Experimental methods, materials & results
- Plans for year 3
- Challenges to be overcome

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

Two analytical approaches

‘Association Genetics’

- Looks for statistical association of each genetic variant with phenotype
- Works well for oligogenic traits controlled by genes with large individual effects
- Leads to understanding of mechanism
- Often has high type II error (false negatives)

‘Kinship-based Analysis’

- Estimates the relative number of “good genes”, not their locations or identities
- Works equally well for polygenic or oligogenic traits
- Leads to predictive power – which individuals will be good parents for breeding
- Predictive power is often less than 100%

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Who cares about Type II errors?

- Testing individual genetic variants for statistical association with phenotypic variation requires thousands to millions of tests for each phenotype
- Multiple testing correction imposes a very high threshold of statistical significance to declare a positive association
- Most genes with small effects on phenotype don't meet that threshold, and are not detected in association studies.
- Human height example – 30,000 subjects genotyped at 2.5 million loci yielded 180 SNPs associated with height
  - All 180 loci combined explained less than 10% of variation
  - Height has heritability of 0.8 – a lot of genes are missed

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Two types of experimental population

‘Association Genetics’

- ADEPT2 population – range-wide sample of loblolly pine
- Clonally-replicated test planting in Florida
- Existing phenotypic data to be supplemented by additional data collection
- ~400 genotypes, not directly connected to breeding populations

“natural population”

‘Kinship-based Analysis’

- Progeny test series
- Based on trees from breeding programs, mated in structured experimental design
- Less diverse than range-wide sample, but more relevant to breeding populations
- >1000 genotypes/site, many sites available

“breeding population”

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

An overview of the pine genome

- ~2.2 x 10^{10} bp per haploid genome, or 7x larger than human or maize genomes
- Diploid – no evidence of polyploidy
- Contains a very diverse collection of transposable elements, retrotransposons, and other multi-copy sequences, but none accounts for > 5% of the genome
- Also contains an abundance of sequences similar to expressed genes, but most are thought to be non-functional
  - a sample of ~10^{6} bp of genomic sequence found 3 putative functional genes and 15 proposed pseudogenes (Kovach et al, BMC Genomics 11:420, 2010)

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

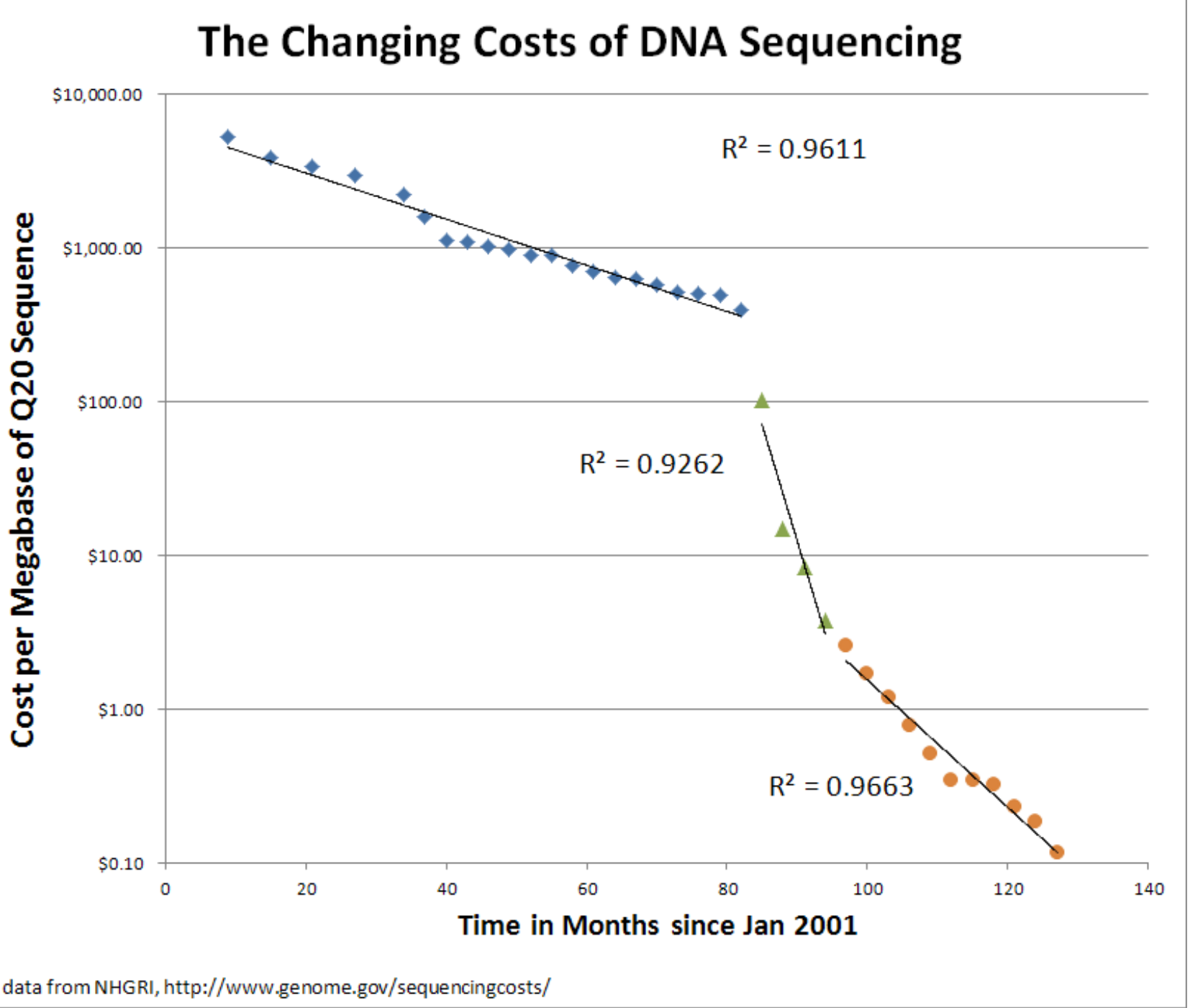
Detecting genetic variation in pine

- DNA sequencing is the most cost-effective way to obtain data on thousands of different genetic variants per individual
- The genome is too large to allow re-sequencing of the entire genome of many individuals at present
- Two methods of selectively recovering a reproducible subset of the genome for sequencing are being tested and compared
- It is possible that no single method exists that is ideal for every experimental purpose – the association genetics experiments to discover mechanism may require a different method than the kinship-based effort to guide breeding programs

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

PINEMAP



data from NHGRI, http://www.genome.gov/sequencingcosts/

Agriculture Project funded by the USDA National Institute of Food and Agriculture

Two genotyping methods

‘Hybrid Capture’

- Uses custom-synthesized DNA ‘bait’ molecules to capture target fragments
- ‘Bait’ sequences designed to capture fragments of expressed genes
- Focuses analysis of genetic variation on sequences in or near known expressed genes

‘Restriction-enzyme’

- Uses restriction enzymes to select a subset of the genome for sequencing
- Methylation-sensitive enzymes deliver fragments from less-methylated subset
- Lower cost per sample because fewer custom reagents are required

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

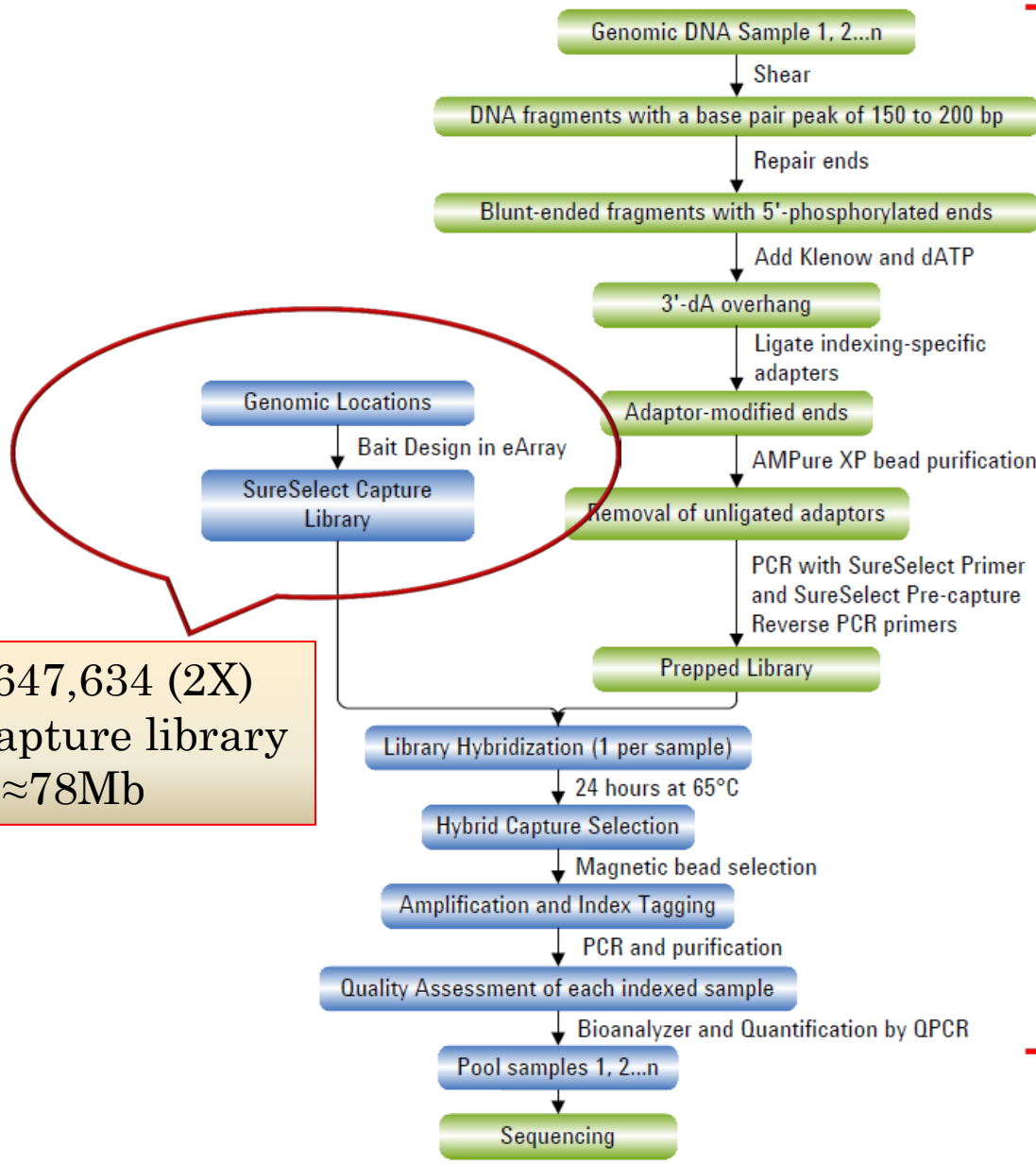


Figure 1 Overall sequencing sample preparation workflow.

#Baits = 647,634 (2X)
Expected capture library size ≈78Mb

Agilent Sure Select Target Enrichment

Slide from Mengmeng Lu
Texas A&M University

- Hybridization of the library
- Capturing the target sequences

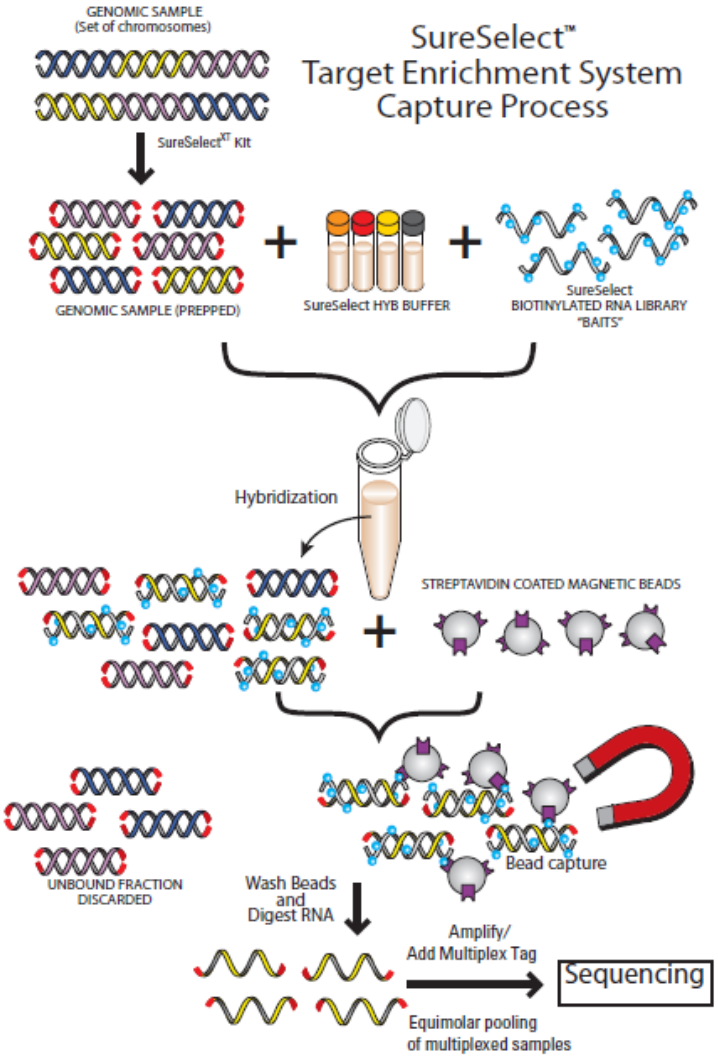


Figure 4 SureSelect Target Enrichment System Capture Process

Slide from Mengmeng Lu
Texas A&M University

Summary of the Sequencing Results

paired-end sequencing, 2×100 bp reads, Illumina Hiseq2000

Result:

The read counts (R1 and R2 have exactly the same number of reads):

Sample_pt20e_L008: 33,166,340 × 2 = 66,332,680 reads ≈ 6.7Gbp

Sample_pt20m_L008: 37,061,838 × 2= 74,123,676 reads ≈ 7.5Gbp

This provides 85-fold to 96-fold coverage of the 78 Mbp capture library

Slide from Mengmeng Lu
Texas A&M University

De novo assembly using megagametophyte reads by CLC Genomics workbench5.5.1

De novo assembly results:
184,907 contigs after scaffolding, total size is ≈58Mb

Baits were designed from 35,550 unigenes, total size about 42 Mb.

	Length
N75	220
N50	331
N25	580
Minimum	150
Maximum	18,824
Average	313
Count	184,907
Total	57,791,398

Slide from Mengmeng Lu
Texas A&M University

SNP detection

When minimum coverage=15, Variant probability=30.0%

Version: CLC Genomics Workbench 5.5.1
User: mlu
Parameters:

```
Minimum coverage = 15
Maximum expected variants = 2
Ignore quality scores = No
Ignore non-specific matches = No
Ignore broken pairs = Yes
Filter 454/Ion homopolymer indels = No
Variant probability = 30.0
Require presence in both forward and reverse reads = No
Genetic code = 1 Standard
Create track = No
Create annotated table = Yes
```

Comments: Edit

Found 68,993 variants

Originates from:

paired-eR1-57 (paired) mapping (history)

68,993 candidate SNPs

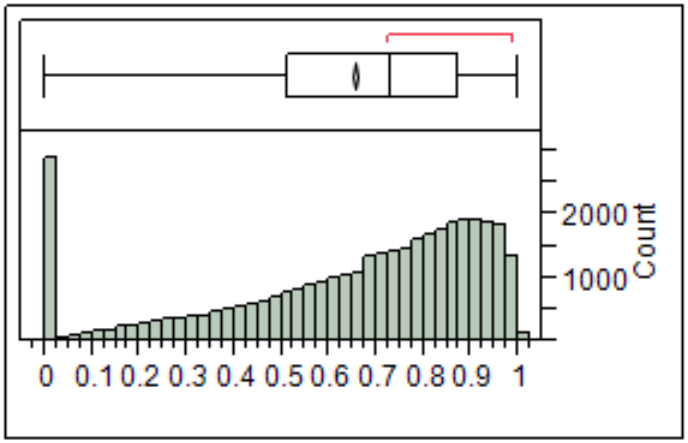
≈ 30.61% of the embryo reads are mapped

Slide from Mengmeng Lu
Texas A&M University

- Only 2674 unigenes are not covered by reads, < 8% percent of the total unigenes.

- Histogram of unigene mapping efficiency (distribution of proportion of unigene length mapped)

Consensus length/ Reference sequence



Quantiles

100.0%	maximum	1
99.5%		0.99621
97.5%		0.98249
90.0%		0.94455
75.0%	quartile	0.87234
50.0%	median	0.73303
25.0%	quartile	0.51238
10.0%		0.19871
2.5%		0
0.5%		0
0.0%	minimum	0

Moments

Mean	0.6545798
Std Dev	0.2802102
Std Err Mean	0.0014862
Upper 95% Mean	0.6574927
Lower 95% Mean	0.6516669
N	35550

Slide from Mengmeng Lu
Texas A&M University

- Relaxed mapping of unmapped reads to uncovered unigenes with the similarity threshold of 0.95 (using reads that were unmappped under similarity threshold of 0.99)

```
Parameters:
          Add conflict annotations = Yes
          Conflict resolution = Ambiguity nucleotides
          Create report = Yes
          Create list of un-mapped reads = Yes
          Mask reference sequences = No
          Match mode = random

input: paired-mR1-55 (paired) un-mapped reads [paired-mR1-55] (paired)
          Similarity = 0.95
          Length fraction = 1.0
          Insertion cost = 3
          Deletion cost = 3
          Mismatch cost = 1
          Color space alignment = No
          Global alignment = Yes
          Override paired distance = Yes
          Min distance = 97
          Max distance = 600

input: paired-mR1-55 (paired) un-mapped reads [single-mR1-59] (single)
          Similarity = 0.95
          Length fraction = 1.0
          Insertion cost = 3
          Deletion cost = 3
          Mismatch cost = 1
          Color space alignment = No
          Global alignment = Yes

input: paired-mR1-55 (paired) un-mapped reads [single-mR2-60] (single)
          Similarity = 0.95
          Length fraction = 1.0
          Insertion cost = 3
          Deletion cost = 3
          Mismatch cost = 1
          Color space alignment = No
          Global alignment = Yes

input: paired-mR1-55 (paired) un-mapped reads [paired-mR1-55] (single)
          Similarity = 0.95
          Length fraction = 1.0
          Insertion cost = 3
          Deletion cost = 3
          Mismatch cost = 1
          Color space alignment = No
          Global alignment = Yes
```

755 unigenes are still not covered.

Slide from Mengmeng Lu
Texas A&M University

BLAST of 755 uncovered unigenes against the 18 Gb draft loblolly pine genome assembly (e-value = e-5)

```
 91
 92
 93  Query=  contig2788
 94  Length=2394
 95
 96
 97  ***** No hits found *****
 98
 99
100
101  Lambda     K      H
102      1.33    0.621     1.12
103
104  Gapped
105  Lambda     K      H
106      1.28    0.460    0.850
107
108  Effective search space used: 41103188044632
109
110
111  Query=  contig3220
112  Length=2308
113
114
115  ***** No hits found *****
116
117
```

168 unigenes (≥120bp) (≈ 0.4%) have no hits

Slide from Mengmeng Lu
Texas A&M University

Two genotyping methods

‘Hybrid Capture’

- Uses custom-synthesized DNA ‘bait’ molecules to capture target fragments
- ‘Bait’ sequences designed to capture fragments of expressed genes
- Focuses analysis of genetic variation on sequences in or near known expressed genes

‘Restriction-enzyme’

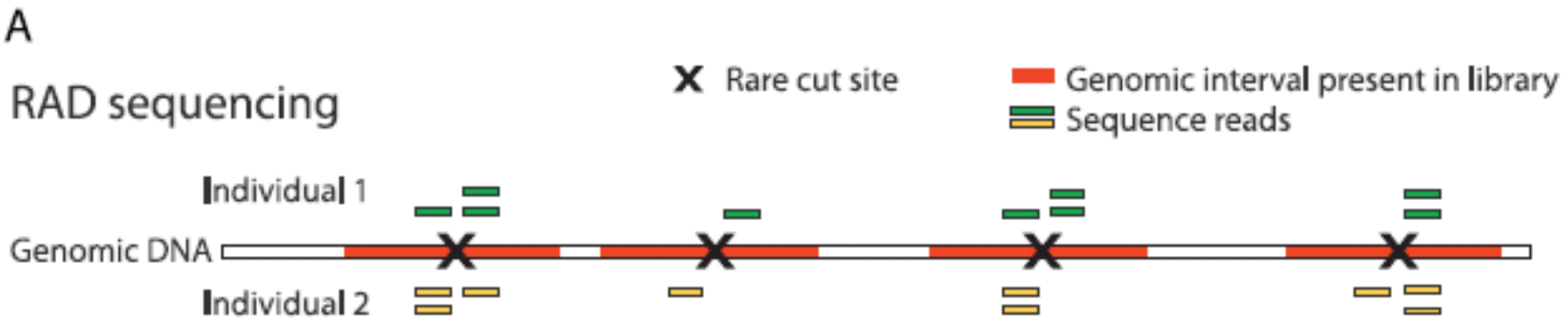
- Uses restriction enzymes to select a subset of the genome for sequencing
- Methylation-sensitive enzymes deliver fragments from less-methylated subset
- Lower cost per sample because fewer custom reagents are required

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Two alternative approaches tested

- Digest pine genomic DNA with a methylation-sensitive restriction enzyme to enrich for single- to low-copy genomic regions
- Attach "barcode sequences" to identify each sample



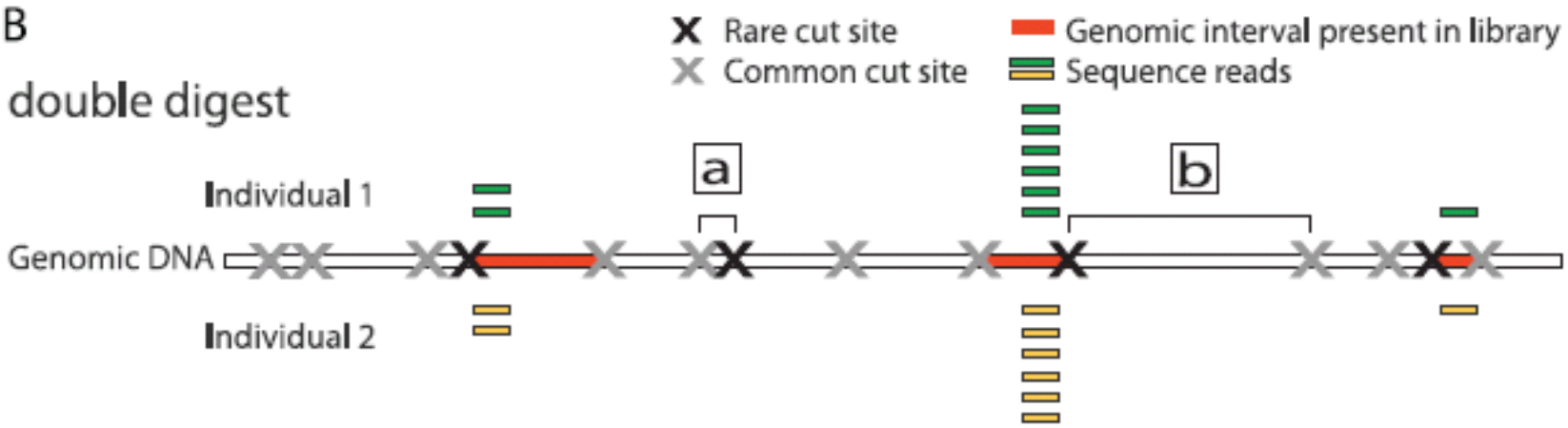
- Pool all samples and randomly shear the DNA to create fragments of appropriate size for Illumina sequencing

Image from Peterson, et al., PLoS ONE 7(5): e37135, 2012
doi:10.1371/journal.pone.0037135

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

Two alternative approaches tested

- Digest pine genomic DNA with TWO methylation-sensitive restriction enzymes to enrich for single- to low-copy genomic regions
- Attach "barcode sequences" to identify each sample



- Pool all samples and select fragments of appropriate size for Illumina sequencing

Image from Peterson, et al., PLoS ONE 7(5): e37135, 2012
doi:10.1371/journal.pone.0037135

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Two genotyping methods

‘Hybrid Capture’

- Uses custom-synthesized DNA ‘bait’ molecules to capture target fragments
- ‘Bait’ sequences designed to capture fragments of expressed genes
- Focuses analysis of genetic variation on sequences in or near known expressed genes

‘Restriction-enzyme’

- Uses restriction enzymes to select a subset of the genome for sequencing
- Methylation-sensitive enzymes deliver fragments from less-methylated subset
- Lower cost per sample because fewer custom reagents are required

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Summary of SureSelect Results

- Bait sequences were synthesized based on DNA sequences of 35,502 expressed genes
- Genomic sequences similar to 92% - 98% of those target genes were recovered by hybrid-capture
- DNA sequence variation in the captured fragments is 0.9%, comparable to previously-reported values for pine, yielding 51,623 candidate SNPs after filtering
- ~30% of the sequences of captured fragments map to target genes; ~70% are non-target sequences
- Over half of the target genes have > 70% coverage
- This method is an efficient way to re-sequence the coding regions of genes

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Summary of RAD-seq results

- Two diploid DNA samples and 10 haploid samples of progeny of one parent were used for sequencing
- RAD sequencing (using random shearing) yielded a large proportion of sequences that contain artifacts from library preparation
- This result is consistent with previous results of others – the RAD-seq technique apparently requires extensive optimization to be useful
- Decision – focus further attention on optimizing the "double digest" and hybrid capture methods

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

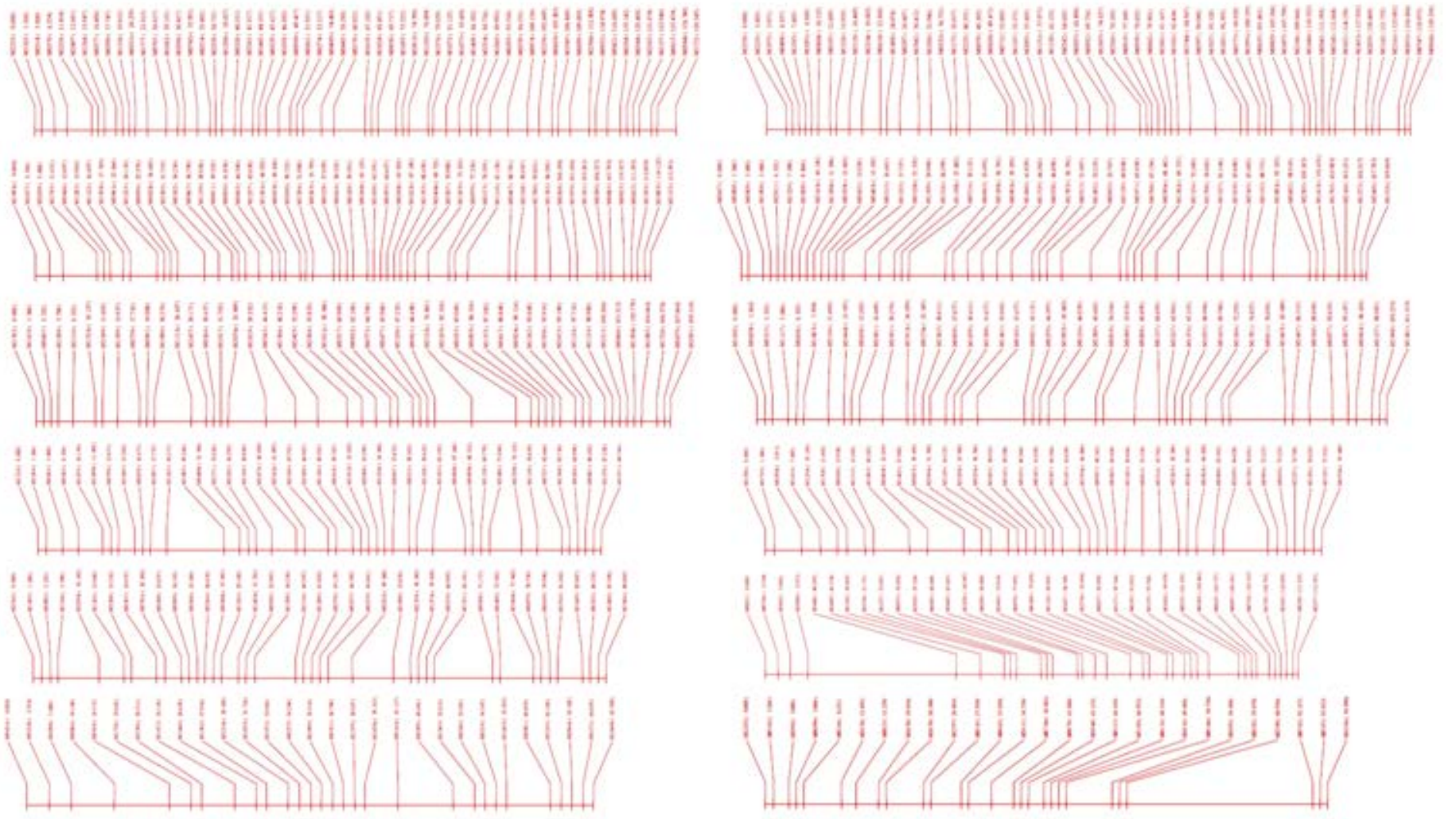
Results from double digest GBS

- Two diploid DNA samples and 90 haploid samples from seeds of one of the diploids were sequenced
- Filter to identify sequences present in about half the samples: 47,131 sequences meet that criterion
- Map those sequences to the pine reference genome draft assembly: 32,809 (70%) sequences align to the genome
- About half (16,236) of those sequences map to a single location in the draft reference genome
- Over 95% (15,494) of the single-copy sequences align to the reference sequence without gaps
- 8,073 of the single-copy sequences align without gaps or mismatches; 7448 align with 1 – 4 candidate SNPs

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

A linkage map of GBS markers



Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

GBS markers and functional genes

- 8199 of 47,131 sequences (17.4%) align to sequences in the draft pine reference transcriptome assembly
- 4962 sequences in the transcriptome assembly are identified by GBS markers
- About 95% of the alignments between GBS markers and transcriptome sequences are ungapped
- 62 different map locations on the preliminary linkage map correspond to GBS markers that align to transcriptome sequences
- Only 375 of the 47,131 candidate markers are derived from chloroplast DNA

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Outline of presentation

- An overview of project objectives , the genetics component, and the focus of this presentation
- Some background in quantitative genetics
- Understanding mechanism vs guiding breeding
  - Are these different objectives or one and the same?
- Experimental methods, materials & results
- Plans for year 3
- Challenges to be overcome

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Plans for year 3 experiments

- Reducing the per-sample cost of the hybrid-capture approach requires innovative methods for combining samples – experiments are underway
- Genotypes of 397 trees sampled from the range-wide distribution of loblolly pine (the "ADEPT2 population") will be obtained using the optimized hybrid-capture protocol
- The same 397 trees will be genotyped using the double-digest GBS procedure
- This will provide SNP markers both within expressed genes and in regions throughout the genome for association analysis with existing or new phenotypic data

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Plans for year 3 experiments

- Kinship-based methods will be tested on progeny sampled from field tests of breeding program
- The double-digest GBS procedure will be used to genotype samples of 960 to 1536 trees per site, for two or more sites with different climate characteristics
- The genotypes will be included in kinship-based analysis of family and individual-tree performance as a function of site characteristics
- Presence of SNP markers within expressed genes is not an advantage for this approach, so the lower cost and higher throughput of the double-digest procedure are expected to be important advantages

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Plans for year 3 experiments (cont'd)

Questions to be answered:

- How many markers are needed to increase the predictive accuracy of kinship-based methods?
- How great an increase in predictive accuracy can be obtained by including marker genotypes in the kinship analysis?
- What is the most cost-effective approach for genotyping many hundreds of progeny that are all descended from <100 parents?
- Can genes discovered by association genetics be integrated into the kinship-based analytical approach to gain the benefit of both approaches for breeding?

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Outline of presentation

- An overview of project objectives , the genetics component, and the focus of this presentation
- Some background in quantitative genetics
- Understanding mechanism vs guiding breeding
  - Are these different objectives or one and the same?
- Experimental methods, materials & results
- Plans for year 3
- Challenges to be overcome

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

Challenges to be overcome

- Dealing with a tsunami of genotyping data
  - Storing and analyzing terabytes of data
  - Separating signal from noise in analysis
  - Recognizing the appropriate domains for application of different data types
- Collecting appropriate phenotypic data in breeding programs
  - Planting progeny tests in areas outside the current zone of adaptation to obtain data on variation in resilience
  - Collecting data on additional phenotypes?
- Communicating results clearly to stakeholders who have to make their own risk-versus-reward decisions

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture

Acknowledgments

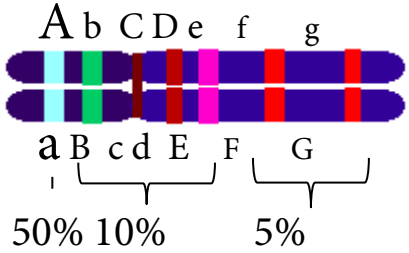
- University of Florida – Gary Peter (investigator), Jianxing Zhang (Ph.D student)
- North Carolina State University – Ross Whetten, Fikret Isik, Steve McKeand (investigators), Alfredo Farjat (Ph.D student), Laura Townsend (M.S. student), Will Kohlway and Ben Rusche (undergrads)
- Texas A&M University – Tom Byram, Carol Loopstra, Kostya Krutovsky (investigators), Tomasz Koralewski (postdoc), Mengmeng Lu (Ph.D student)
- Virginia Tech – Jason Holliday (investigator), Rajesh Bawa (Ph.D student)
- Dana Nelson – US Forest Service (investigator)

Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

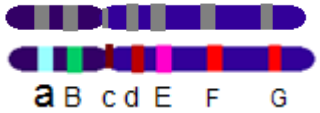
USDA United States Department of Agriculture National Institute of Food and Agriculture

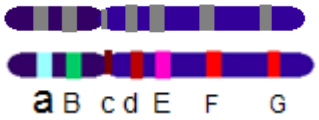
Detecting the average effect of an allele is easier if effects are unequal

Alleles have unequal effects



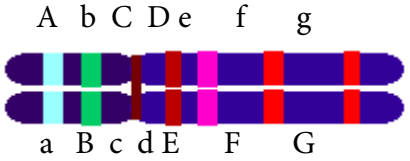
Progeny with a 'good' allele that accounts for 50% of genetic variation are easily detected in experimental populations

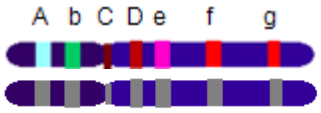


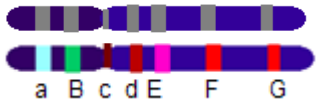


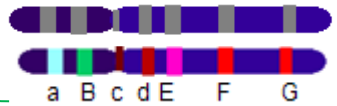
If allelic effects are all equal, the best estimator of genetic value is the total number of 'good' alleles, rather than the presence of a specific allele

All alleles have equal effects









Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) is a Coordinated Agriculture Project funded by the USDA National Institute of Food and Agriculture

USDA United States Department of Agriculture National Institute of Food and Agriculture