

HOS 6236 Molecular Marker Assisted Plant Breeding

Fall 2017

Last Class:

Introduction to Genomic in Breeding

Today's Class

How to use Genomic data in breeding

GWS – Developed in the context of animal breeding

- Haley and Visscher 1998 “cheap and high-density markers would move selection based on polygenes to effective genomic selection”
- Mewissen et al 2001. developed the analytical framework.
- Originally only simulations with parameters not reflecting the reality of most traits (e.g. 10 QTLs and large populations size).
- Animal breeding similarities and differences

GWS – US dairy

- US dairy industry unofficial ranking (VanRaden 2007, 2008)
- Officially in 2009 (Wiggans et al 2011)
- “Integration of genotype database was critical to achieve acceptable accuracies”
- By 2011 a couple of companies were marketing bulls based solely in GEBV
- It took them less than 10 years to apply the new methodology

GWS – Rationale – data point of view

- Massive phenotypic (and pedigree) data exist!!
- Massive genomic data exist... And keep increasing!!
 - 1990s Hundreds - thousands
 - 2000s Thousands – hundred thousands
 - >2016 hundred thousands – millions
- Use of whole genome
- (Hayes et al. www.1000bulldogenomes.com) but does not make a difference for prediction (ICQG 2012)

Resources and Strategies

- Dairy cattle low-density chips: 384 ; Bovine3K ; 6K
- High density: BovineSNP50 ; BovineHD (800K)
- HD used to find how many genes controlled a trait in 20,000 cattle ...1000's genes with small effect (Goddard 2012)
- Use of the 50K and either 3K or 6K with **imputation** (Mulder et al 2012)

GWS – Use of markers Statistics view

- **Before:**
Markers used as **Fixed effects**
Statistic thresholds used to declare significance of the effect
Usually **One marker at the time**
Simple Regression
- **Current:**
Markers used as **Random effects**
No statistics thresholds but **Cross validation**
All markers at the same time Linear Mixed Models/Bayesian
- These simple changes brings statistics consequences that require a new framework

GWS – Accuracy depends of:

1. Linkage disequilibrium between markers and QTL (gene), which depends on the effective population size (N_e) and marker density used in the genotyping
2. Size of the training population used for estimating the marker effects
3. The heritability of the trait
4. Distribution of the QTLs

J. Dairy Sci. 92:433–443

doi:10.3168/jds.2008-1646

© American Dairy Science Association, 2009.

Invited review: Genomic selection in dairy cattle: Progress and challenges

B. J. Hayes,^{*1} P. J. Bowman,^{*} A. J. Chamberlain,^{*} and M. E. Goddard^{*†}

^{*}Biosciences Research Division, Department of Primary Industries Victoria, 1 Park Drive, Bundoora 3083, Australia

[†]Faculty of Land and Food Resources, University of Melbourne, Parkville 3010, Australia

GWS – Accuracy depends of:

Review this paper to learn more about the factors affecting the accuracy of GWS

Tree Genetics & Genomes (2011) 7:241–255
DOI 10.1007/s11295-010-0328-4

ORIGINAL PAPER

Genomic selection in forest tree breeding

Dario Grattapaglia · Marcos D. V. Resende

Why GWS in perennials? Breeding cycle

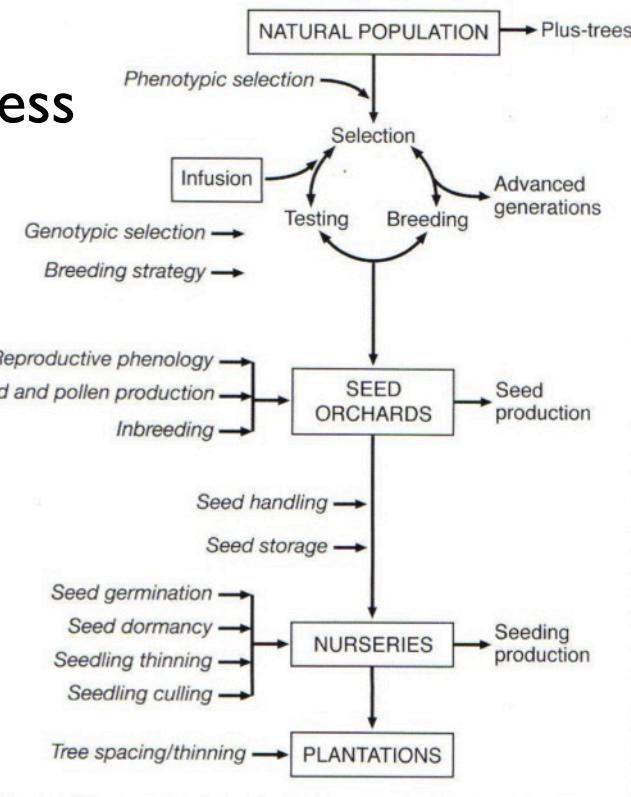
Example

Pine Breeding is a long multi-step process

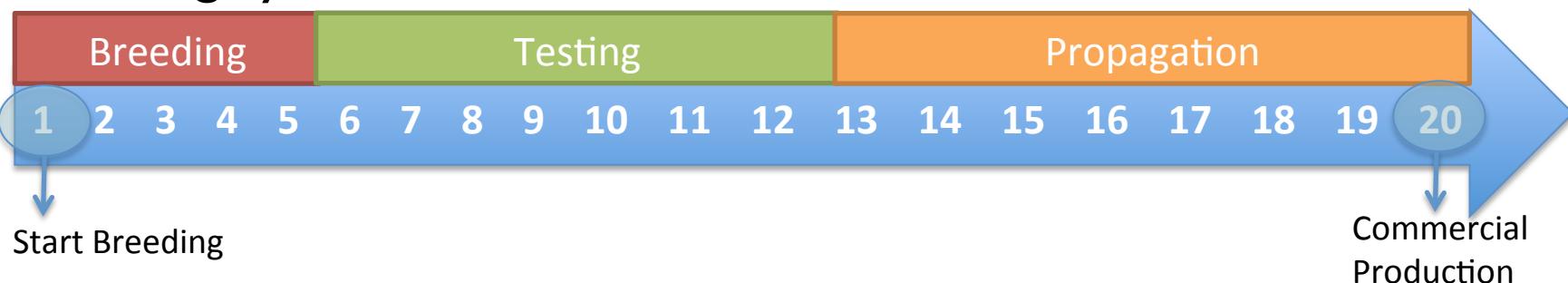
More than 30 years 1st to 2nd

Reduction of 2nd-3rd ~15 years:

- Early Selection
- Smaller populations
- Top-grafting



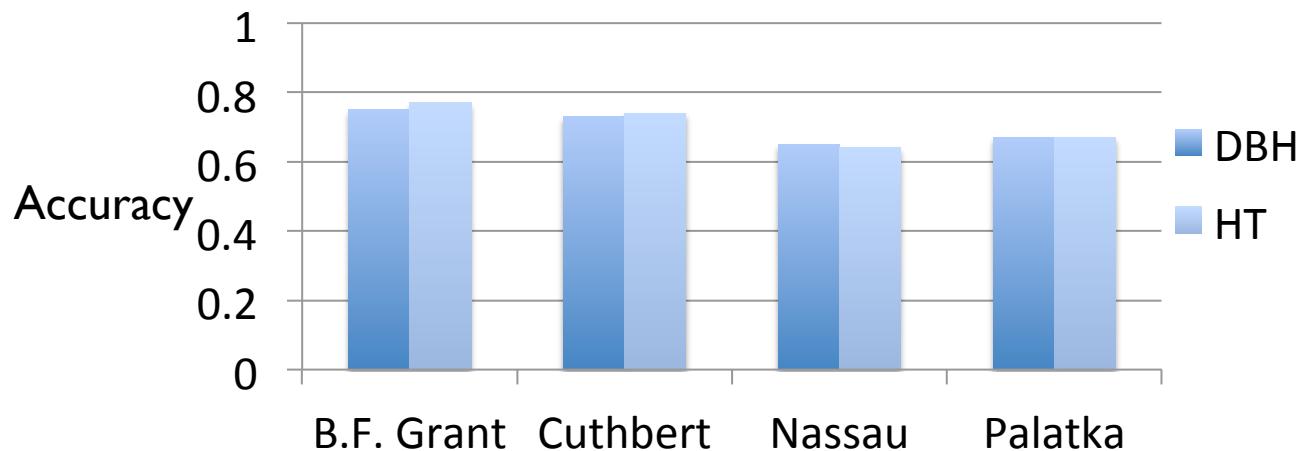
Breeding cycle



Genomic Selection – In Pine

Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments

M. F. R. Resende Jr^{1,3}, P. Muñoz^{2,3}, J. J. Acosta³, G. F. Peter^{3,4}, J. M. Davis^{3,4}, D. Grattapaglia^{5,6},
M. D. V. Resende^{7,8} and M. Kirst^{3,4}



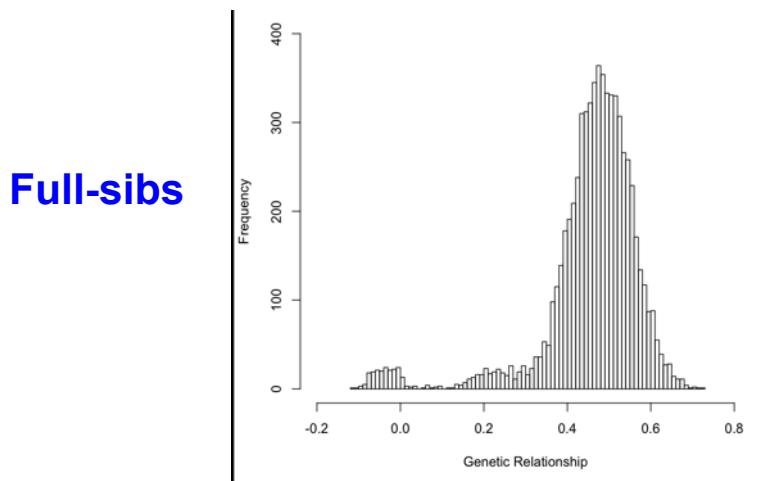
Efficiency in the Selection		
	Phenotypic Selection (14 years)	Genomic Selection (7 Years)
DBH	100%	192%
Height	100%	212%

Pedigree Errors – Detection and Correction

RESEARCH

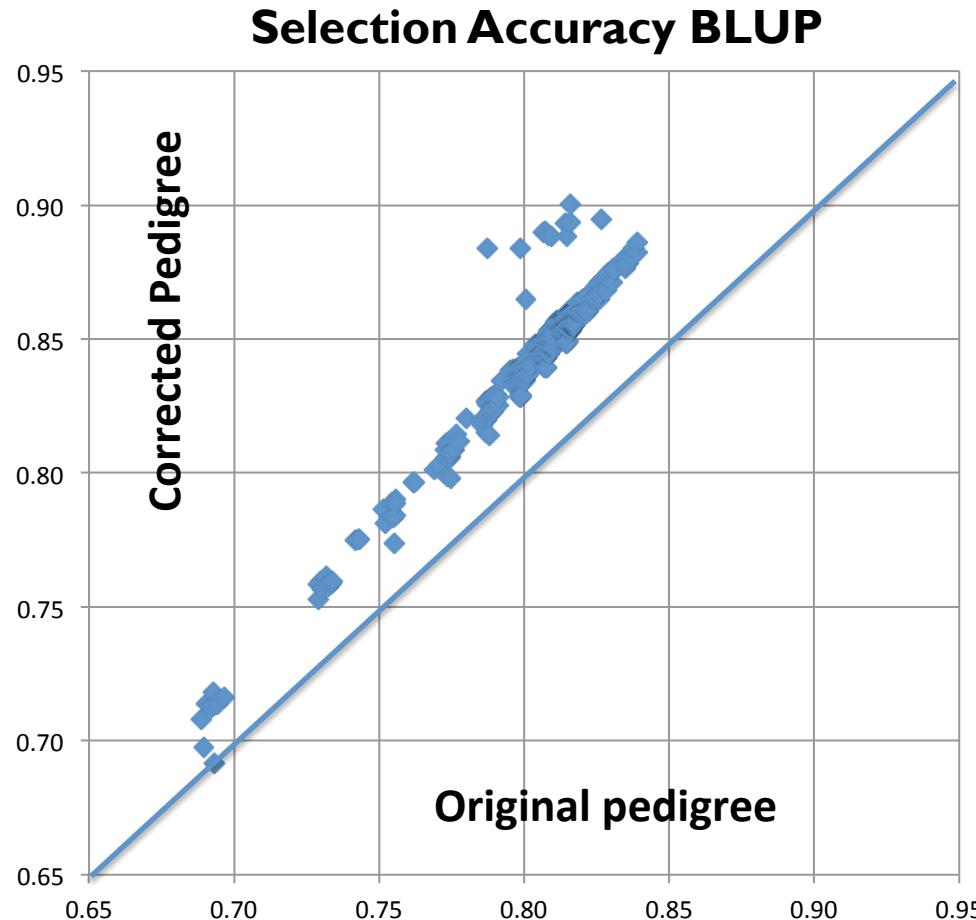
Genomic Relationship Matrix for Correcting Pedigree Errors in Breeding Populations: Impact on Genetic Parameters and Genomic Selection Accuracy

Patrício R. Munoz, Marcio F. R. Resende Jr., Dudley A. Huber, Tania Quesada, Marcos D. V. Resende, David B. Neale, Jill L. Wegrzyn, Matias Kirst, and Gary F. Peter*



Effect of Errors on Predictions

- More than 50% reduction in the standard error of genetic parameters
- Large improvement in fit of the model



- Increase of 6% (15%) in the GWS model accuracy
- Using RR-BLUP

Different Methods to Predict Marker Effect

- **BLUP-Based:** G-BLUP, RR-BLUP, RR-BLUP_B
- **Bayes-Based:** BayesA, BayesB, BayesC π , BayesR
- **LASSO-Based:** Bayesian Lasso Regression, Improved Lasso
- **Semi-Parametric Regression:** RKHS
- **Non-Parametrics:** Suport Vector Machine, Neural-Networks
- **Others...**

Meuwissen et al 2001; Habier et al 2011; De los Campos et al 2009; Legarra et al 2011; Gianola et al 2006; Long et al 2011; Gianola et al 2011

Trait Specific Analytic Methodologies

Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.)

M. F. R. Resende, Jr., *^{†,1} P. Muñoz, ^{‡,†,1} 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^{†,††,2}

*Genetics and Genomics Graduate Program, [†]Plant Molecular and Cellular Biology Graduate Program, and [†]School of Forest Resources and Conservation, University of Florida, Gainesville, Florida 32611, [§]EMBRAPA Forestry, Estrada da Ribeira, Colombo, PR 83411-000 Brazil, ^{**}Department of Forest Engineering, Universidade Federal de Viçosa, Viçosa, MG 36571-000 Brazil, ^{††}Department of Animal Science, Iowa State University, Ames, Iowa 50011, and ^{††}University of Florida Genetics Institute, University of Florida, Gainesville, Florida 32611

¹These authors contributed equally to this work.

²Corresponding author: University of Florida, Newins-Ziegler Hall Rm. 367, University of Florida, Gainesville, FL 32611. E-mail: mkirst@ufl.edu

17 traits with difference in h^2 and predicted genetic architecture

Test four different original methods that differ with respect to assumptions regarding distribution of marker effects:

RR-BLUP

Bayesian LASSO

Bayes A

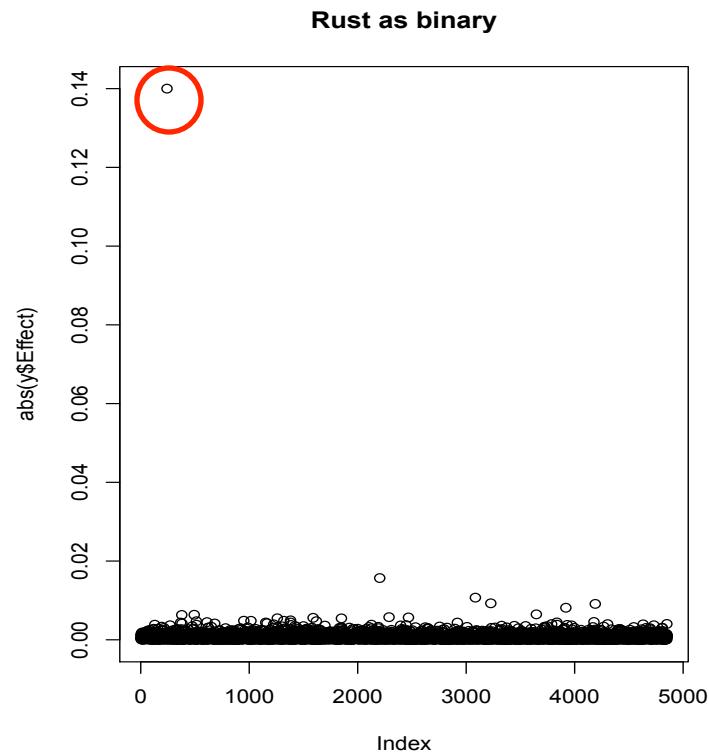
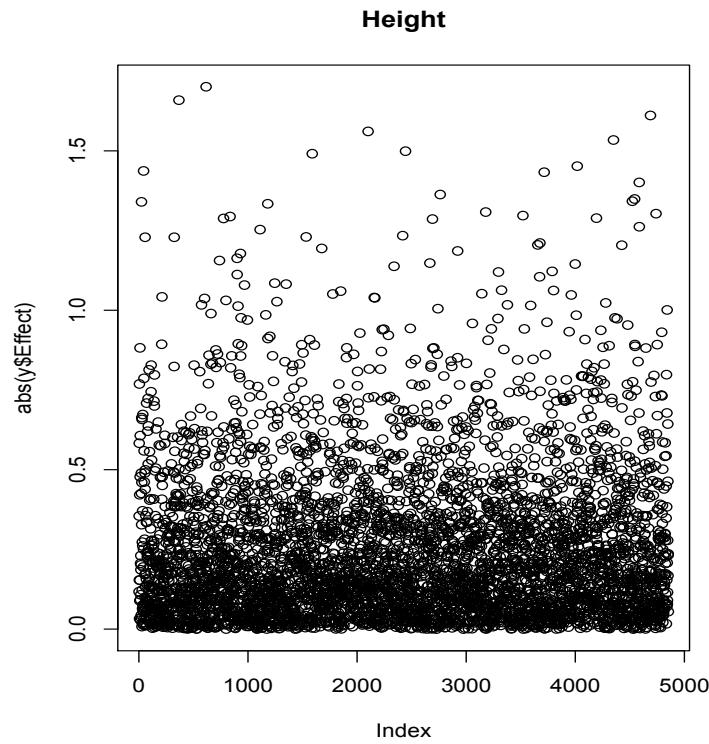
Bayes Cpi

Genomic Selection- Predictive Ability Different Methods

Trait category	Trait
Growth	HT
	HTLC
	BHLC
	DBH
Development	CWAL
	CWAC
	BD
	BA
Disease resistance	Rootnum_bin
	Rootnum
	Rust_bin
	Rust_gall_vol
Wood quality	Stiffness
	Lignin
	LateWood
	Density
	C5C6



Genomic Selection- Major Effects Traits



Genomic Selection- Rust in Pines

Forests 2014, 5, 347-362; doi:10.3390/f5020347

OPEN ACCESS

forests

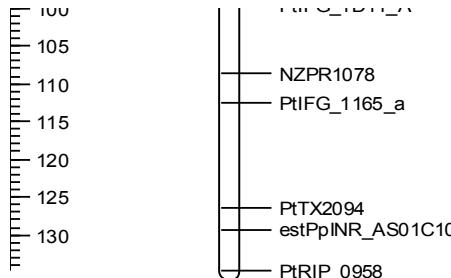
ISSN 1999-4907

www.mdpi.com/journal/forests

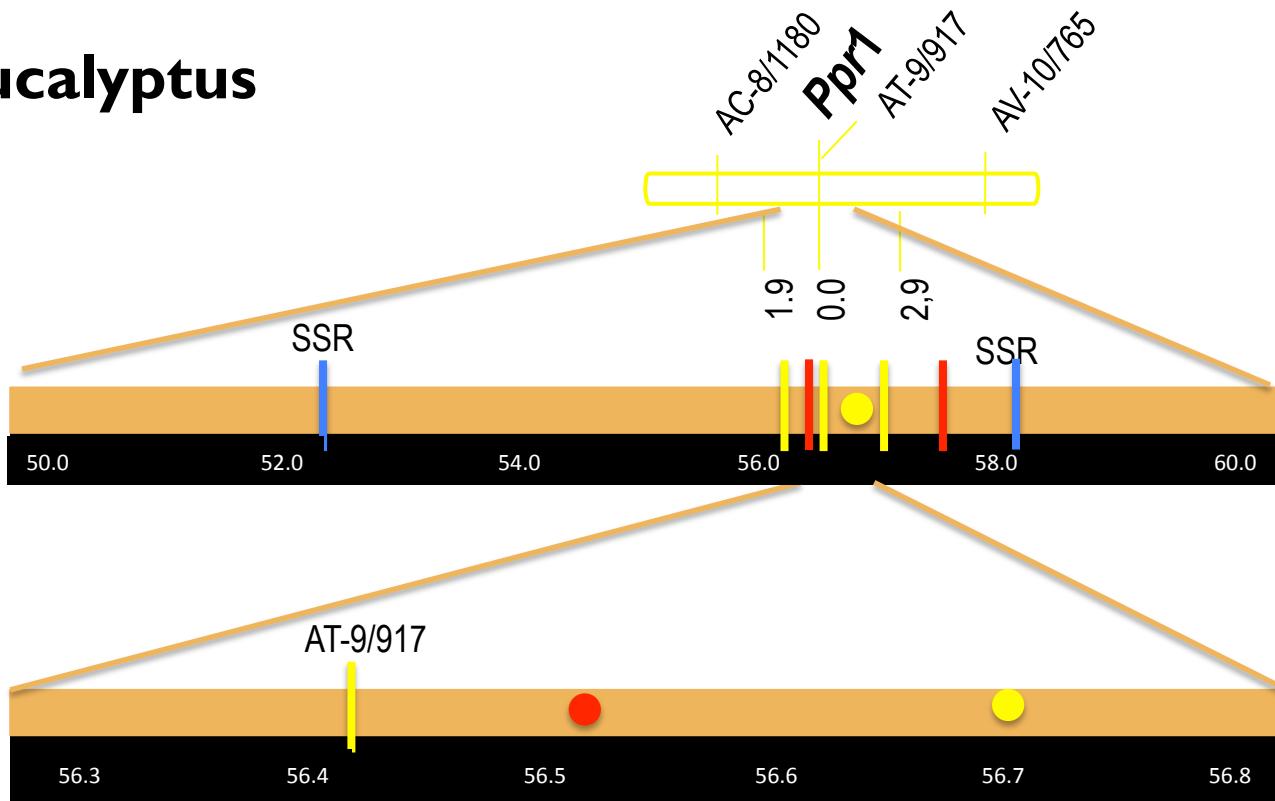
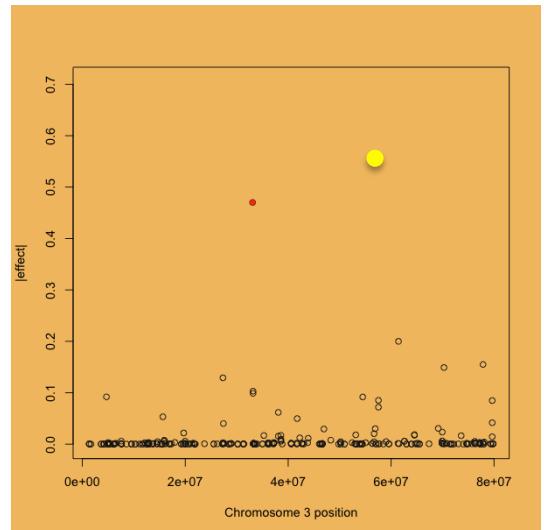
Article

Mapping Fusiform Rust Resistance Genes within a Complex Mating Design of Loblolly Pine

Tania Quesada ¹, Marcio F.R. Resende Jr. ², Patricio Muñoz ³, Jill L. Wegrzyn ⁴,
David B. Neale ⁵, Matias Kirst ^{1,6}, Gary F. Peter ^{1,6}, Salvador A. Gezan ¹, C. Dana Nelson ⁷
and John M. Davis ^{1,6,*}



GWS- Rust in Eucalyptus



Identification of a novel QTL contributing to rust resistance in *Eucalyptus* ■

Bruno M Lima^{1*}, Juliana EC Teixeira², Rodrigo Gazaffi¹, Antonio AF Garcia¹, Dario Grattapaglia³, Raphaelle KD Valle⁴, Luis EA Camargo⁴

Genetic mapping provides evidence for the role of additive and non-additive QTLs in the response of inter-specific hybrids of *Eucalyptus* to *Puccinia psidii* rust infection ■

Alexandre Alonso Alves · Carla Cristina Gonçalves Rosado ·
Danielle Assis Faria · Lúcio Mauro da Silva Guimarães · Douglas Lau ·
Sérgio Hermínio Brommonschenkel · Dario Grattapaglia · Acelino Couto Alfenas

Disease resistance protein (TIR-NBS-LRR class) family
BLAST Top Hit:

Poplar MER locus (Resistance to Populus Rust)

Resende et al. *In preparation*

Genomic Selection- In association

Full paper

Association genetics of oleoresin flow in loblolly pine: discovering genes and predicting phenotype for improved resistance to bark beetles and bioenergy potential



Jared W. Westbrook¹, Marcio F. R. Resende Jr¹, Patricio Munoz¹, Alejandro R. Walker^{1,2}, Jill L. Wegrzyn³, C. Dana Nelson⁴, David B. Neale³, Matias Kirst^{1,2}, Dudley A. Huber², Salvador A. Gezan², Gary F. Peter^{1,2} and John M. Davis^{1,2,*}

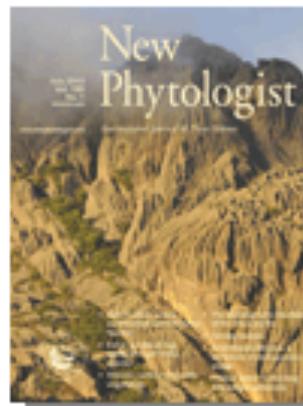
Article first published online: 28 MAR 2013

DOI: 10.1111/nph.12240

© 2013 The Authors. New Phytologist © 2013
New Phytologist Trust



Issue



New Phytologist

Volume 199, Issue 1, pages
89–100, July 2013

USDA – DOE:

“Accelerated Development of Optimal Feedstock For Bioenergy Using Genome-Wide Selection”

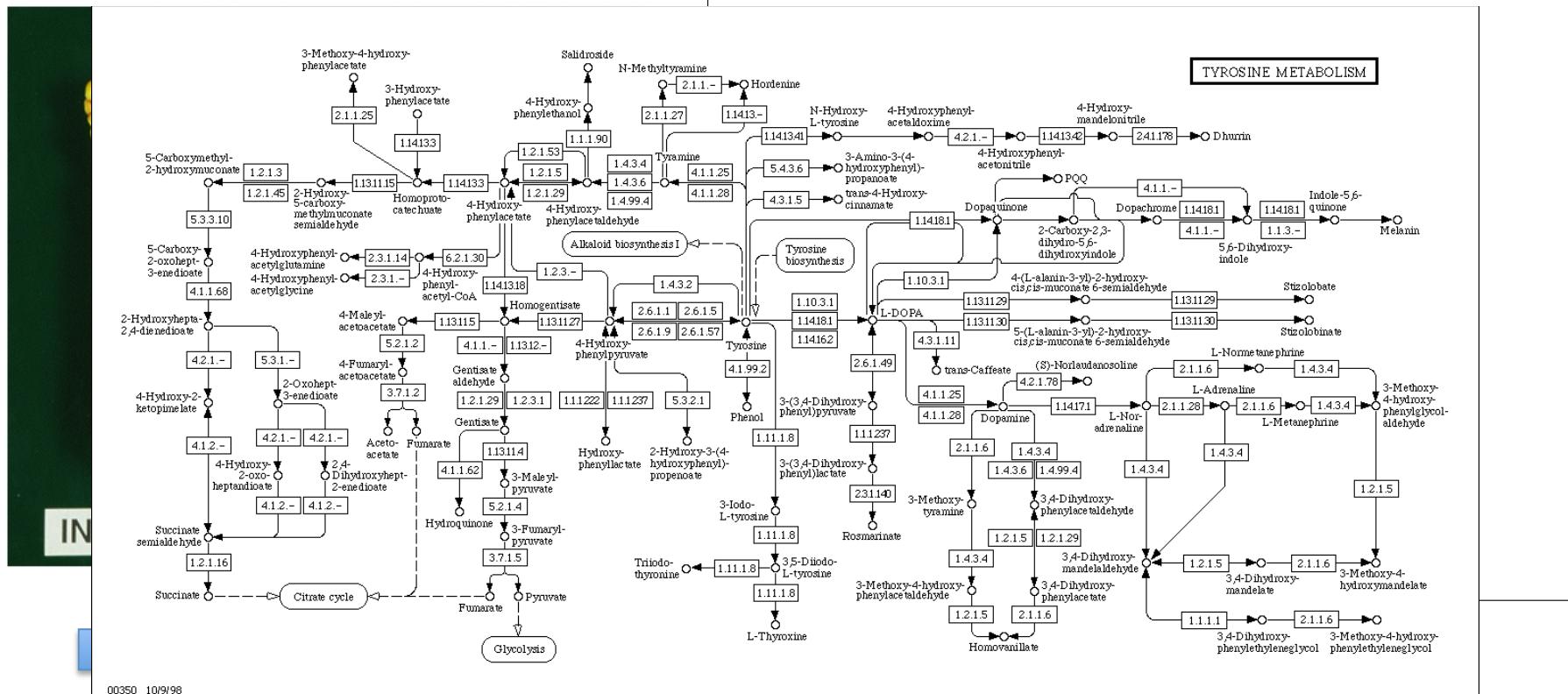
Non-Additive Effects Exist! - Evidence

Molecular evidence

Inborn metabolic defects

Interaction of different genes

DO WE REALLY THINK EVERYTHING IS ADDITIVE!!



00350 10/9/98

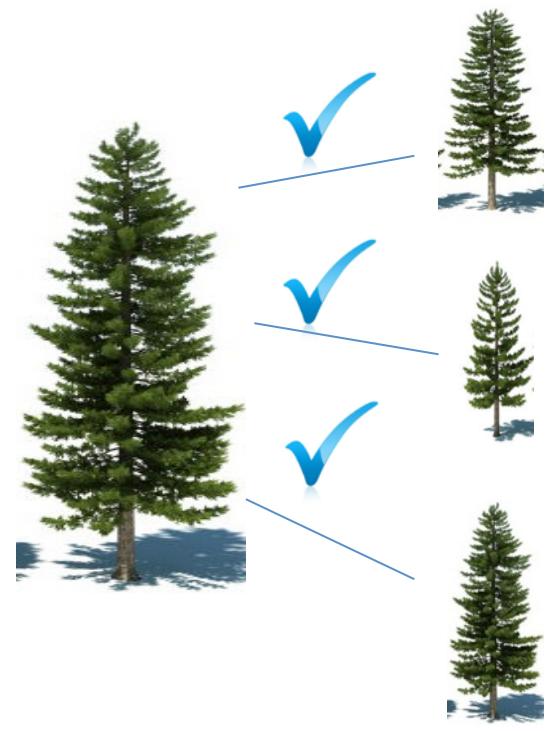
Hybrid Vigor

<http://www.genome.jp/kegg/pathway.html>

McKeand et al 2007

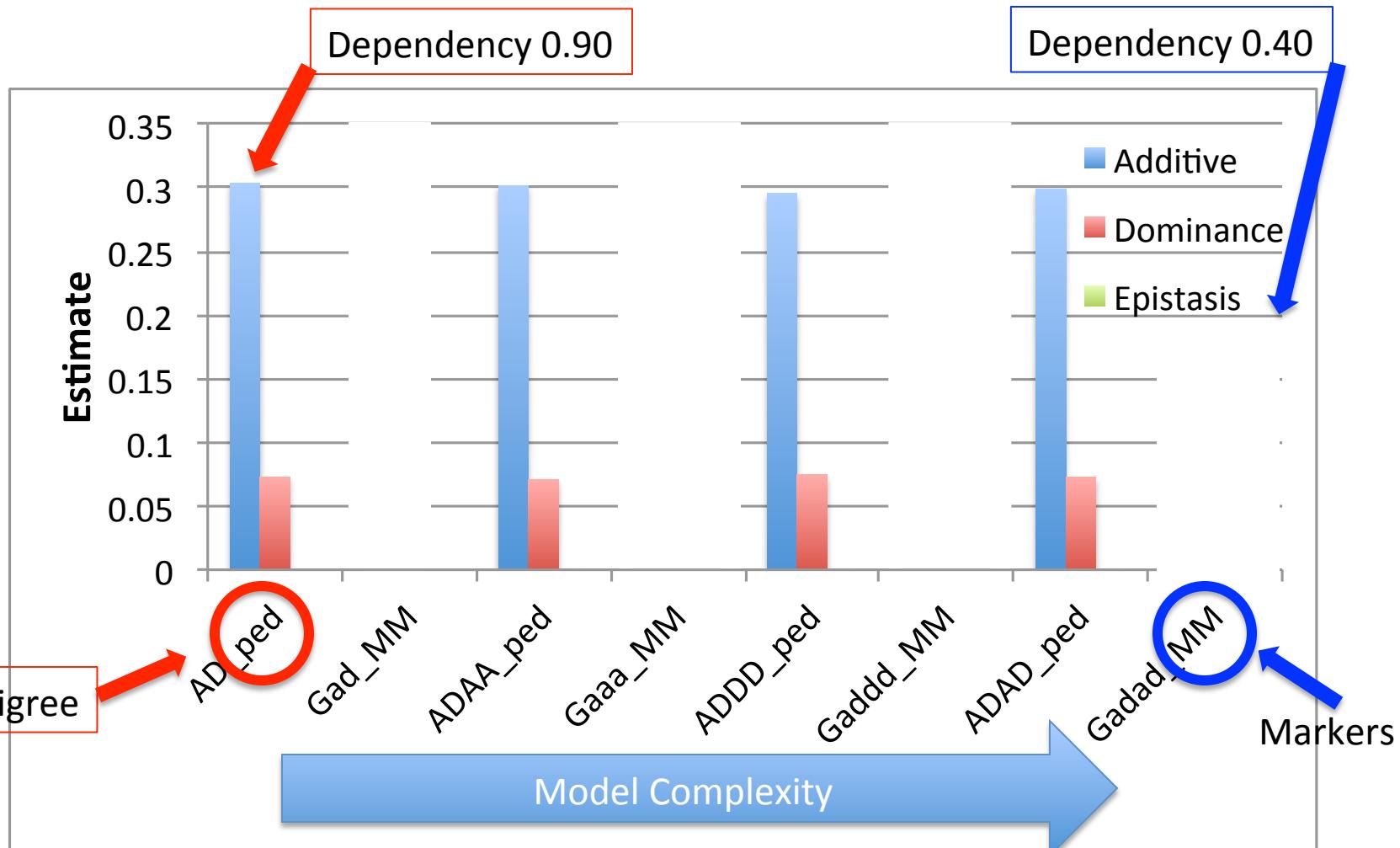
Markers-derived Relationships

Siblings



Additive Pedigree Relationships	Dominance Pedigree Relationships
0.5	0.25
0.5	0.25
0.5	0.25

Estimates of Variance Components



Even when complexity of the model increases partition is the same with pedigree.

A different architecture is observed with marker-derived matrices

Prediction in 10-fold Cross Validation

Model	Cor(RBV,PBV)	MSE(RBV,PBV)	Top10%RankCor
A_ped	0.640	1335.800	0.17
Ga_MM	0.670	1291.800	0.34
ADAD_ped	0.727	657.258	0.16
Gadad_MM	0.872	108.240	0.37
ADDD_ped	0.732	638.464	0.18
Gaddd_MM	0.873	151.199	0.32

RBV → Breeding Value using all data within model

PBV → Breeding Value in cross validation

Top10%RankCor → correlation between R-ranking and P-ranking
for top 10% performance

Estimates of Variance Components

GENOMIC SELECTION

Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices

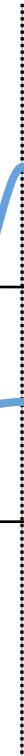
Patricio R. Muñoz,^{*,†,1,2} Marcio F. R. Resende, Jr.,^{†,‡,1} Salvador A. Gezan,[†] Marcos Deon Vilela Resende,^{§,***}
Gustavo de los Campos,^{††} Matias Kirst,^{†,‡‡} Dudley Huber,[†] and Gary F. Peter^{†,‡‡,2}

*Agronomy Department, [†]School of Forest Resources and Conservation, [‡]Genetics and Genomics Graduate Program, and
^{‡‡}University of Florida Genetics Institute, University of Florida, Gainesville, Florida 32611, [§]EMBRAPA Forestry, Estrada da Ribeira,
Colombo, PR 83411-000 Brazil, ^{**}Department of Forest Engineering, Universidade Federal de Viçosa, Viçosa, MG 36571-000
Brazil, and ^{††}Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama 35294

What should you cross?

- ~950 individuals
- ~450,000 possible crosses

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	
1		16	17			18												19															
2	6		18	17													20													18	19	22	
3	34	23		19	19																												
4		68	37			15																											
5			58		19	18					20	21																					
6				15	10	19	19											18															
7	45				11	64		21	19																								
8					50	56		18	19																					21			
9						40	26		20																								
10						25	49		21																								
11			39			44		18										21															
12								28	21	19																			21				
13	35		41					19	19	21																							
14								46	65		18																						
15						7					68		20	22																			
16										70	38		22	21																			
17	16									36	42		3	31																			
18								27			60	0	21	22																			
19											1	22		22	15																		
20											30	57		21	20															21			
21											13	29				19																	
22												8		19	18															20			
23						48							67	20	18	20																	
24													51	55		17																	
25											66																						
26													43	62		21	19																
27	54													9		18	19																
28	33													31	32		20																
29														52	21		14	22															
30	12													3		2																	
31															61																		
32																4																	



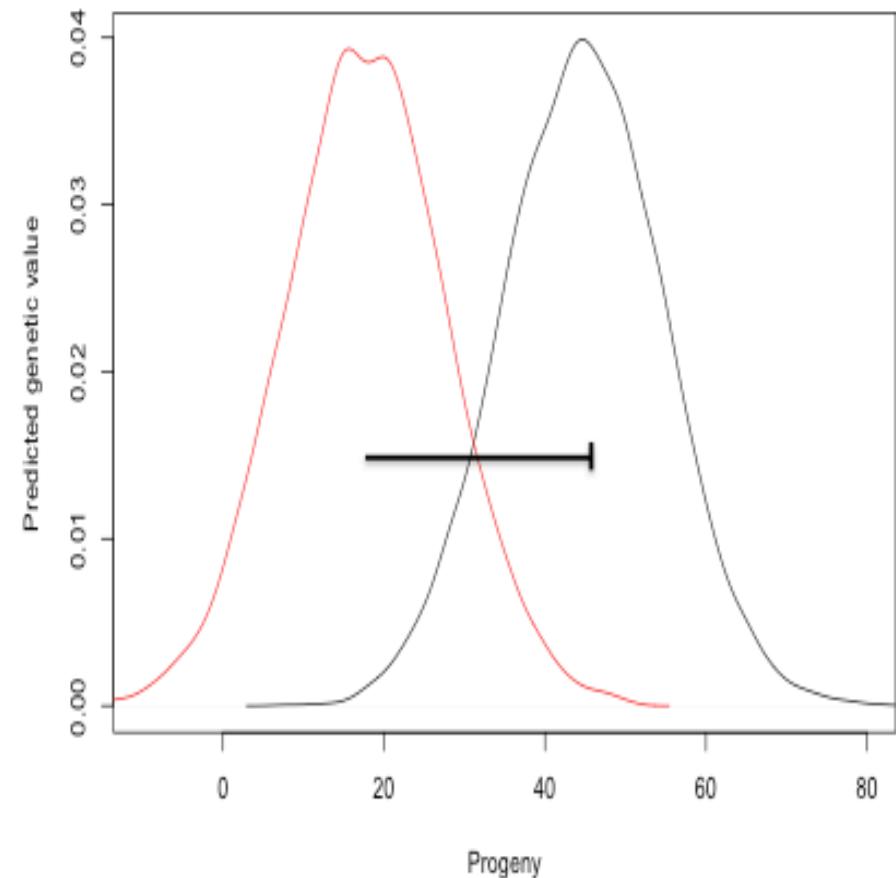
Seed production



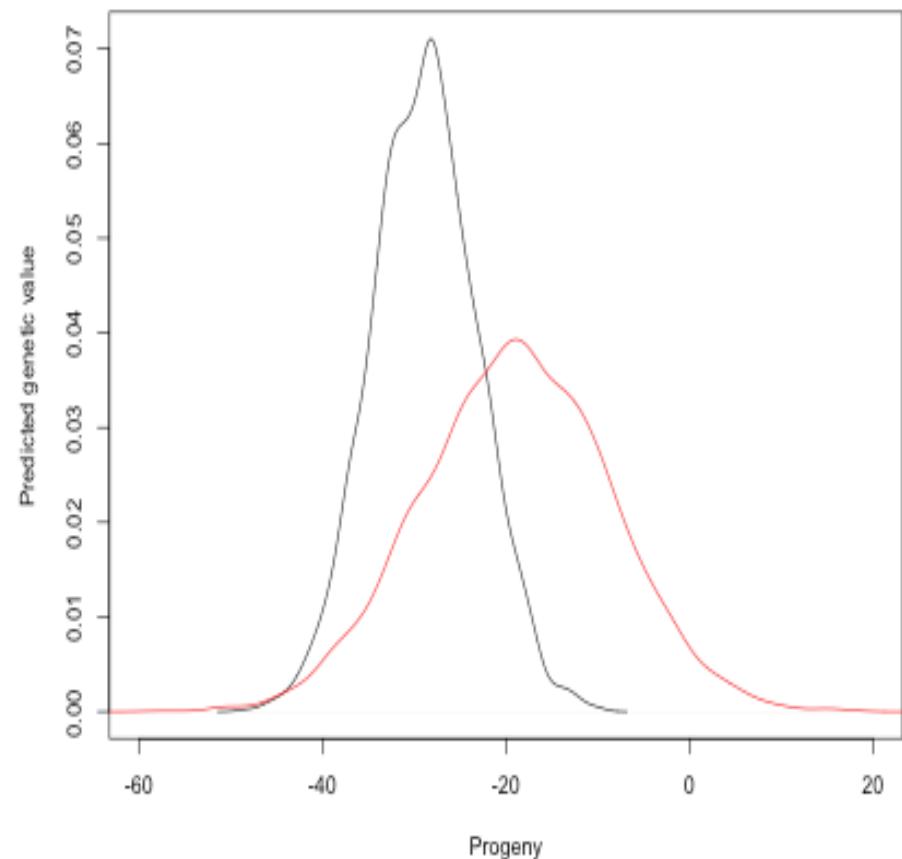
Breeding

GWS to Guide Crosses

Predicted distribution for two given crosses considering dominance effects



Predicted distribution for two given crosses considering dominance effects



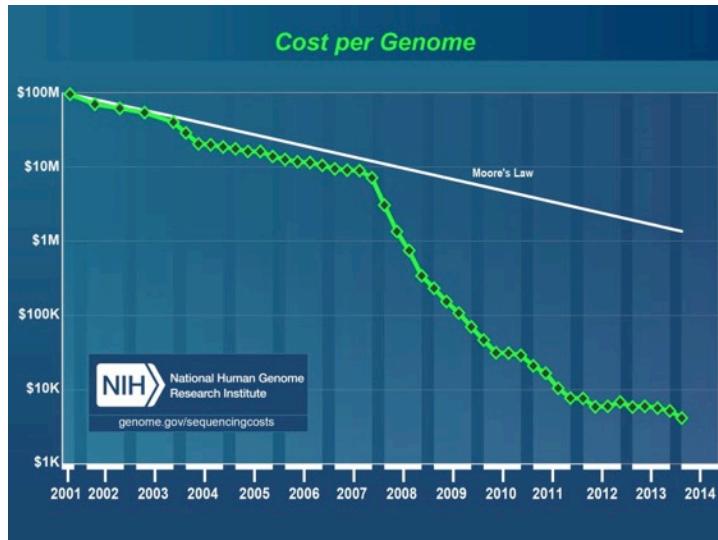
USDA – NIFA

“Accelerated Breeding By Improved Accuracy and Mate Allocation Using Genome-Wide Selection”

Family-Bulk Genome-wide Prediction

Goal: perform predictions using bulk genotyping and phenotyping

In many species “cultivars” corresponded to a population of genotypes (e.g. forages).



Even when genotyping cost has decreased significantly using NGS, it is still too high for most breeding programs.

Model Species and Method

Pine breeding (CCLONES) population with 71 families and genotyped with Chip for ~4,700 SNPs markers

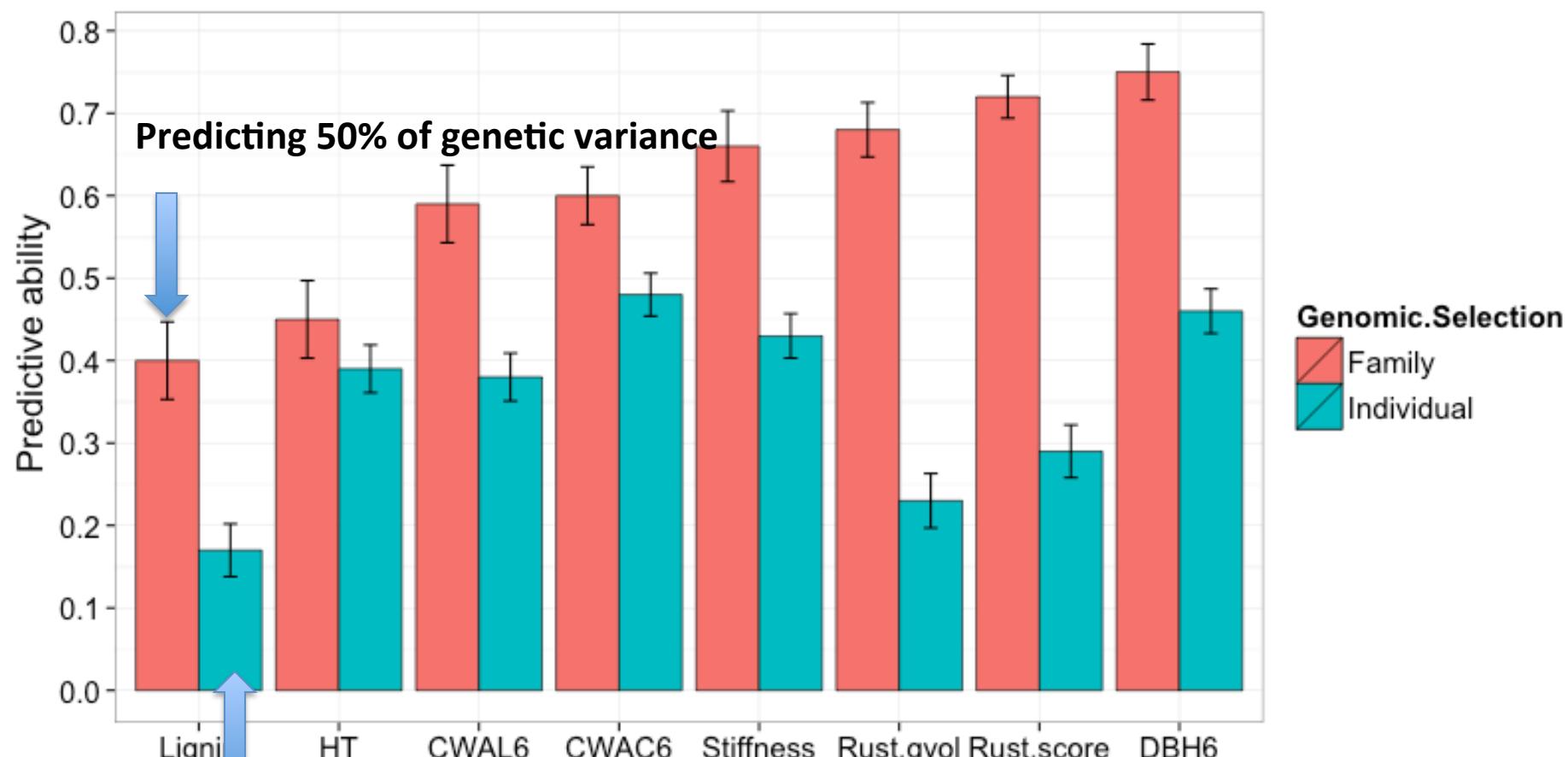


Phenotype: family phenotype mean

Genotype: family allele frequency



Family-Bulk Genotyping vs Traditional GWS Pine

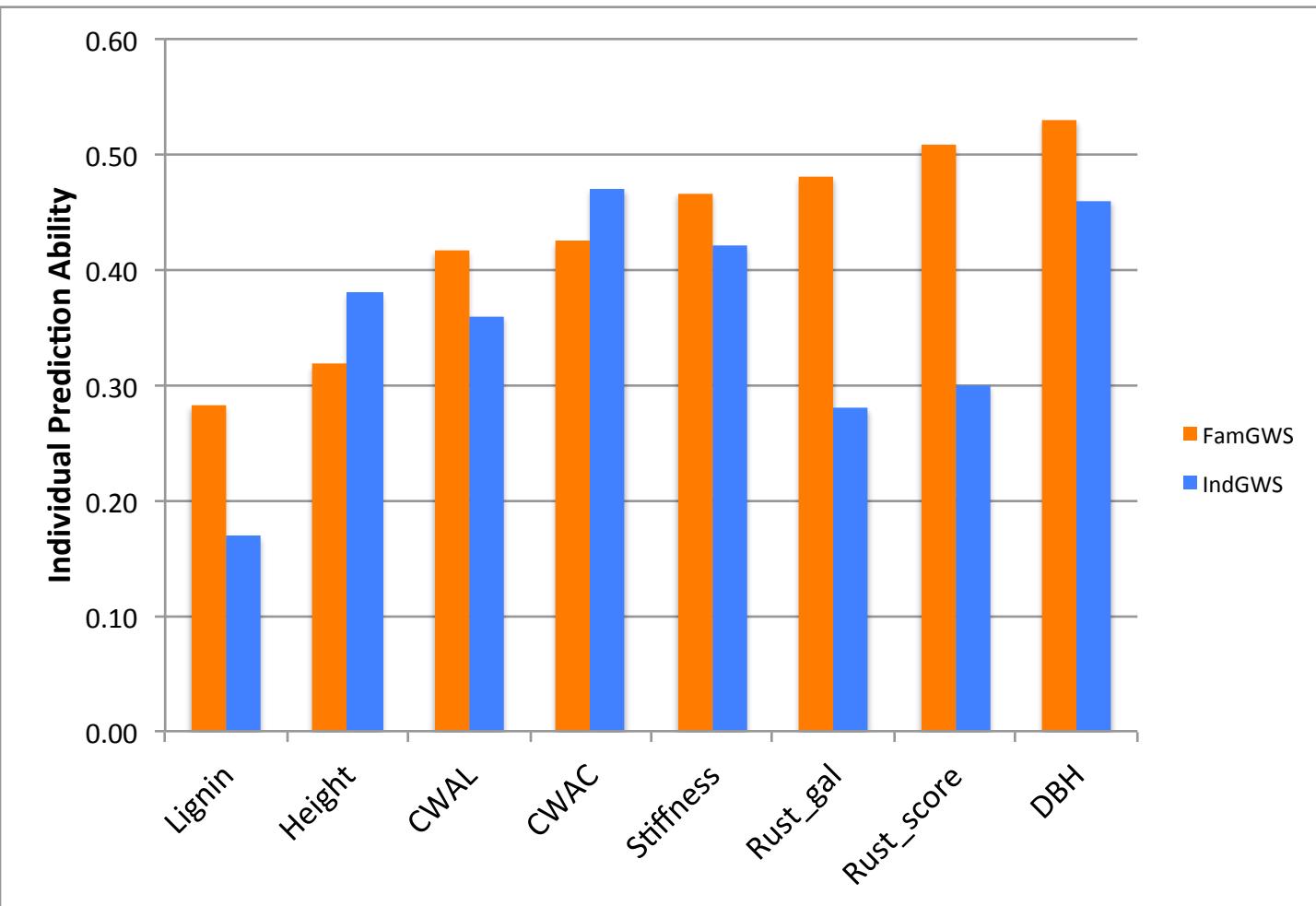


Capitalizing information at the mean level, which is more precise than at the individual level. But...
Predicting 100% of genetic variance

Family-Bulk Genotyping vs Traditional GWS Pine

To n
fam

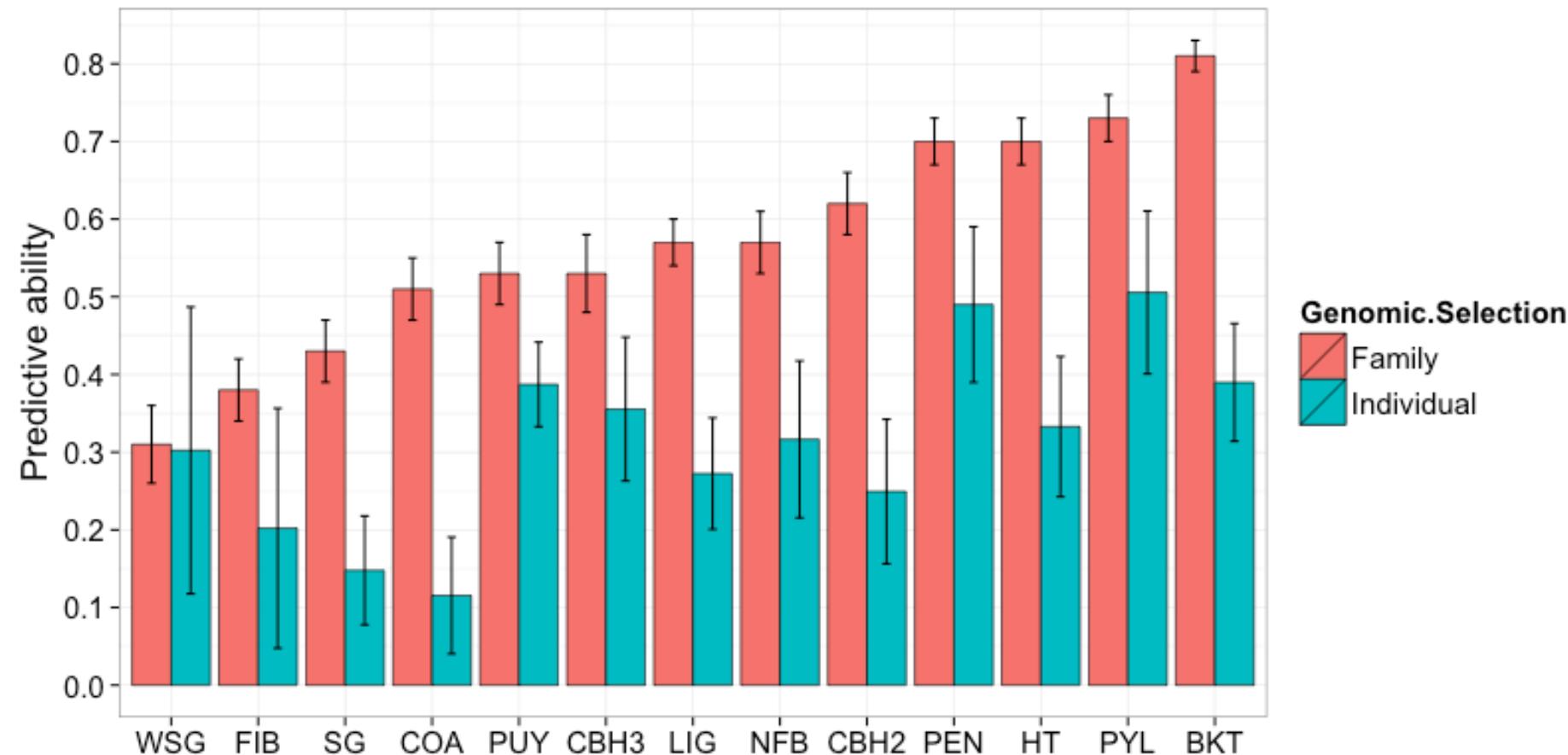
e



Still better in 6 out of 8 traits

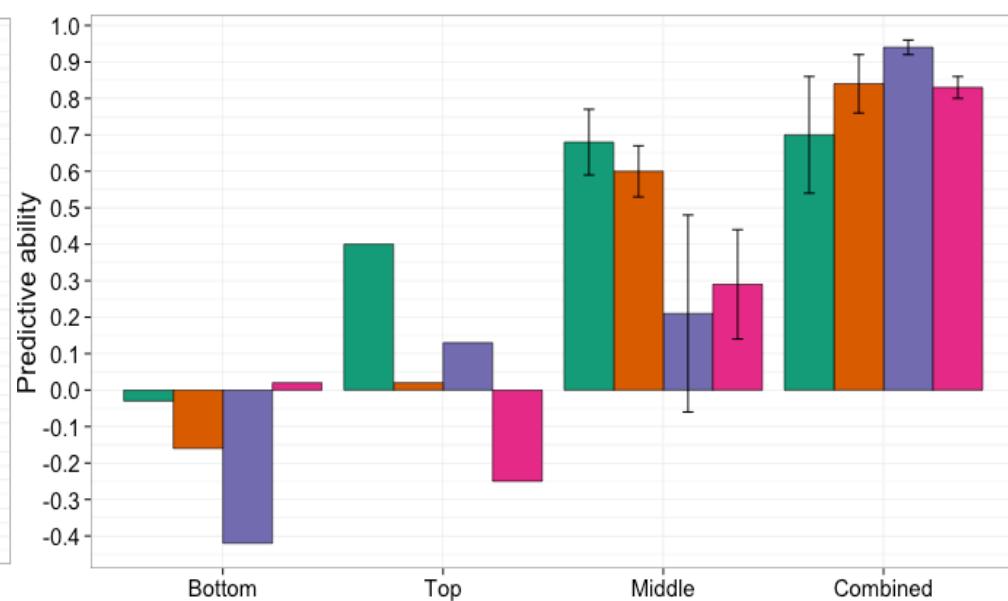
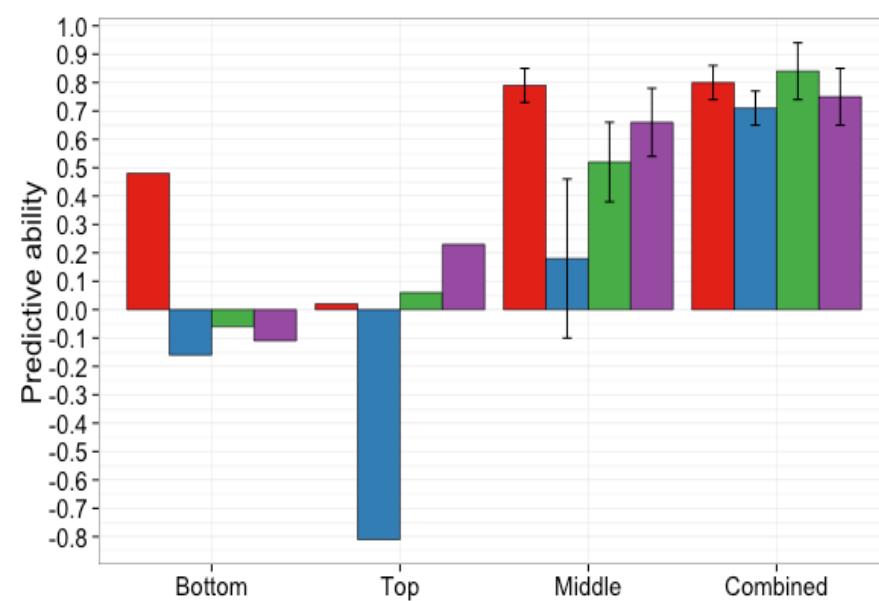
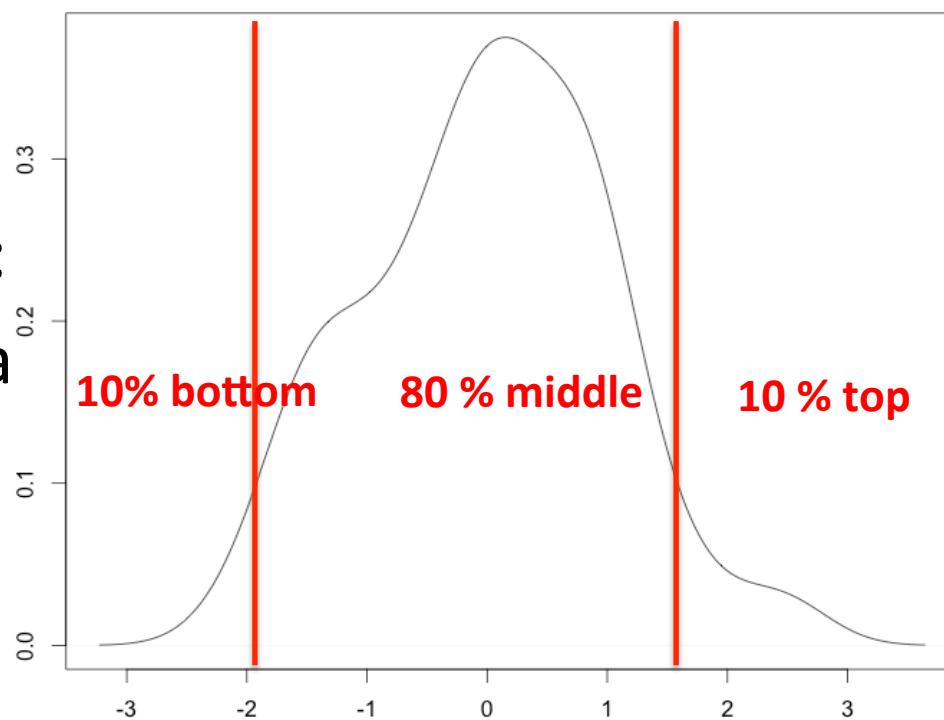
Family-Bulk Genotyping vs Traditional GWS Eucalyptus

Eucalyptus breeding population with 68 families and genotyped with sequence capture for 500,000 SNPs markers.



Still better in 8 out of 13 traits at the individual level

Challenges:
Very contra
ation folds



Summary

- An important investment will be needed to genotype a large number of individuals for GWS, thus maximizing the use of the markers is a **MUST**
- Using the genomic information we can:

Construct GWS models

Pedigree corrections

Genetic architecture

Predicting Crosses