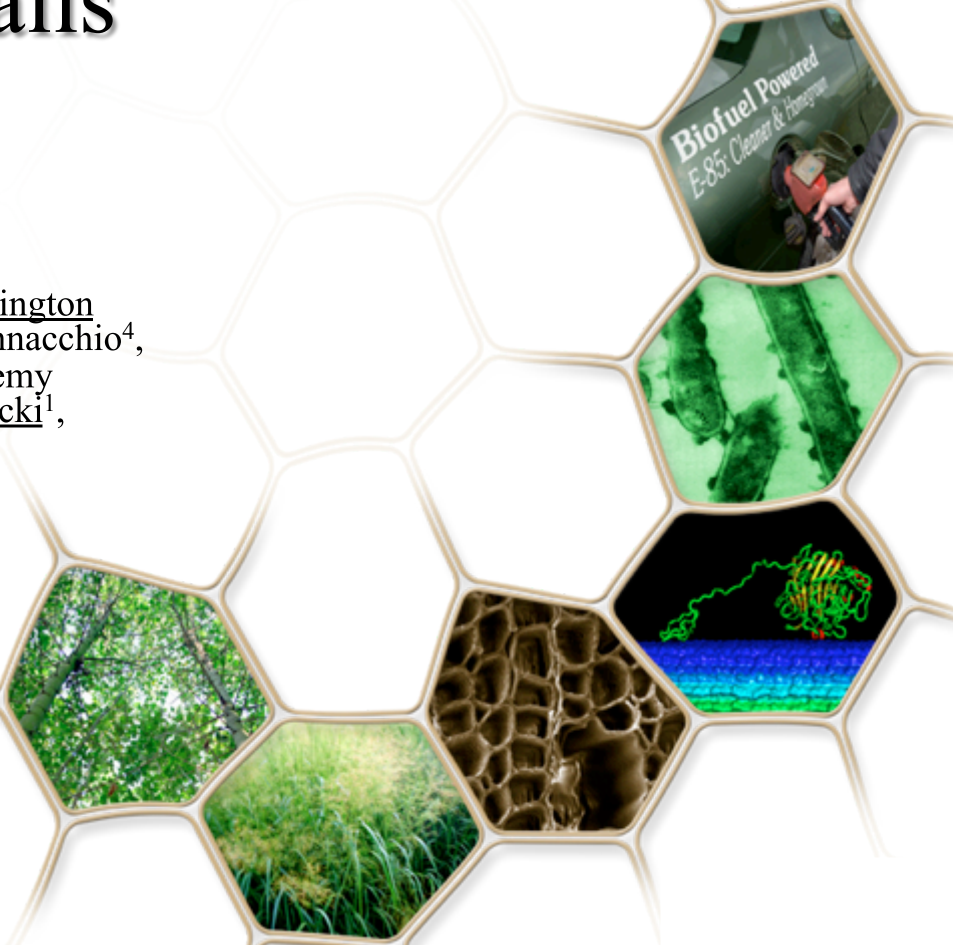
New Technologies for New Markets:

What is Modern Genetics Telling Us About Cell Walls

Gerald Tuskan[1,4]

Paul Abraham[1,2], Rachel Adams[1,2], Priya Ranjan[1], Wellington Muchero[1], Stephen DiFazio[3], Gancho Slavov[3], Len Pennacchio[4], Wendy Schackwitz[4], Joel Martin[4], , Jin-Gui Chen[1], Jeremy Schmutz[5], Dan Rokshar[4], Robert Hettich[1], Olaf Czarnecki[1], Jianjun Guo[1]

1 Oak Ridge National Laboratory, Oak Ridge, TN
2 Genome Science and Technology, University of Tennessee, Knoxville, TN
3 West Virginia University, Morgantown, WV
4 Joint Genome Institute, Walnut Creek, CA
5 HudsonAlpha, Huntsville, AL

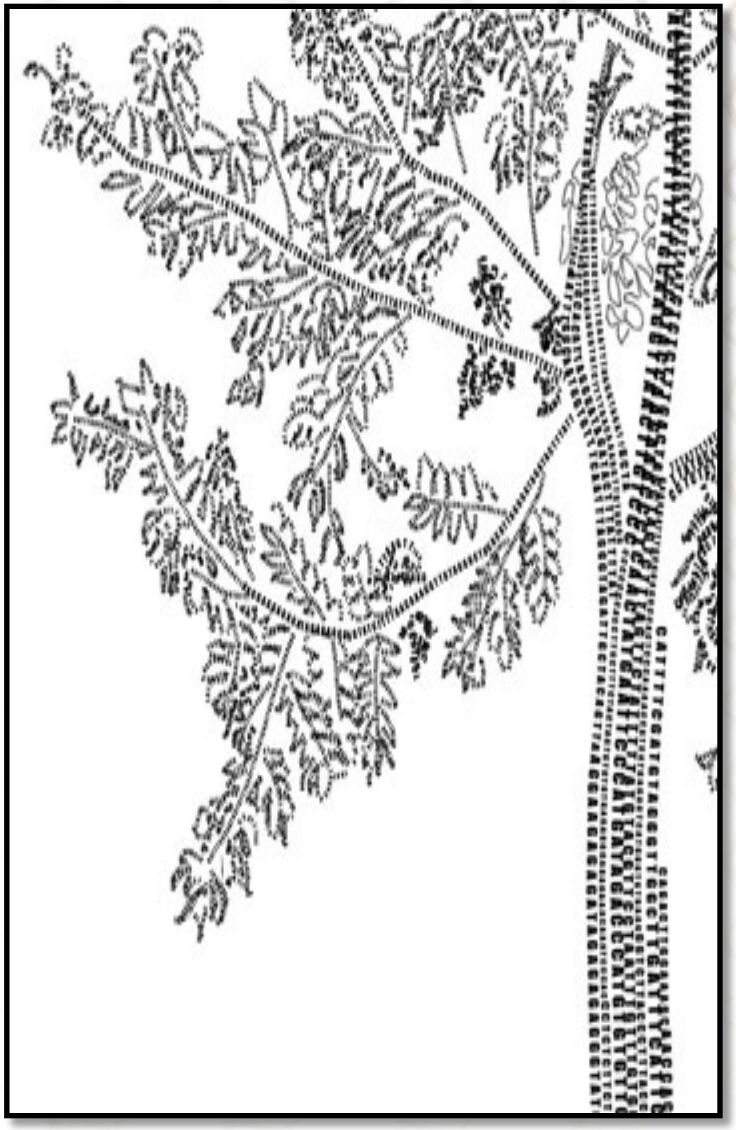


Scope of the Talk

I. Association Mapping
  - SNP detection and distribution
  - Genome-wide Association Genetics

II. Activation Tagging
  - TAIL PCR Identification
  - Arabidopsis Confirmation and Recapitulation

III. Stable Transformation
  - Projected Transgenics
  - Early Phenotypes

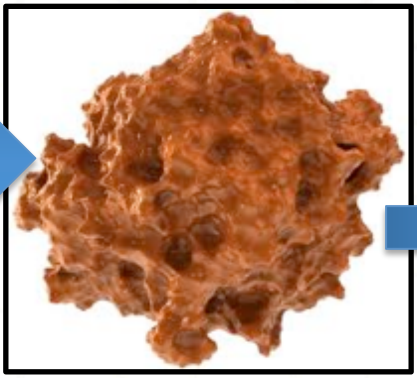


Lignocellulose by any other name . . .

Cell Wall Deconstruction and Product Output

Dedicated Feedstock
Wood Chips

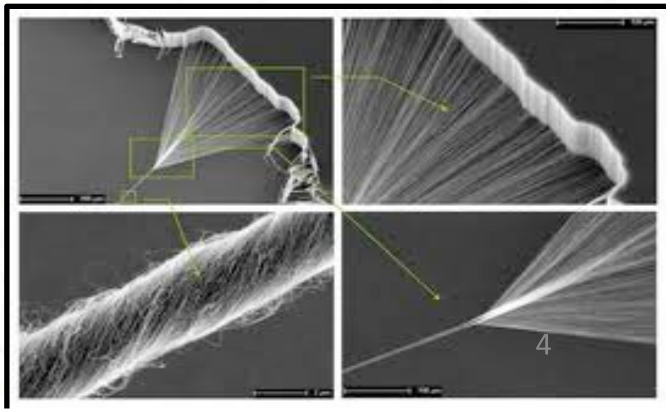
Cellulose



Lignin

Ethanol,
Isobutanol,
etc.

Spun
Carbon
Fibers

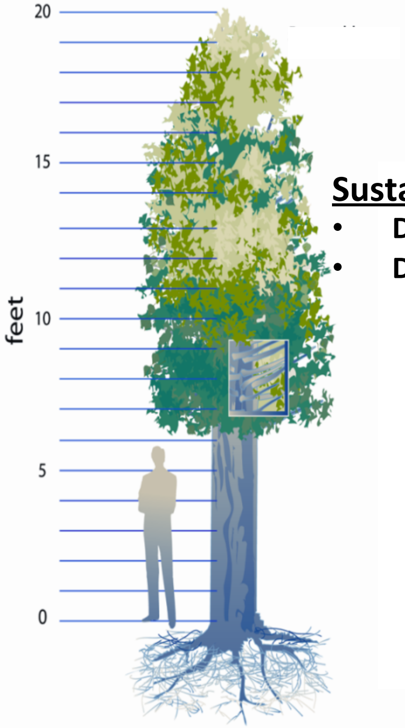


Populus as a Model for Cell Wall Formation

Enhancing Traits for Energy Applications

Capture and Allocation of Carbon

- Increased photosynthetic efficiency
- Optimized crown and leaf architecture

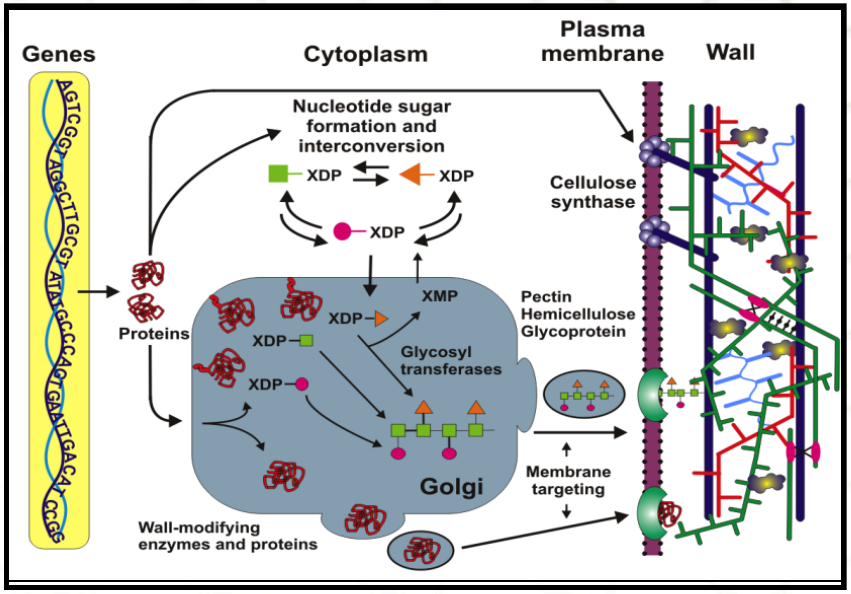


Biomass

- Tailored biomass
- Enhanced production

Sustainability

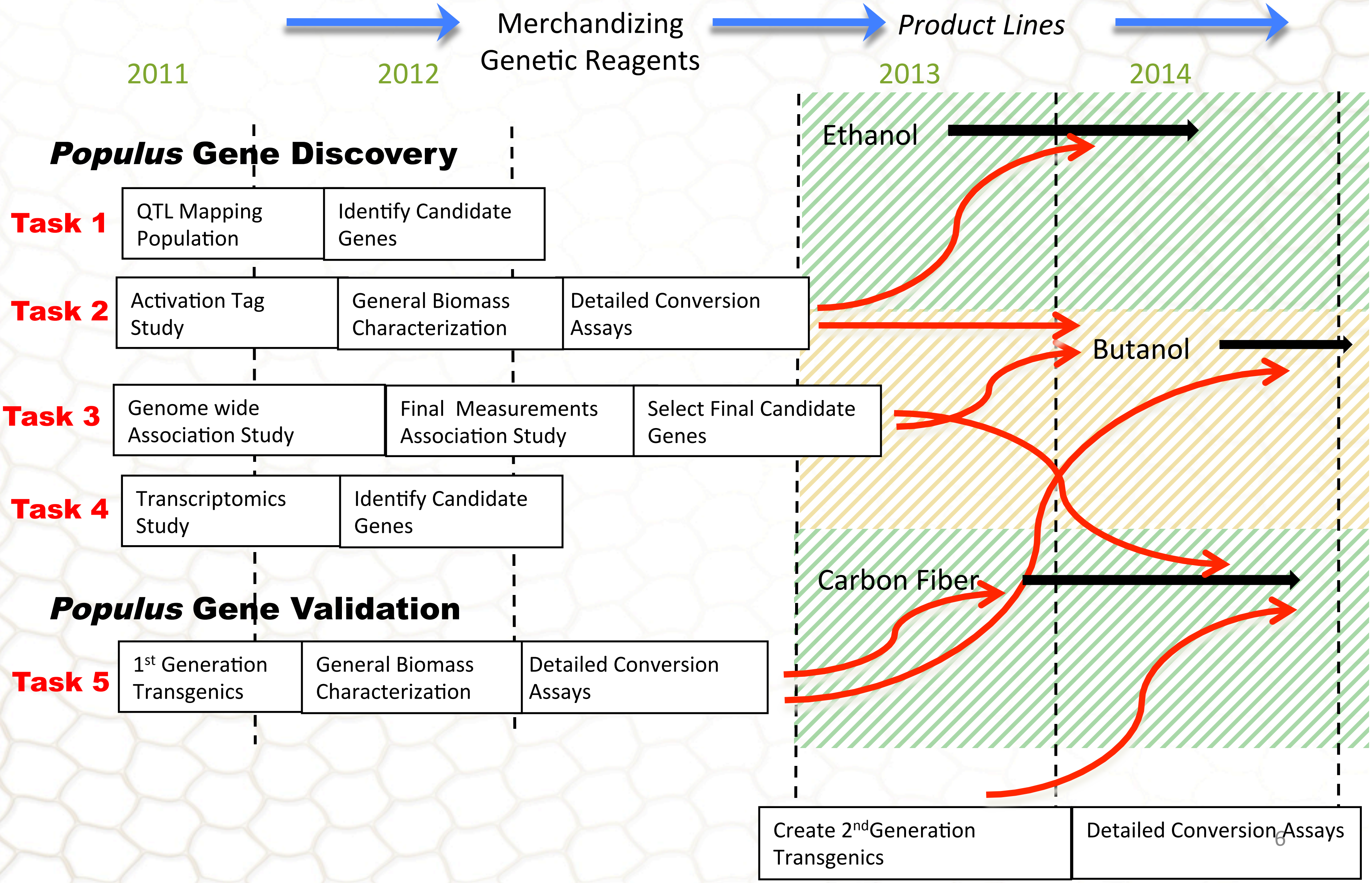
- Disease resistance
- Drought tolerance



Goal: To reduce the recalcitrance of lignocellulosic biomass for economic production of biofuels

Focus: To understanding basic mechanisms of cell wall formation in *Populus*

De Novo Gene Discovery – BioEnergy Science Center



Merchandizing Genetic Reagents

Product Lines

2011

2012

2013

2014

Populus Gene Discovery

Task 1
- QTL Mapping Population
- Identify Candidate Genes

Task 2
- Activation Tag Study
- General Biomass Characterization
- Detailed Conversion Assays

Task 3
- Genome wide Association Study
- Final Measurements Association Study
- Select Final Candidate Genes

Task 4
- Transcriptomics Study
- Identify Candidate Genes

Populus Gene Validation

Task 5
- 1st Generation Transgenics
- General Biomass Characterization
- Detailed Conversion Assays

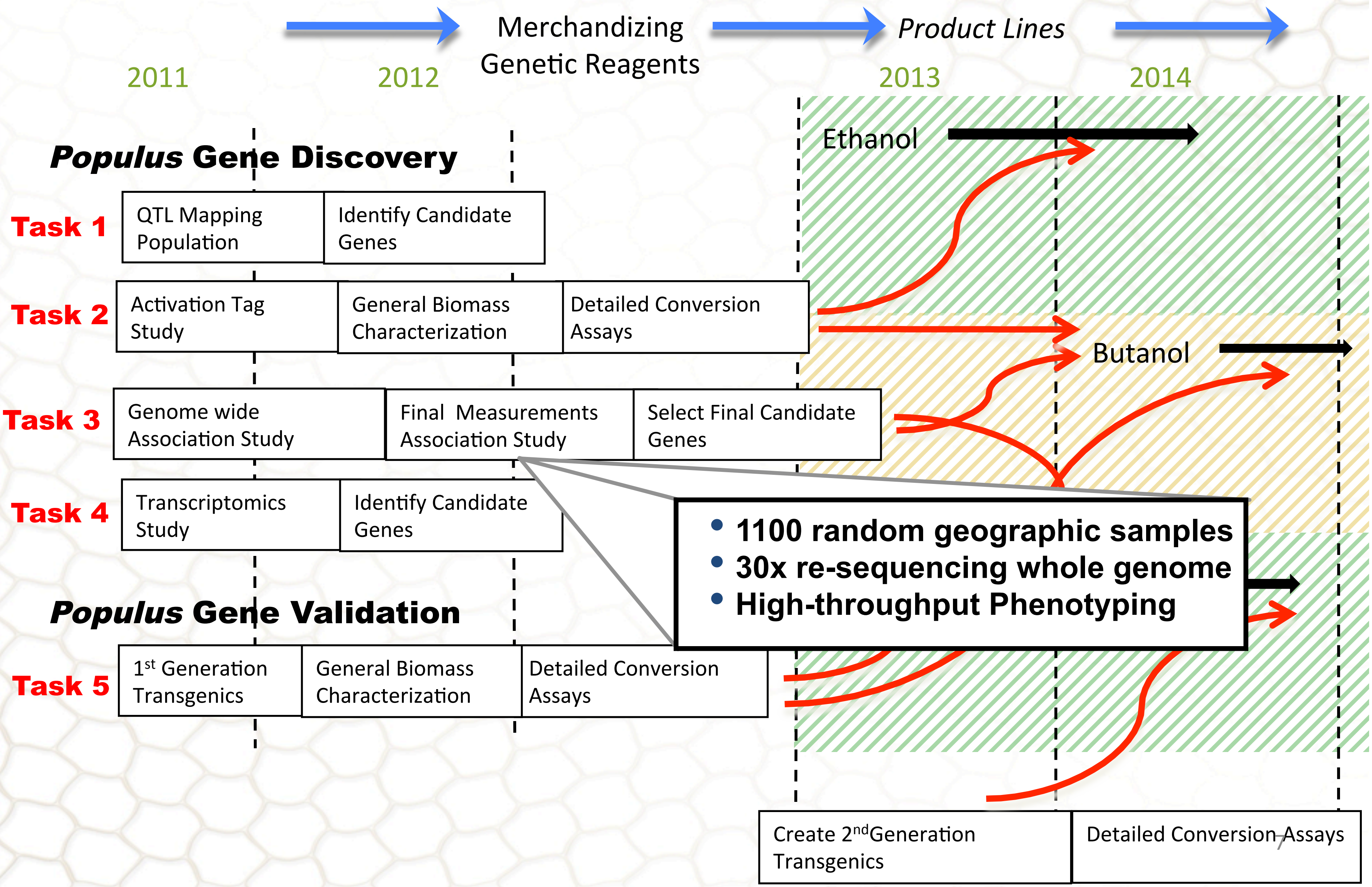
- Create 2nd Generation Transgenics
- Detailed Conversion Assays

Ethanol

Butanol

Carbon Fiber

De Novo Gene Discovery – Association Genetics



Merchandizing Genetic Reagents

Product Lines

2011

2012

2013

2014

Ethanol

Butanol

***Populus* Gene Discovery**

Task 1
- QTL Mapping Population
- Identify Candidate Genes

Task 2
- Activation Tag Study
- General Biomass Characterization
- Detailed Conversion Assays

Task 3
- Genome wide Association Study
- Final Measurements Association Study
- Select Final Candidate Genes

Task 4
- Transcriptomics Study
- Identify Candidate Genes

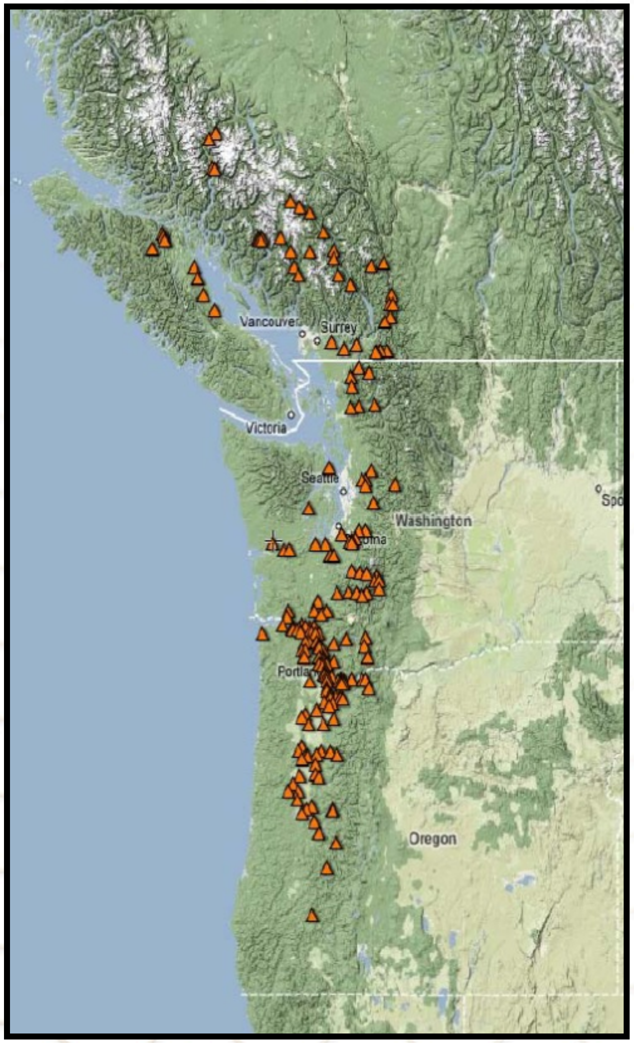
- **1100 random geographic samples**
- **30x re-sequencing whole genome**
- **High-throughput Phenotyping**

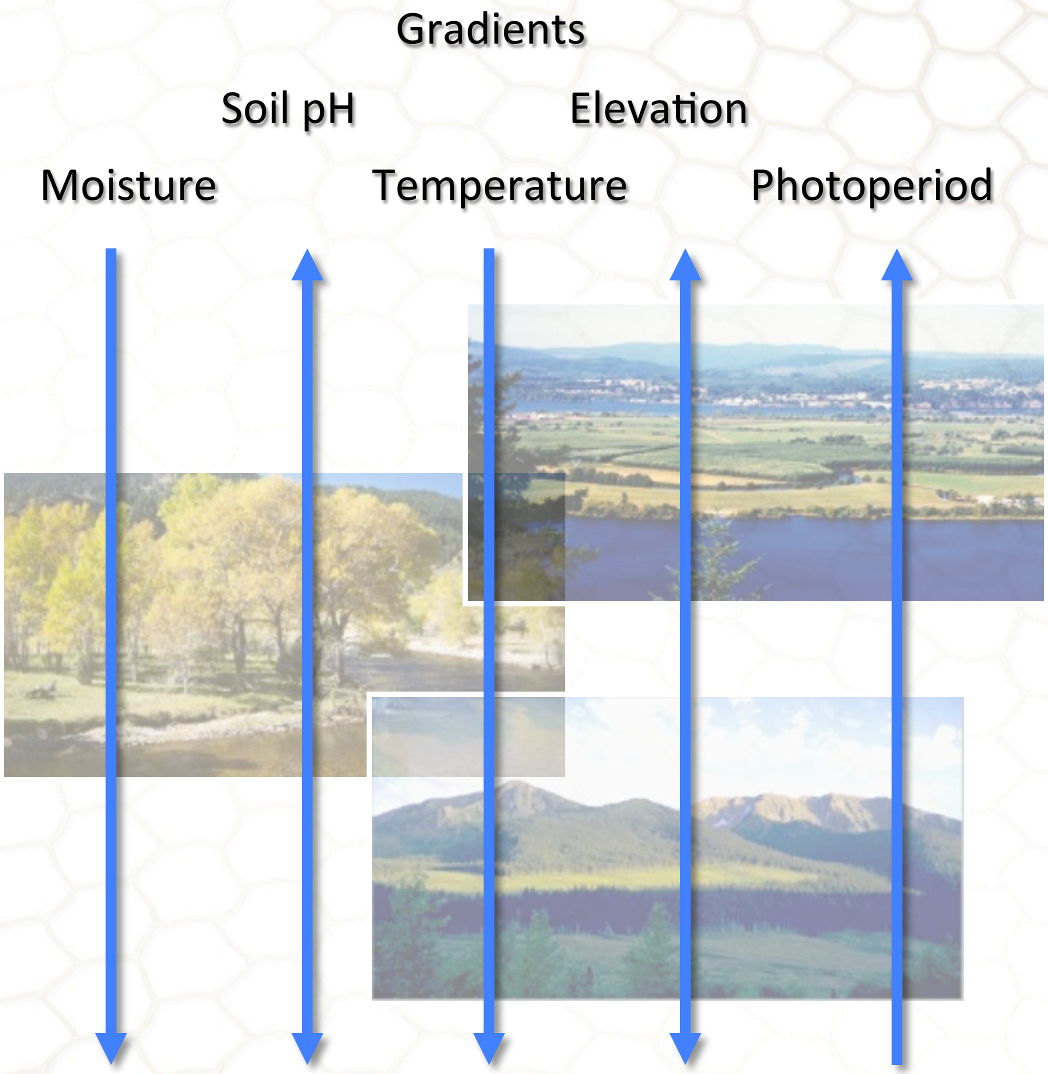
***Populus* Gene Validation**

Task 5
- 1st Generation Transgenics
- General Biomass Characterization
- Detailed Conversion Assays

- Create 2nd Generation Transgenics
- Detailed Conversion Assays

GWAS -- Mining Adaptive Allelic Variation





Gradients

Soil pH

Elevation

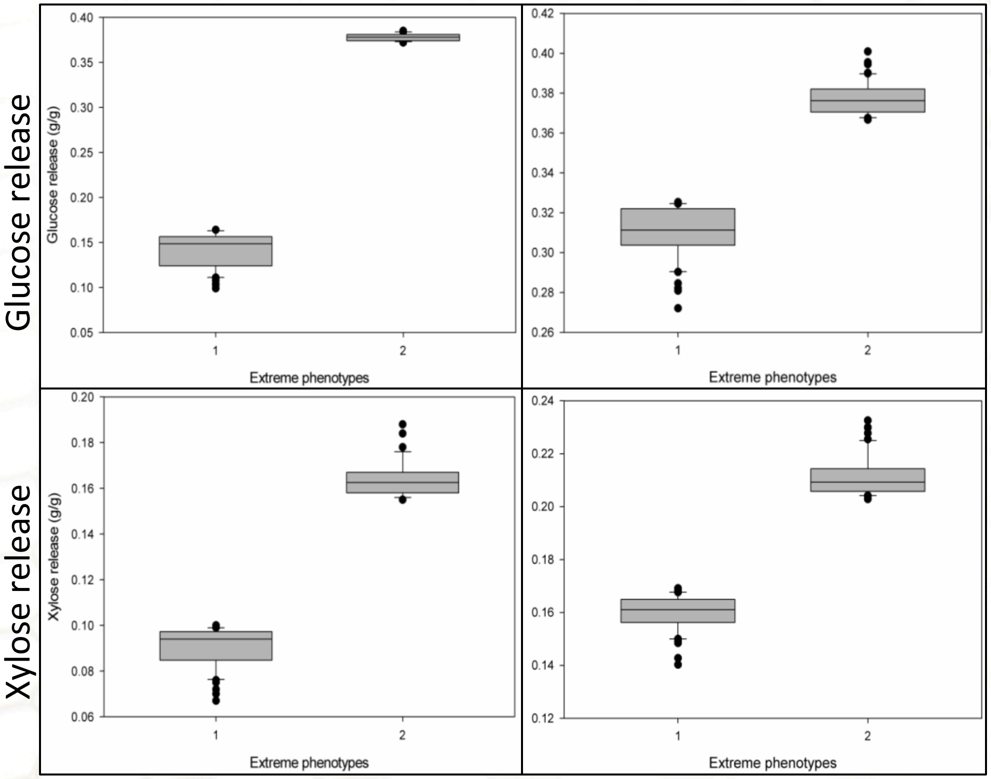
Moisture

Temperature

Photoperiod

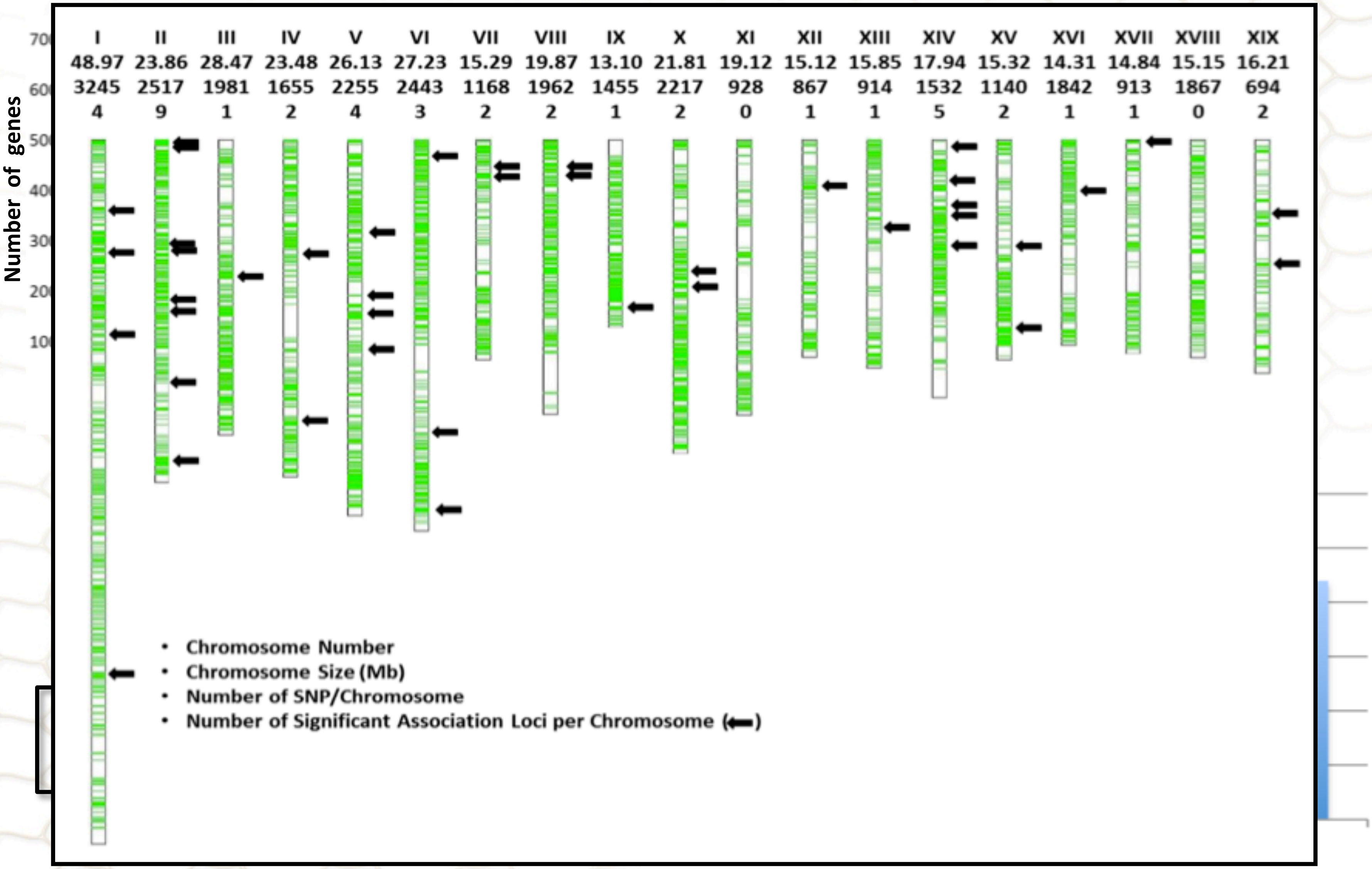
Natural Variation in Sugar Release

Native environment   Corvallis, OR field site



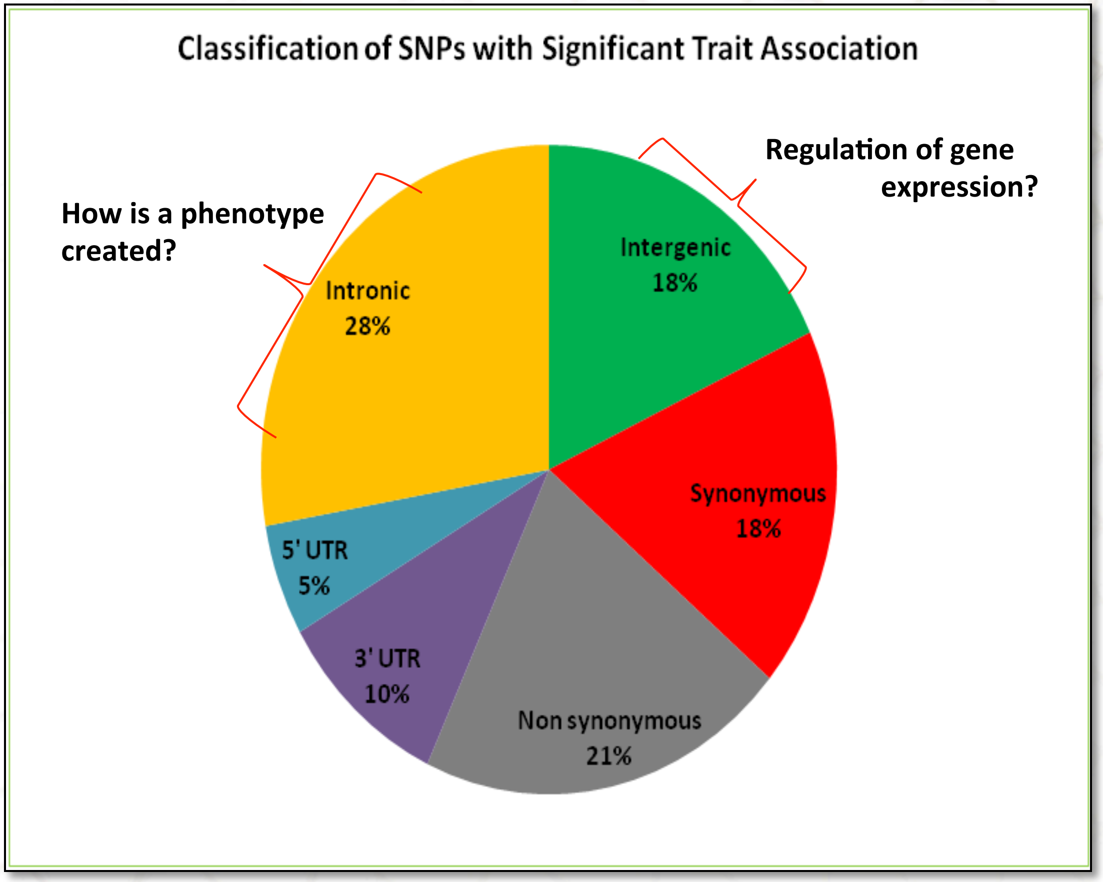
Phenotypic extremes

GWAS Analysis -- SNP distribution by genic region



Association Mapping: Results

- **42** SNPs had significant associations with cell wall chemistry and/or sugar release
- **6** were intergenic
- **34** SNPs mapped in **15** genes **– 18** were non-synonymous
- Of the 15 genes, **7** were **transcription factors**

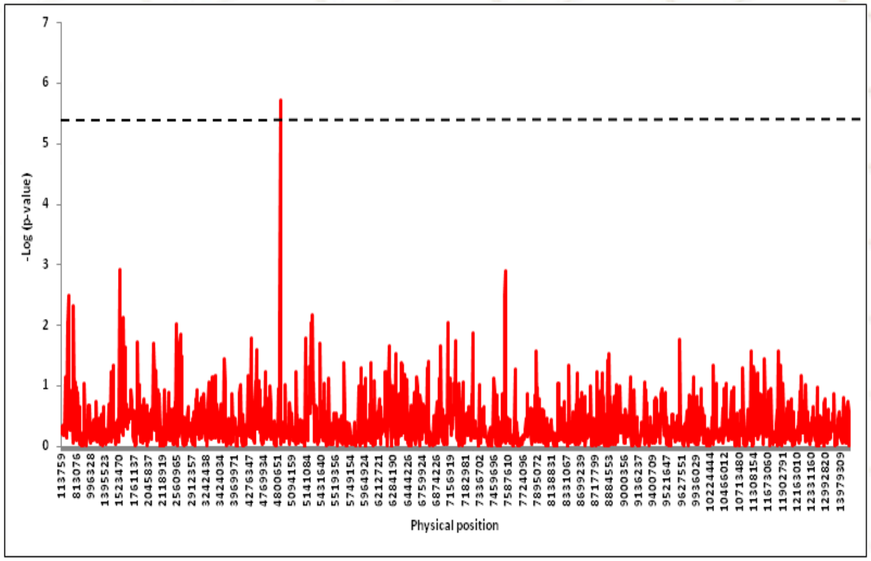

Classification of SNPs with Significant Trait Association

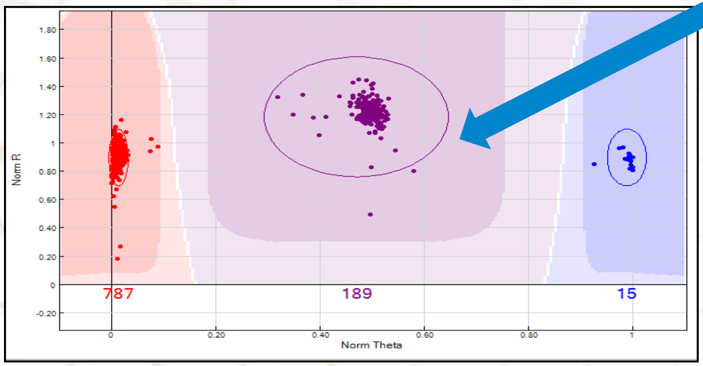
How is a phenotype created?

Regulation of gene expression?

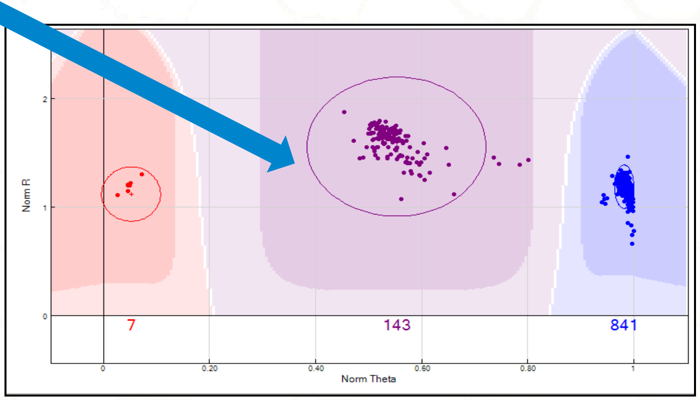
Association Mapping: Results

- *Populus* has 2 copies of a CW Synthase gene as a result of a whole-genome duplication event
- The two paralogs are highly identical (90% similarity at the sequence level)
- Preliminary evidence suggests that paralogs cannot complement each other
- Rare natural variants exhibit profound metabolite changes



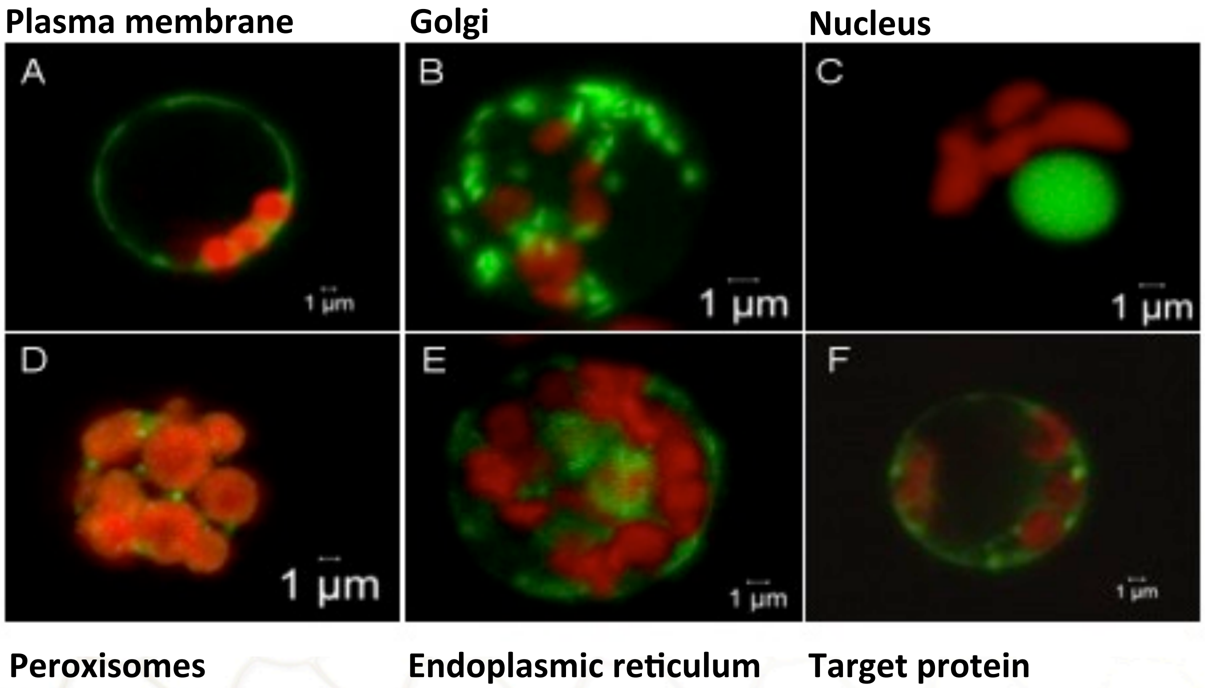


Scaffold_14_CW Synthase



Scaffold_2_CW Synthase

Transient Expression – Protoplast Reporter Genes



<u>Lignin</u>
-- 6 reporter genes
<u>Cellulose</u>
-- 5 reporter genes
<u>Hemicellulose</u>
-- 3 reporter genes

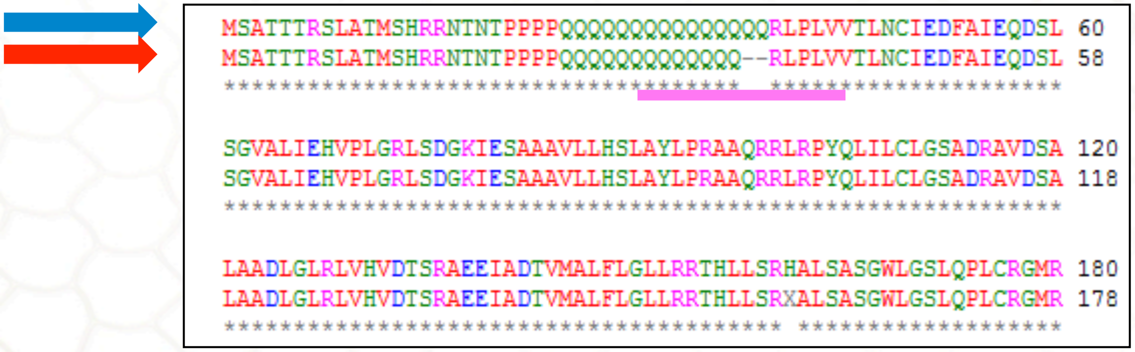
Transient transformation + GFP (or RT PCR)
= Initial verification of gene effect on targeted phenotype

Jay Chen, Jim Guo

Association Mapping: Transcription Factor CtBP

- cDNA cloning and sequencing revealed a **polyglutamine (PolyQ)** length polymorphism
- **PolyQ polymorphisms have been associated with numerous human degenerative disease**

Alternate allele amino acid alignment

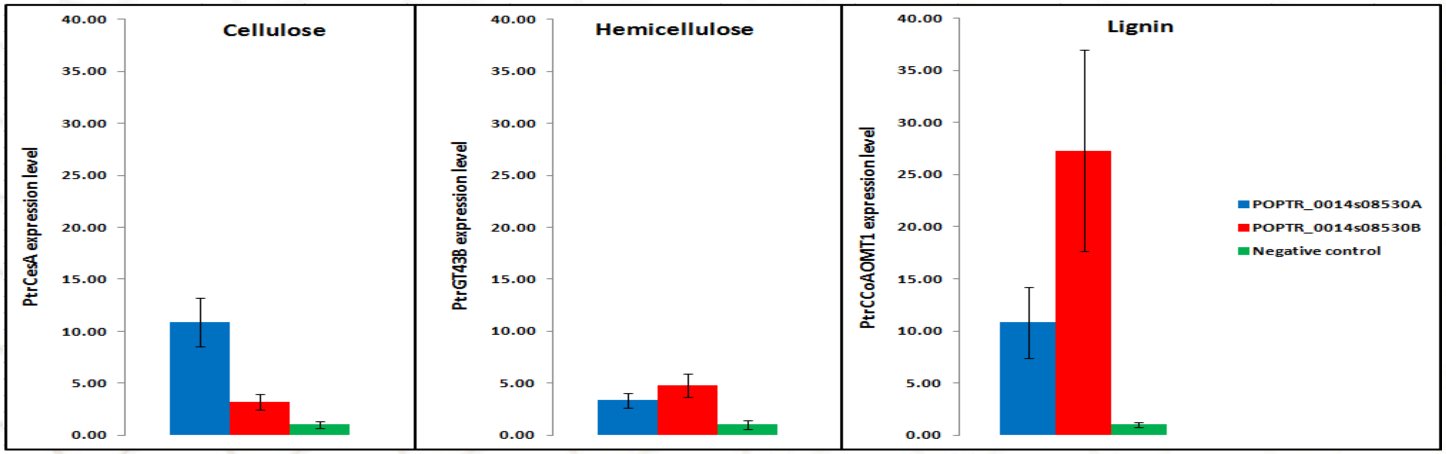


```
MSATTTRSLATMSHRRNTNTPPPPQQQQQQQQQQQQQQQRLPLVVTLNCIEDFAIEQDSL 60
MSATTTRSLATMSHRRNTNTPPPPQQQQQQQQQQQQQ--RLPLVVTLNCIEDFAIEQDSL 58
*************************************  *********************

SGVALIEHVPLGRLSDGKIESAAAVLLHSLAYLPRAAQRRLRPYQLILCLGSADRAVDSA 120
SGVALIEHVPLGRLSDGKIESAAAVLLHSLAYLPRAAQRRLRPYQLILCLGSADRAVDSA 118
************************************************************

LAADLGLRLVHVDTSRAEEIADTVMALFLGLLRRTHLLSRHALSASGWLGSLQPLCRGMR 180
LAADLGLRLVHVDTSRAEEIADTVMALFLGLLRRTHLLSRXALSASGWLGSLQPLCRGMR 178
**************************************** *******************
```

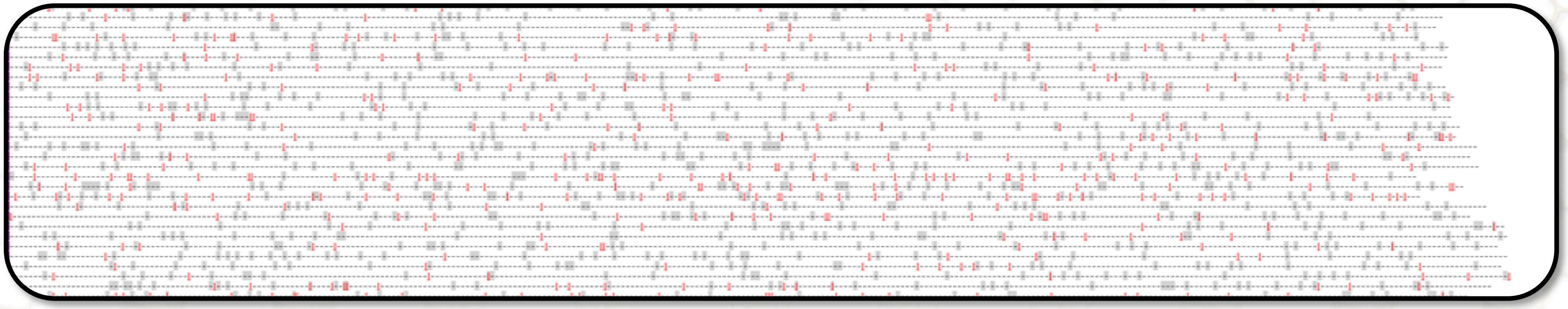
Side-by-side transfection of *Populus* protoplasts using alternate alleles

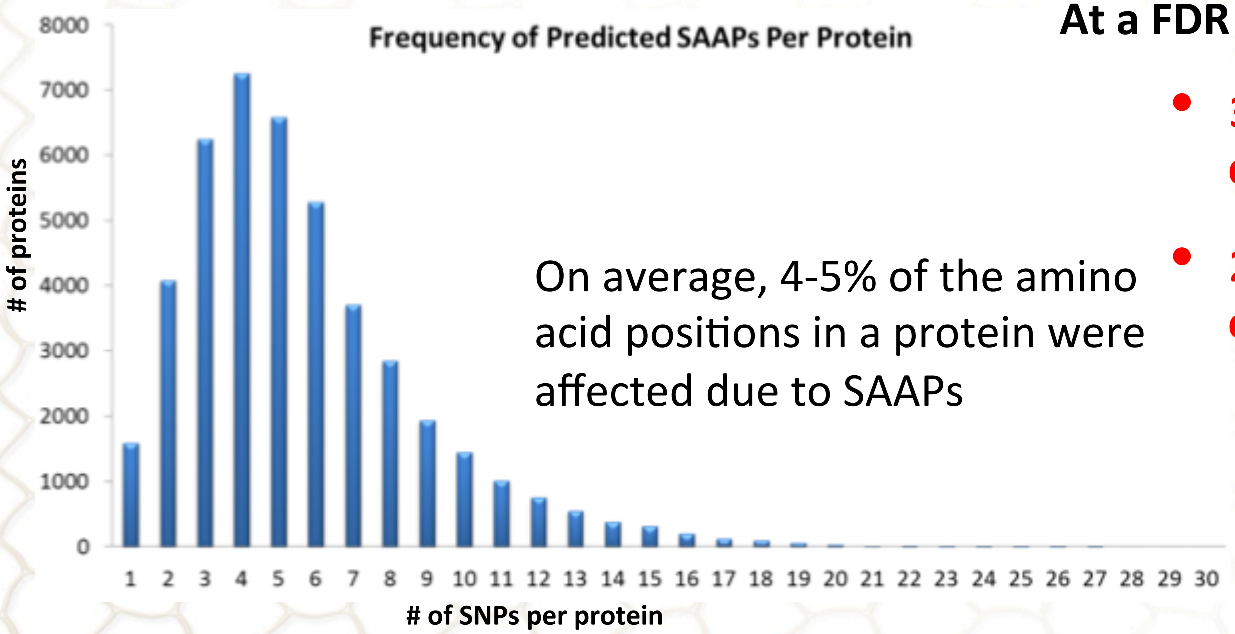


Association Mapping: Single Amino Acid Polymorphisms

Database includes:

- Over 400,000 SAAP positions
- Over 7.2 million SAAP peptides





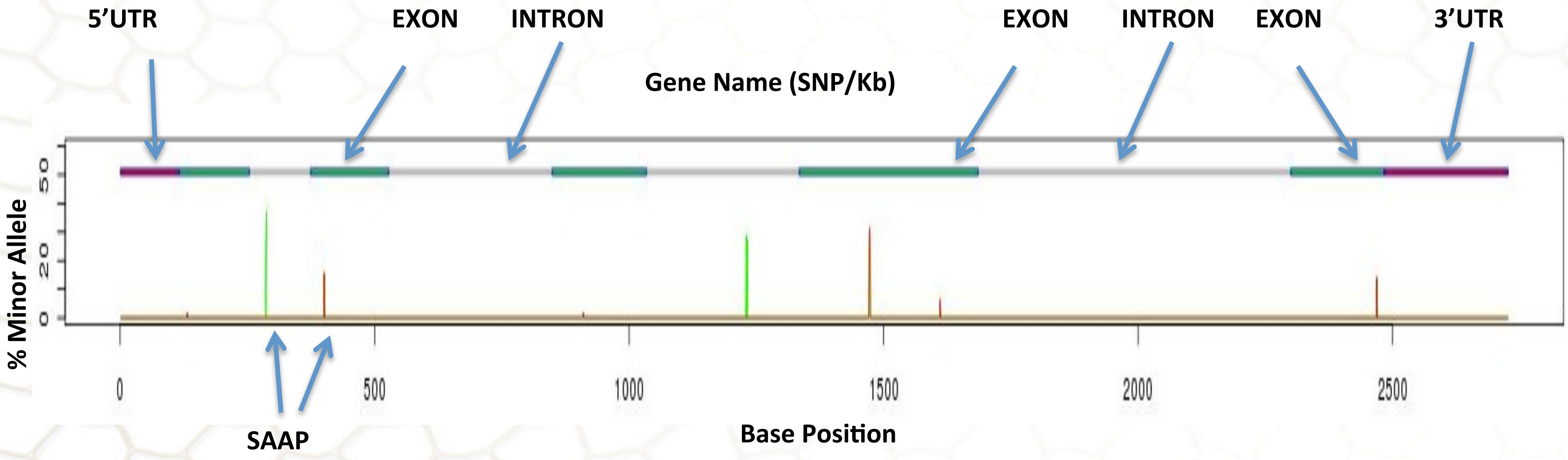
On average, 4-5% of the amino acid positions in a protein were affected due to SAAPs

At a FDR of <1%:

- 3% increase in detected proteins
- 25% increase in peptide coverage per protein

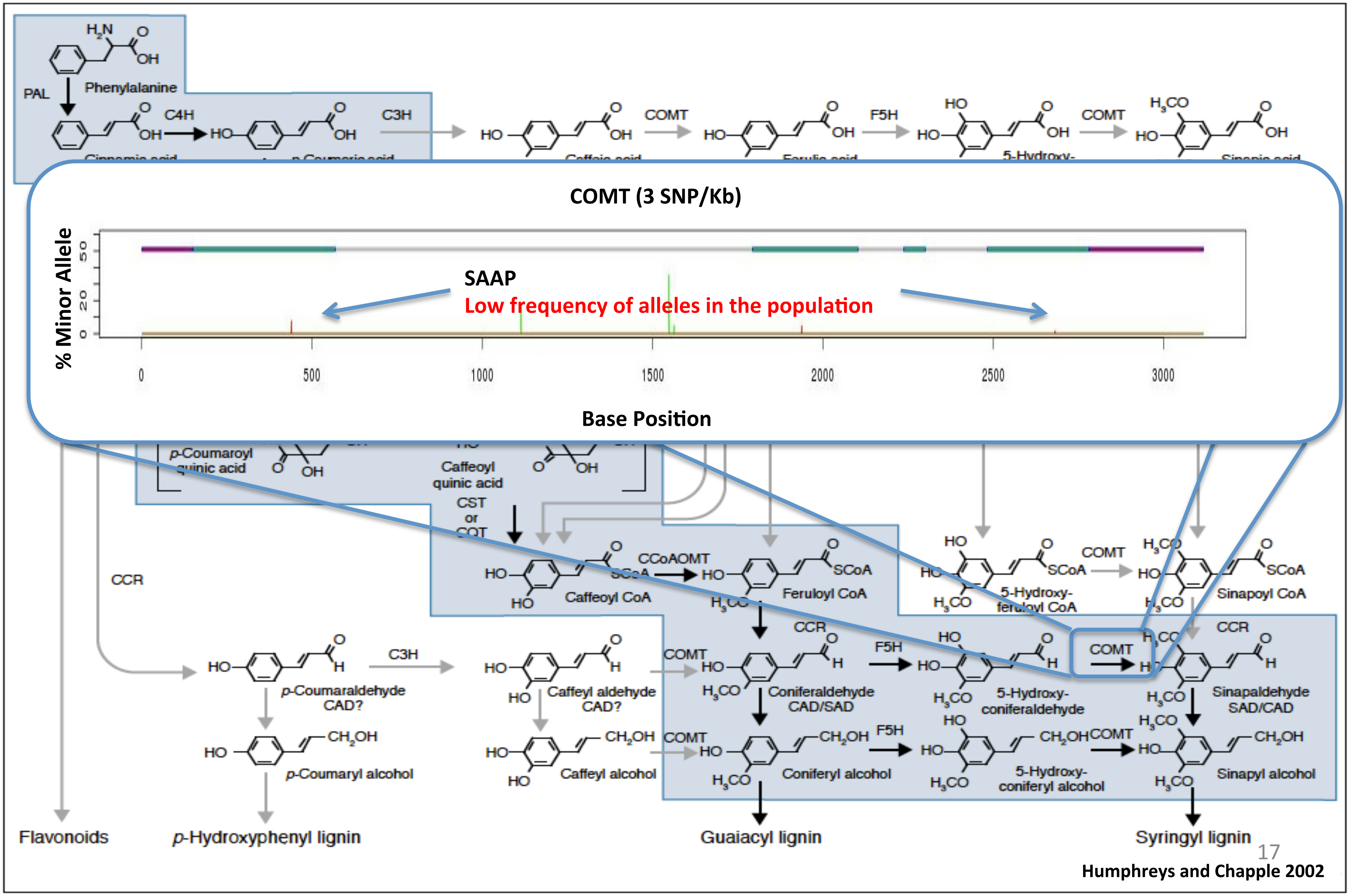
SAAP-based Functional Protein Modifications

- Non-synonymous substitutions
- Splice site mutants
- Premature stop codons
- Domain disruptions
- Conformational changes in protein folding
- Post-Translational Modifications



Ranjan Priya, Paul Abraham, Rachel Adams, Bob Hettich

Lignin Biosynthetic Pathway -- COMT



COMT (3 SNP/Kb)

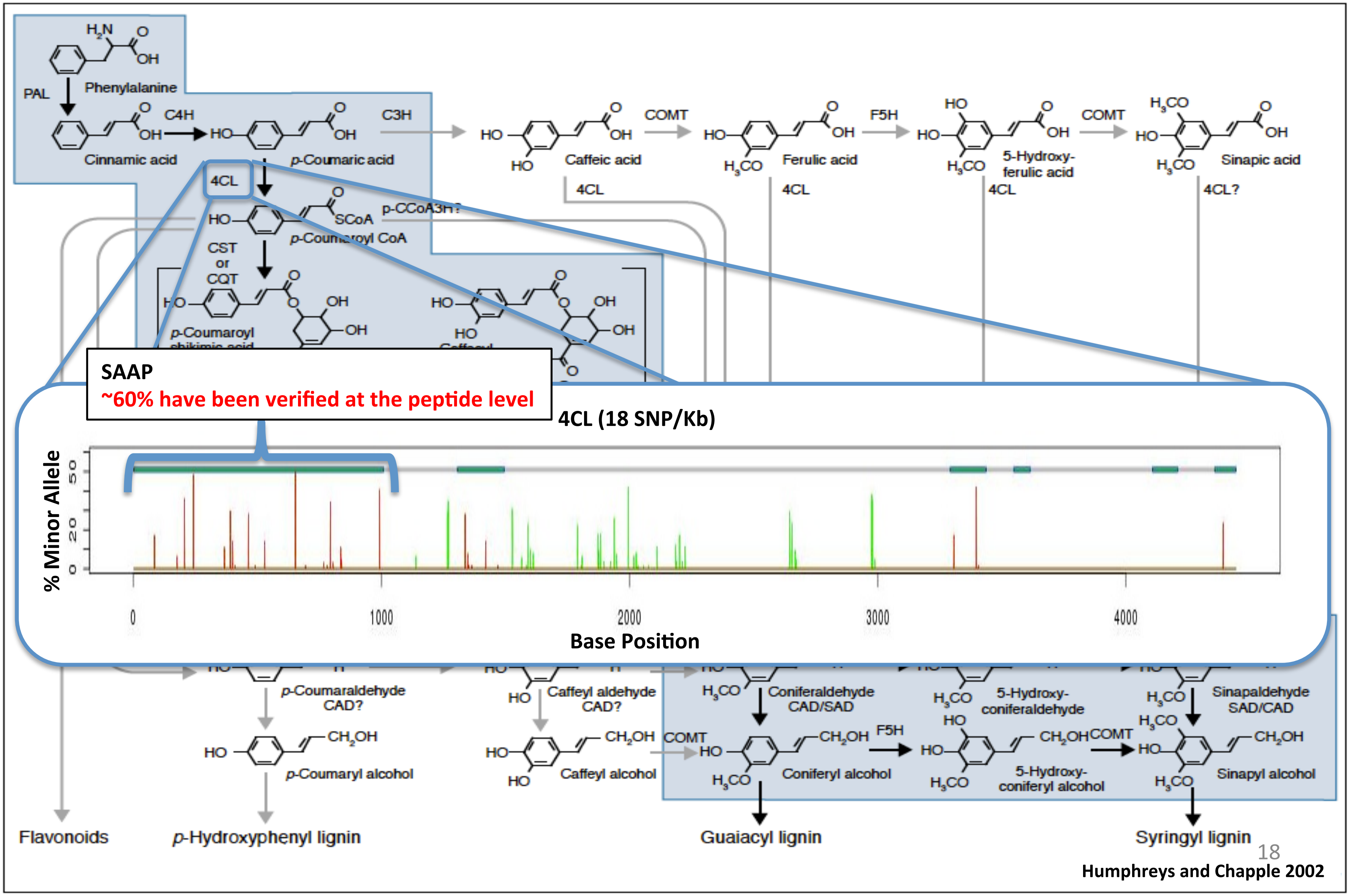
SAAP

Low frequency of alleles in the population

Base Position

Humphreys and Chapple 2002

Lignin Biosynthetic Pathway – 4CL



SAAP

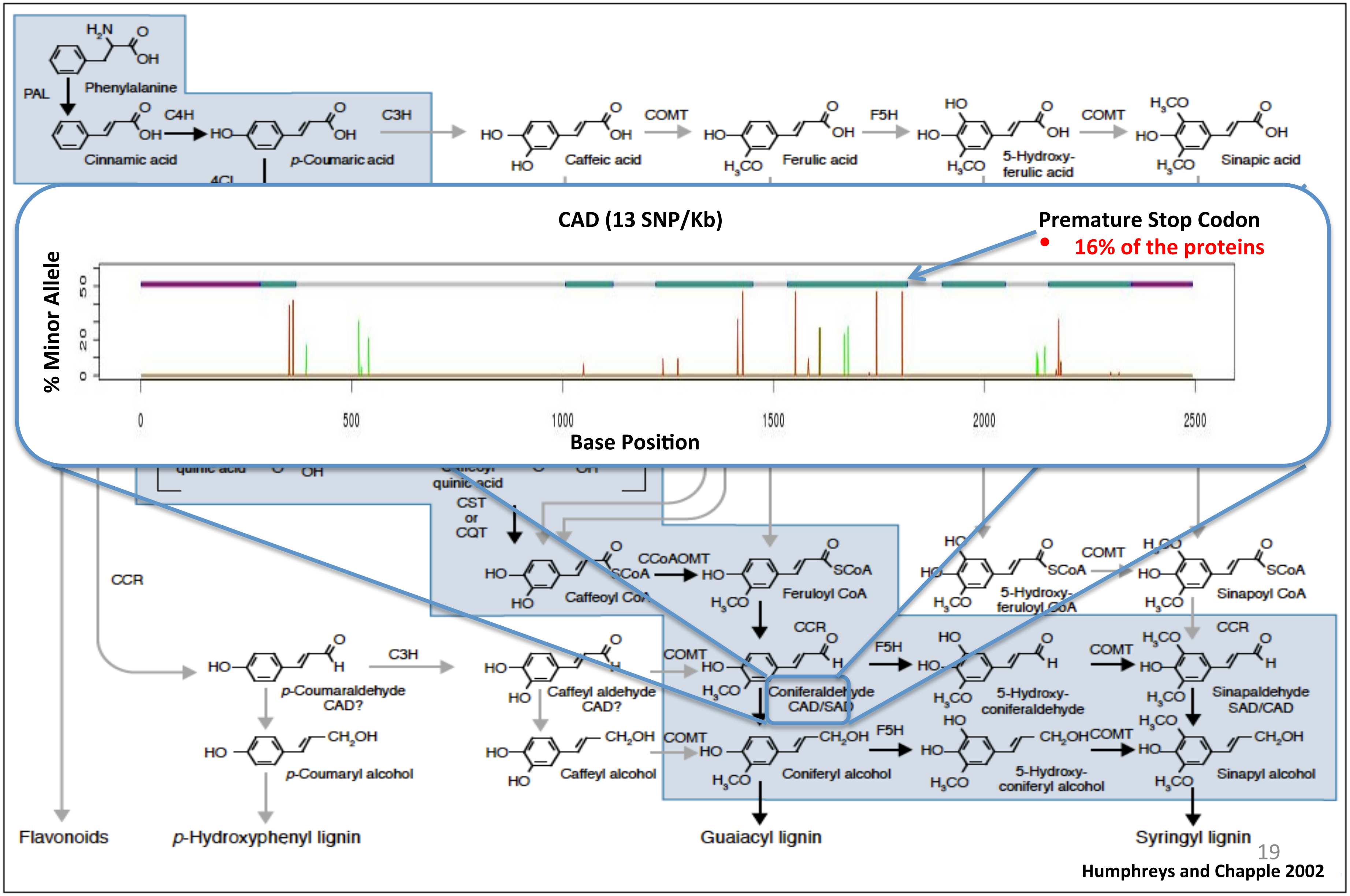
~60% have been verified at the peptide level

4CL (18 SNP/Kb)

Base Position

Humphreys and Chapple 2002

Lignin Biosynthetic Pathway -- CAD



CAD (13 SNP/Kb)

Premature Stop Codon

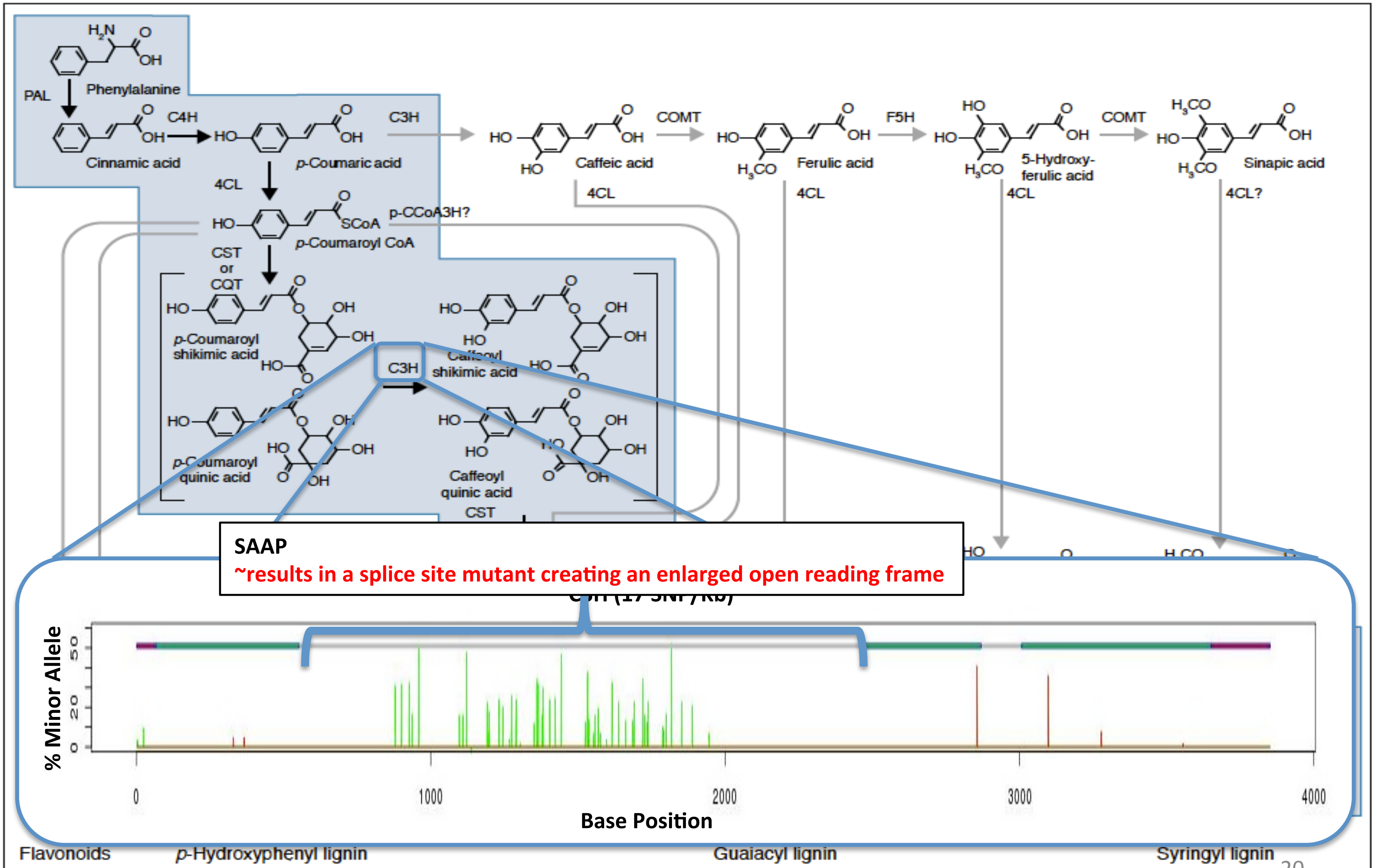
- **16% of the proteins**

% Minor Allele

Base Position

Humphreys and Chapple 2002

Lignin Biosynthetic Pathway – C3H

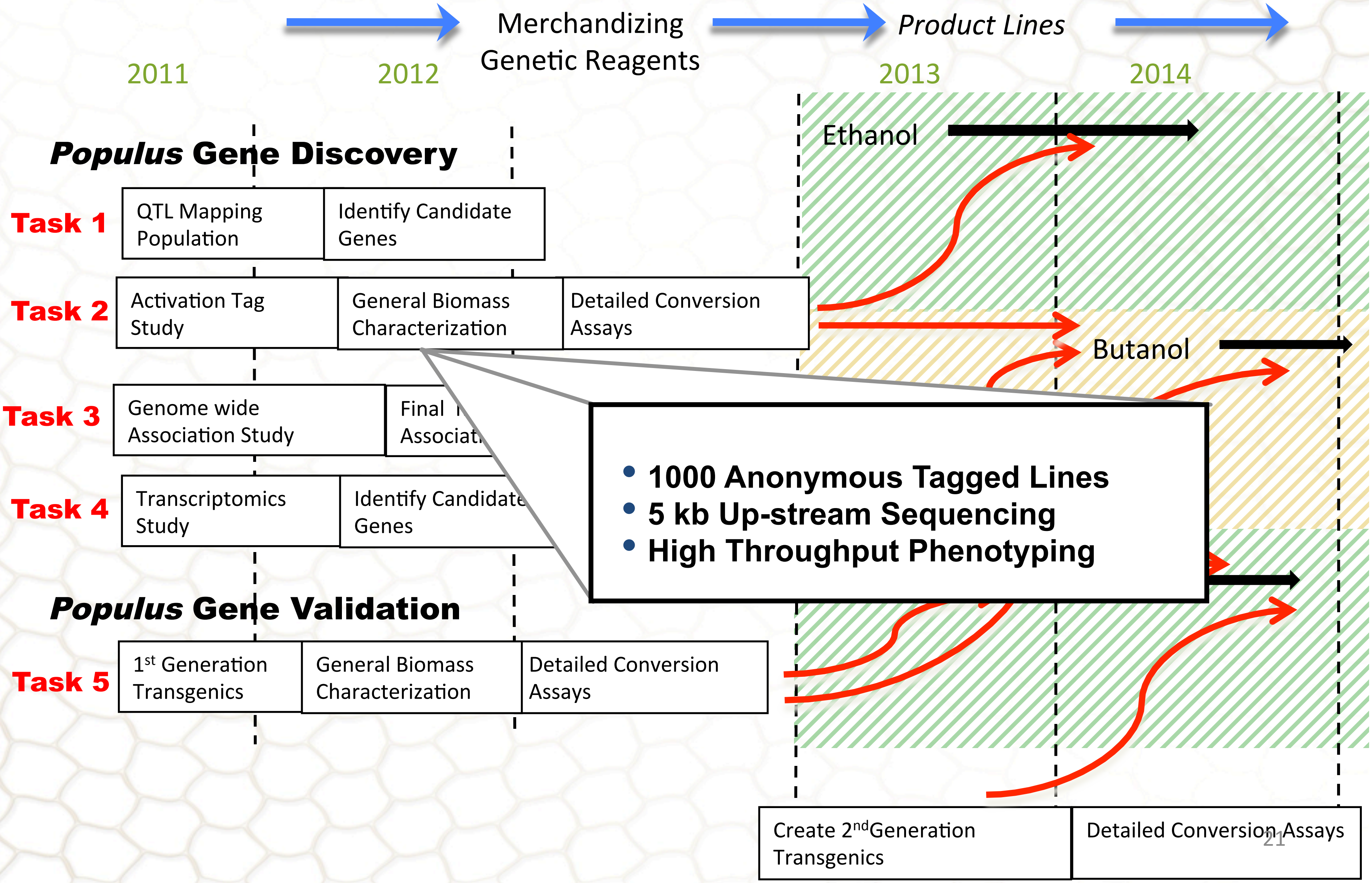


SAAP

~results in a splice site mutant creating an enlarged open reading frame

Humphreys and Chapple 2002

De Novo Gene Discovery – Activation Tagging



Merchandizing Genetic Reagents

Product Lines

2011

2012

2013

2014

***Populus* Gene Discovery**

Task 1
QTL Mapping Population
Identify Candidate Genes

Task 2
Activation Tag Study
General Biomass Characterization
Detailed Conversion Assays

Task 3
Genome wide Association Study
Final
Associat

Task 4
Transcriptomics Study
Identify Candidate Genes

***Populus* Gene Validation**

Task 5
1st Generation Transgenics
General Biomass Characterization
Detailed Conversion Assays

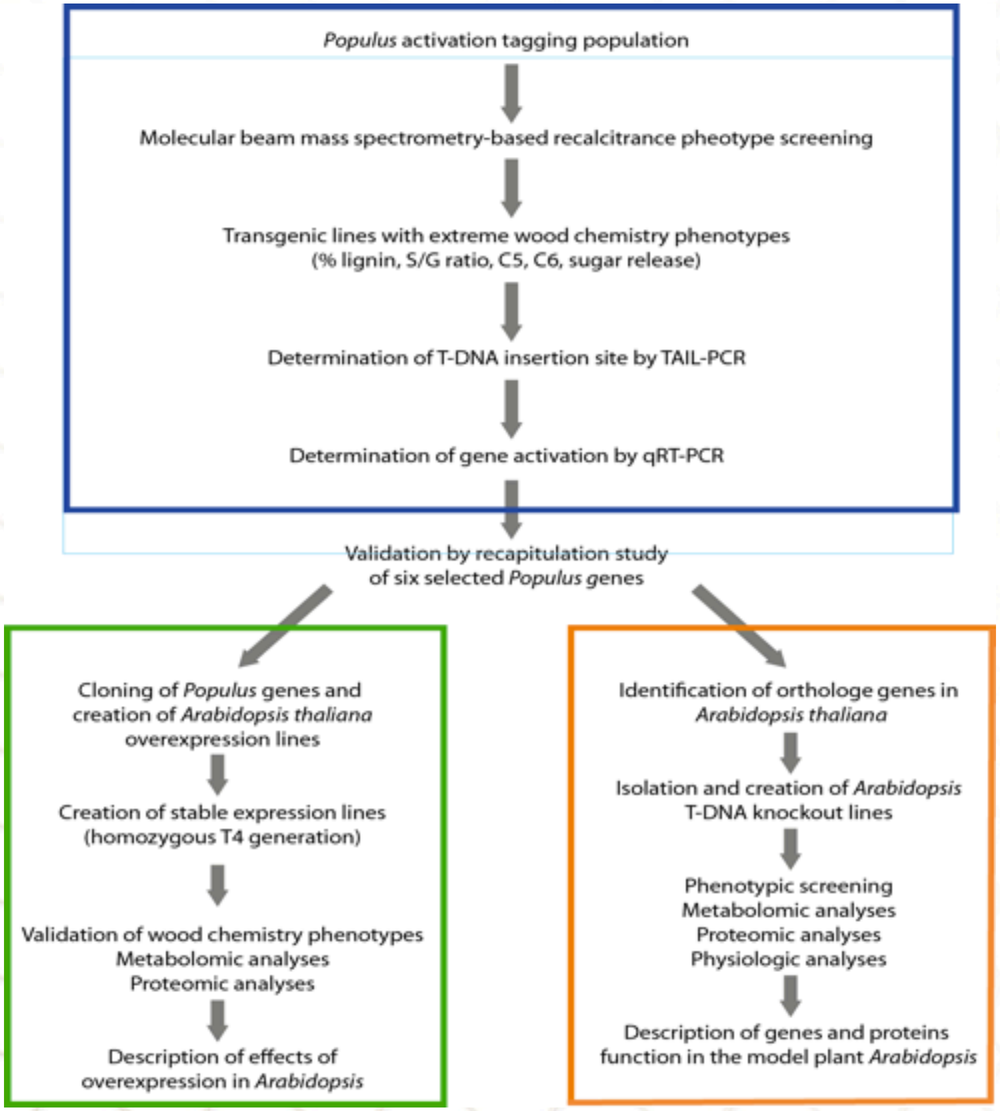
Create 2nd Generation Transgenics
Detailed Conversion Assays

Ethanol

Butanol

- **1000 Anonymous Tagged Lines**
- **5 kb Up-stream Sequencing**
- **High Throughput Phenotyping**

Activation Tagging: Results



Populus activation tagging population

Molecular beam mass spectrometry-based recalcitrance pheotype screening

Transgenic lines with extreme wood chemistry phenotypes
(% lignin, S/G ratio, C5, C6, sugar release)

Determination of T-DNA insertion site by TAIL-PCR

Determination of gene activation by qRT-PCR

Validation by recapitulation study
of six selected *Populus* genes

Cloning of *Populus* genes and
creation of *Arabidopsis thaliana*
overexpression lines

Creation of stable expression lines
(homozygous T4 generation)

Validation of wood chemistry phenotypes
Metabolomic analyses
Proteomic analyses

Description of effects of
overexpression in *Arabidopsis*

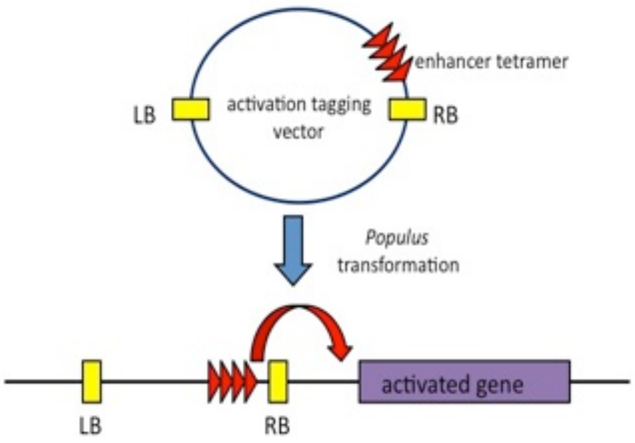
Identification of orthologe genes in
Arabidopsis thaliana

Isolation and creation of *Arabidopsis*
T-DNA knockout lines

Phenotypic screening
Metabolomic analyses
Proteomic analyses
Physiologic analyses

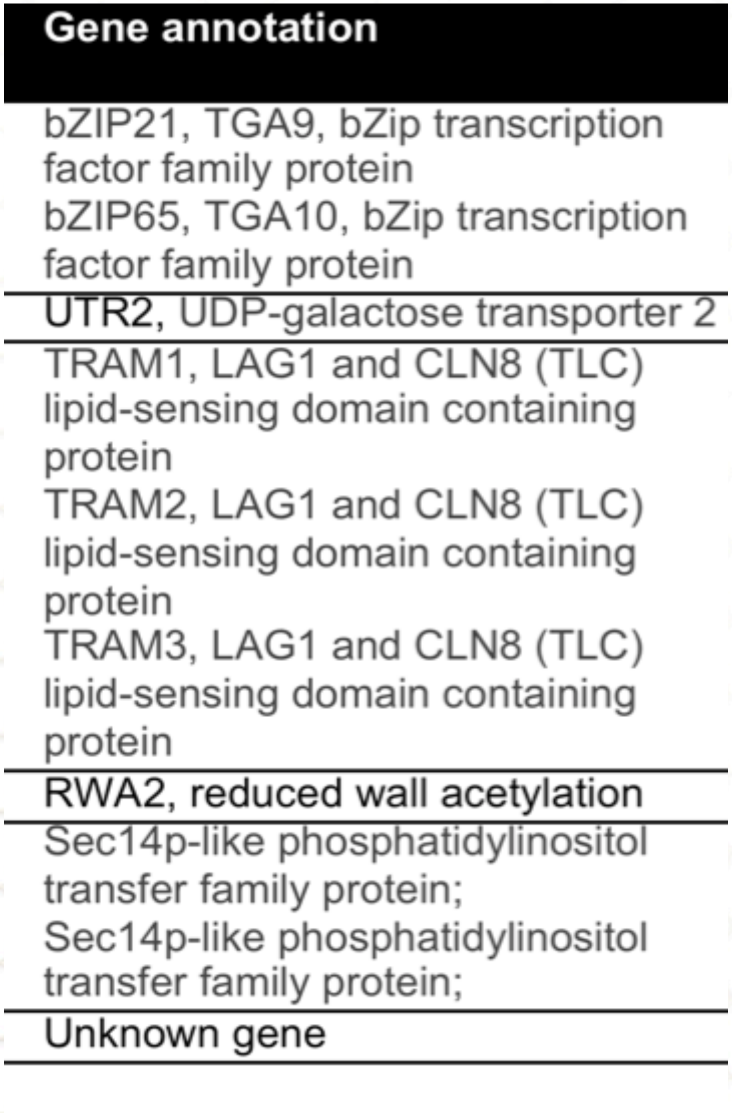
Description of genes and proteins
function in the model plant *Arabidopsis*

Research strategy to discover genes regulating cell wall biosynthesis and biomass recalcitrance



Activation tagging: an insertional mutagenesis technique that results in the dominant up-regulation of an endogenous gene.

Activation Tagging: Results



Gene annotation
bZIP21, TGA9, bZip transcription factor family protein bZIP65, TGA10, bZip transcription factor family protein
UTR2, UDP-galactose transporter 2
TRAM1, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein TRAM2, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein TRAM3, LAG1 and CLN8 (TLC) lipid-sensing domain containing protein
RWA2, reduced wall acetylation
Sec14p-like phosphatidylinositol transfer family protein; Sec14p-like phosphatidylinositol transfer family protein;
Unknown gene

Arabidopsis genes putatively involved in cell wall biosynthesis and biomass recalcitrance

Six *Populus* genes, Ten *Arabidopsis* homologues, 28 *Arabidopsis* T-DNA insertion lines . . .

Jay Chen, Olaf Czarnecki

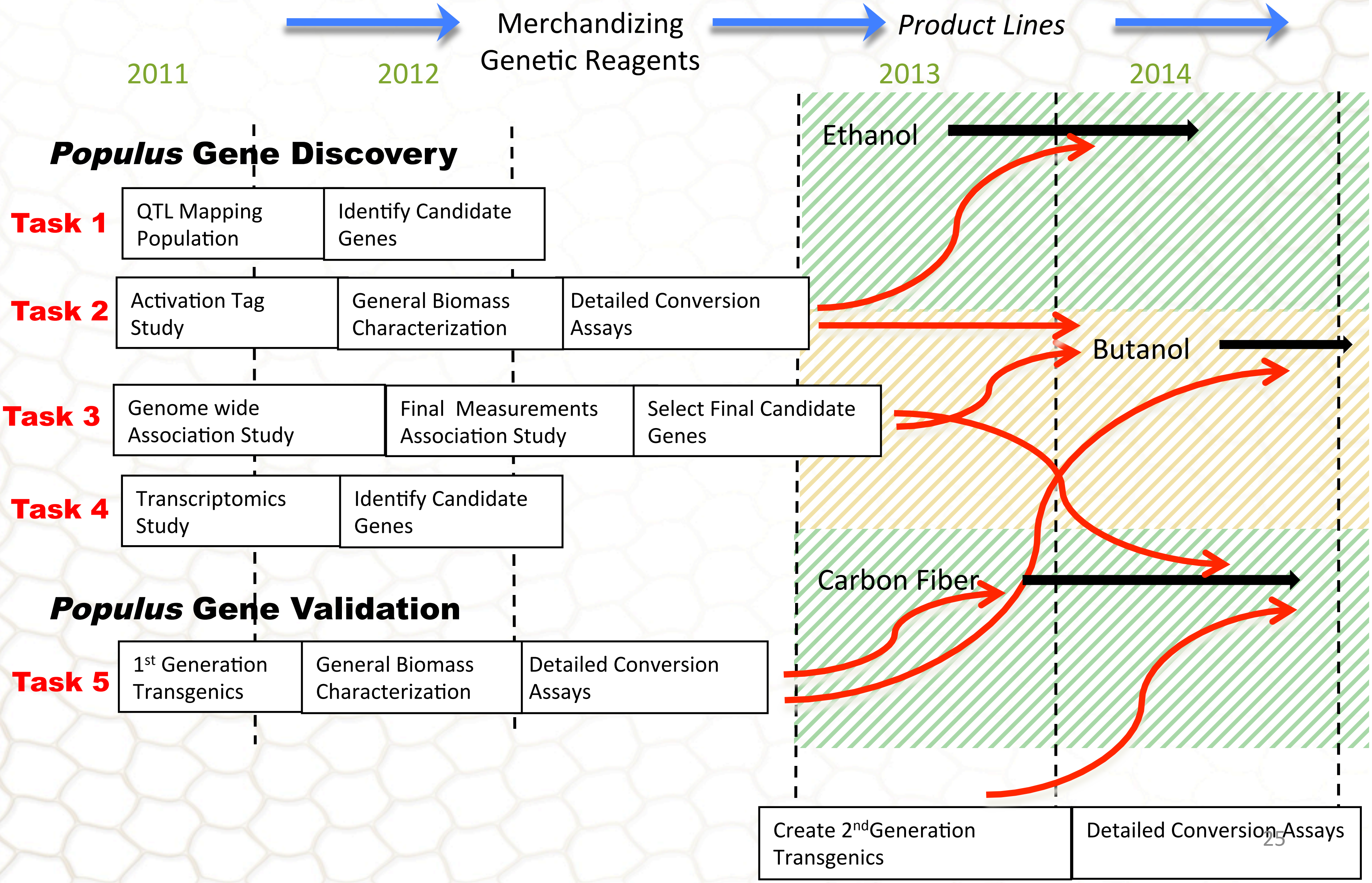
Activation Tagging: Results

One clear morphological phenotype so far…

2-week-old
Arabidopsis
mutant

4-week-old
Arabidopsis
mutant

De Novo Gene Discovery – Transformation Efforts



Merchandizing Genetic Reagents

Product Lines

2011

2012

2013

2014

***Populus* Gene Discovery**

Task 1

QTL Mapping Population

Identify Candidate Genes

Task 2

Activation Tag Study

General Biomass Characterization

Detailed Conversion Assays

Task 3

Genome wide Association Study

Final Measurements Association Study

Select Final Candidate Genes

Task 4

Transcriptomics Study

Identify Candidate Genes

***Populus* Gene Validation**

Task 5

1st Generation Transgenics

General Biomass Characterization

Detailed Conversion Assays

Ethanol

Butanol

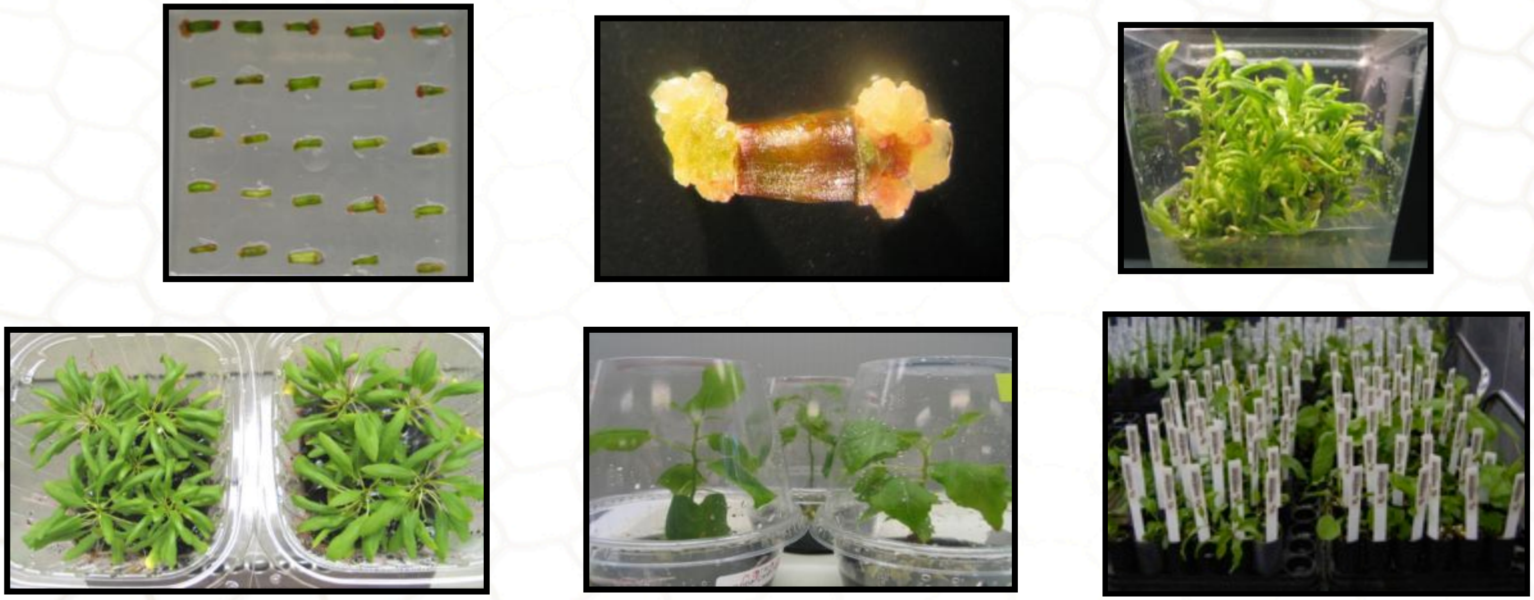
Carbon Fiber

Create 2nd Generation Transgenics

Detailed Conversion Assays

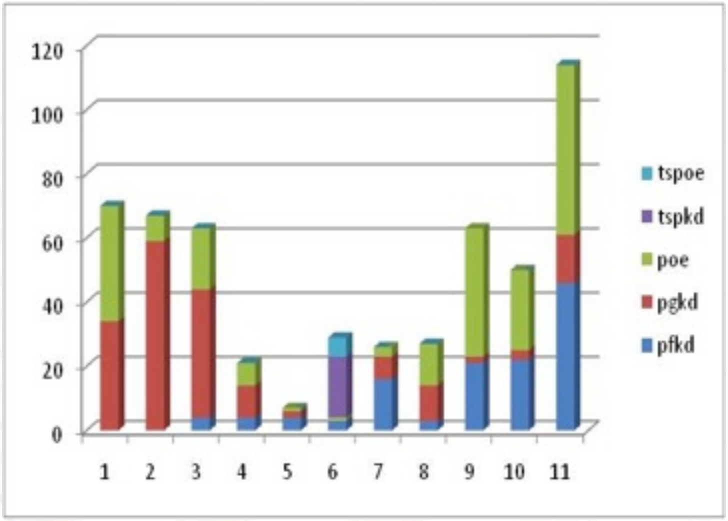
Stable Transformation Results

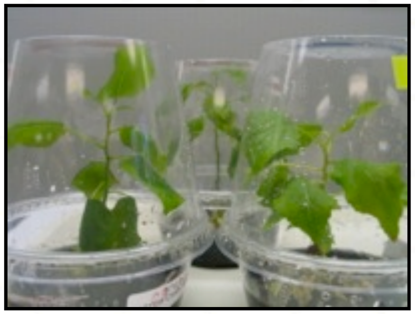
Constitutive Over-Expression and RNAi Knockdown lines have been created



Maud Hinchee, Will Rottmann, Lee Gunter

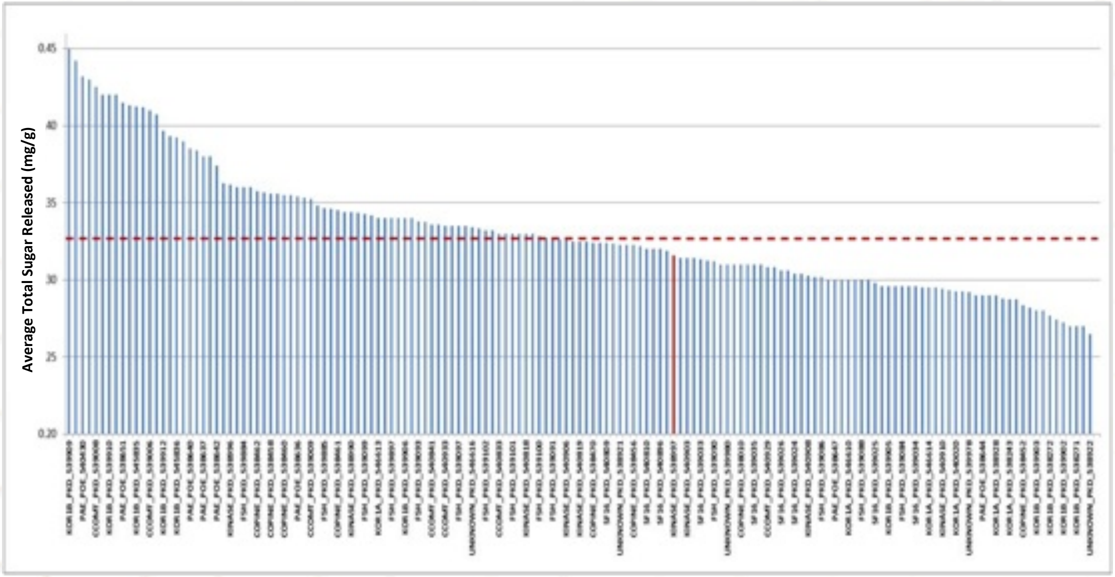
Stable Transformation Results



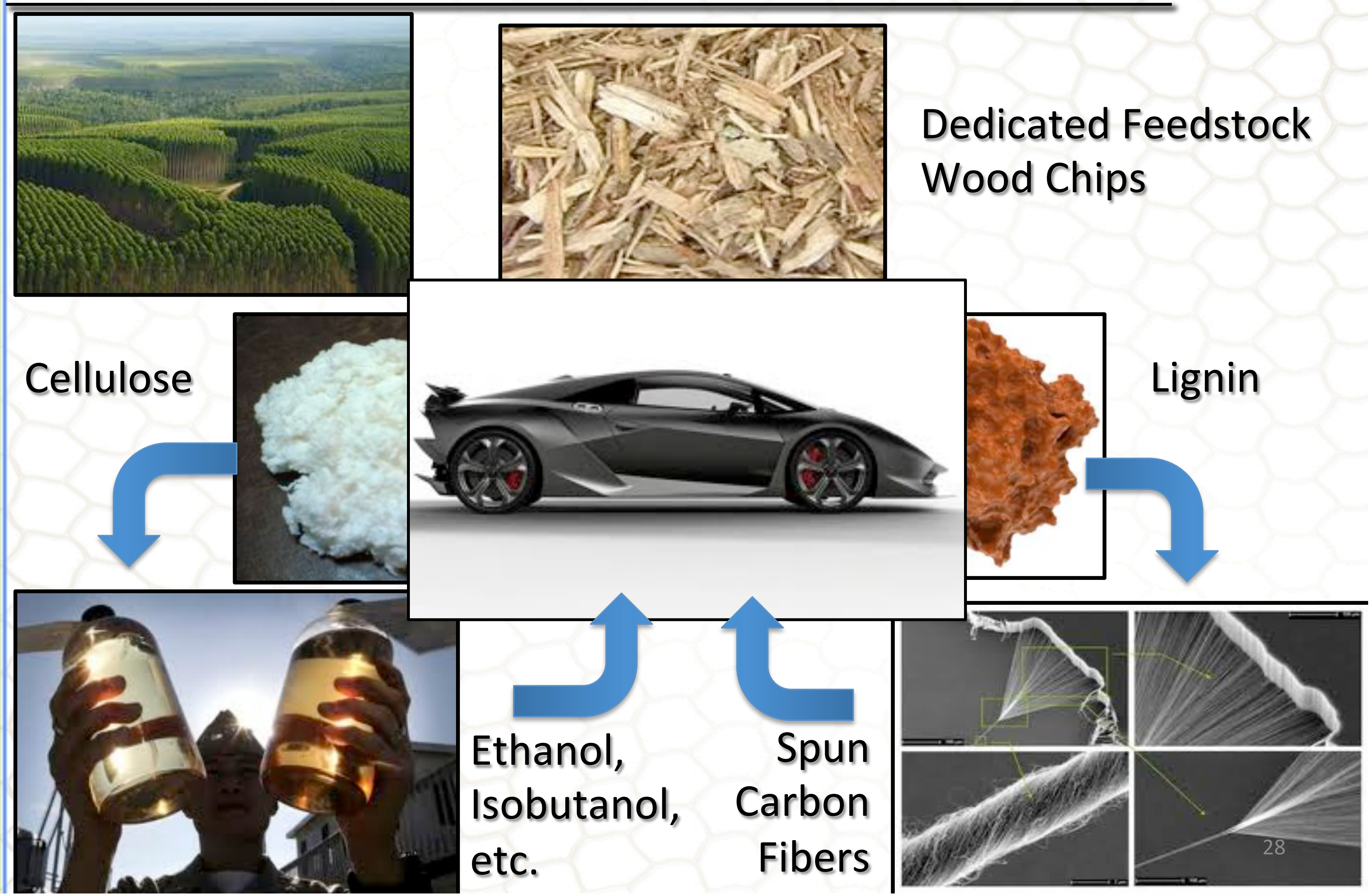


Candidate gene submission and analysis

- **Primer design for 537 gene expression constructs**
- Sequence verification for **273** constructs
- Ramets/constructs received (**24,508/127**)
- Analyzed for wood chemistry (**3,717**) and sugar release (**1,654**)
- **42** elite lines selected for qPCR expression and metabolite analysis
- Propagated **98** constructs in stool beds



Cell Wall Deconstruction and Product Output



Dedicated Feedstock
Wood Chips

Cellulose

Lignin

Ethanol,
Isobutanol,
etc.

Spun
Carbon
Fibers

Acknowledgments – Thank You

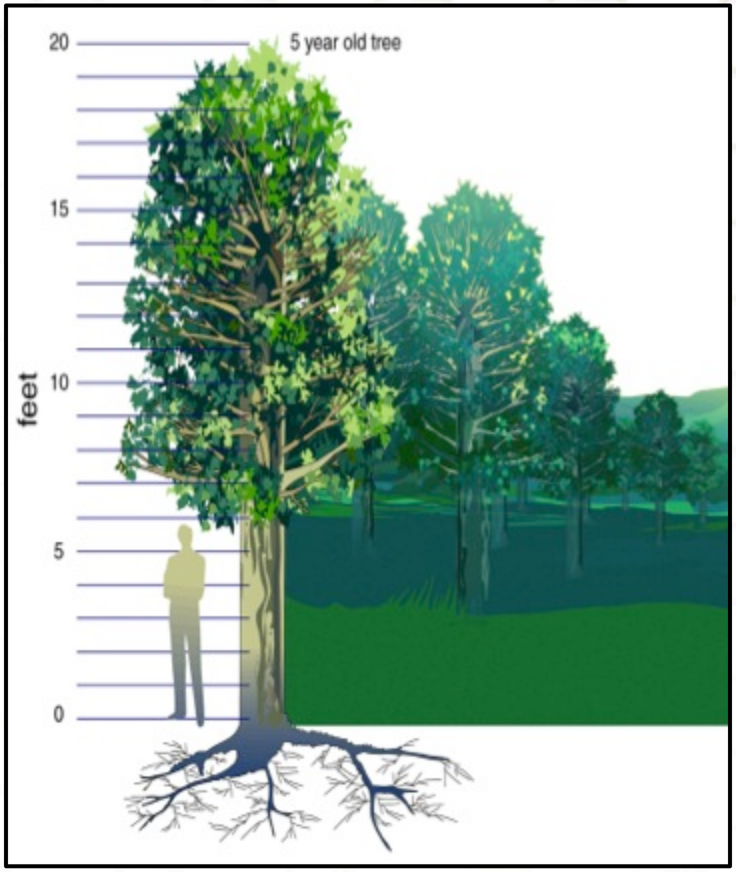
SNP Detection -- Wendy Schackwitz, Joel Martin, Len Pennacchio

SNP Array -- Stephen DiFazio, Gancho Slavov

Association Genetics -- Priya Ranjan, Wellington Muchero

Proteomics -- Paul Abraham, Rachel Adams, Robert Hettich

Genome Assembly & Annotation – Dan Rokshar, Jeremy Schmutz



OAK RIDGE National Laboratory

JGI DOE JOINT GENOME INSTITUTE US DEPARTMENT OF ENERGY OFFICE OF SCIENCE

HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY

ARBORGEN

U.S. DEPARTMENT OF ENERGY Office of Science

BESC BioEnergy Science Center