Genomic Selection and Genome-Wide Association Study in *Populus trichocarpa* and *Pinus taeda*

Abdullah Kaplan

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Forestry

Thesis committee:

Jason Holliday

Harold Burkhart

M. A. Saghai Maroof

August 4, 2016

Blacksburg, VA

Keywords: Genomic Selection, GEBV, GWAS, Pinus taeda,Populus trichocarpa

ABSTRACT

Forest tree breeding methods rank among the most efficient ways to increase productivity and quality of forests. With the advent of high-throughput genotyping technology, genome-enabled breeding has started to gain importance and may overcome some weaknesses of traditional tree breeding. Genomic Selection (GS), which involves using genome-wide markers to predict breeding values of individuals in a population, has been proposed for animal and plant breeding programs. GS enables very accurate selection decisions through estimation of genomic estimated breeding values (GEBVs). While the goal of GS is to predict phenotype from genotype, it does not identify the underlying genes that have important roles in a trait. Genome-Wide Association Studies (GWAS) approaches are therefore complementary to GS, enabling identification of these genes, which may be useful for marker-assisted selection in some traits.

In this study, we first estimated heritability for several adaptive traits (cold hardiness, dbh, bud flush, height, and bud set) in a population of *Populus trichocarpa* and for height, diameter, and stem straightness in *Pinus taeda*. GEBVs accuracies were estimated using a ridge regression–best linear unbiased prediction (rrBLUP) model, and these accuracies were compared with estimated heritabilities. GWAS was also performed for the both imputed and non –imputed data of *P. taeda* population using TASSEL (Trait Analysis by aSSociation Evolution and Linkage) software, as well as rrBLUP and FFBSKAT (Fast Family-Based Sequence Kernel Association Test) packages in R.

Heritabilities ranged from 0.34 to 0.56 for *P. trichocarpa* and 0.14 to 0.37 for *P.taeda.* GWAS identified 3244 associations for dbh, 4077 associations for stem straightness, and 5280 SNPs for height *($p \leq 0.05$)* in TASSEL using the reduced model (marker data only), whereas 2729, 3272

and 3531 associations were found with the full model where we also included population structure as a covariate. FFBSKAT showed a similar number of SNP associations (2989, 3046 and 3058). There was an inflation of SNP associations (~20k) found in rrBLUP, which suggests population structure was not effectively controlled.

The GEBVs accuracies ranged from 0.09 and 0.22 for *P.trichocarpa* and 0.09 to 0.23 for *P.taeda* using rrBLUP method. Testing the effect of repetation on the accuracy of GEBV for poplar showed that there was no significant difference between the number of cycles. Also, there was no significant difference the accuracy of GEBVs in pine between two different imputation methods, the marker mean value and Beagle software.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would first like to thank my committee supervisor Dr. Jason Holliday for his support, encouragement, patience, guidance, and commitment to the success of my master thesis. I also would like to express my gratitude towards my committee members, Harold Burkhart and M. A. Saghai Maroof for their valuable contribution to my thesis development, assistance, and constructive feedbacks.

I would also like to thank the members of Holliday and Brunner Lab, Haktan, Rajesh, Mihir, Mandy, Bidisha, Xiaoyan, and Lecong, for their help and friendship. A special thank goes to Haktan and his wife, Leman, for their help, advice, encouragement, and providing a family atmosphere to work in.

This thesis was made possible thanks to the Ministry of National Education of the Republic of Turkey, which has given me this opportunity through their funding.

Finally, and most importantly, I would like to thank my mother, Sevim, and my father, Cavus, and my sisters, Dilek, and Hacer, believing in me and for always helping me reaching my goals.

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

AM	Association Mapping
BLUP	Best Linear Unbiased Prediction
DArT	Diversity Array Technology
DBH	Diameter At Breast Height
FFBSKAT	Fast Family-Based Sequence Kernel Association Test
GEBV	Genomic Estimated Breeding Value
GLM	General Linear Model
GS	Genomic Selection
GWAS	Genome-Wide Association Study
LD	Linkage Disequilibrium
MAS	Marker-Assisted Selection
MDS	Multidimensional Scaling
QTL	Quantitative Trait Loci
rrBLUP	Ridge Regression–Best Linear Unbiased Prediction
SNP	Single Nucleotide Polymorphisms
SSR	Simple Sequence Repeat
STRT	Straightness
TASSEL	Trait Analysis By Association Evolution And Linkage

I. Introduction

Forest trees species harbor a great deal of genetic diversity both within- and among-species. While substantial genetic improvement of tree populations has been achieved through conventional breeding, harnessing the abundance of natural variation for tree improvement has been hindered by large genomes and long generation times. Recent developments in genomics and associated statistical tools have the potential to significantly shortening the breeding cycle.

Over the past several decades, traditional plant and animal breeding techniques have been adopted, modified, and applied to many important tree species. In contrast with most plant breeding approaches, tree breeders must consider several unique ecological, population and quantitative genetics issues in selection programs. For example, the time involved in tree breeding from investment to harvest is much greater than other breeding programs (Libby, Stettler, & Seitz, 1969), and controlling external factors is difficult, and sometimes impossible. At the same time, forest tree breeding programs rank among the most efficient way to increase productivity and quality of planted forests (Namkoong, Kang, & Brouard, 1988). For these reasons, tree breeding strategies that leverage molecular genetic information, such as marker-assisted selection (MAS) and genomic selection, have gained importance in recent years (Muranty et al., 2014; Neale et al., 1992; Plomion, Durel, & Verhaegen, 1996).

Marker-assisted selection (MAS) is a type of indirect selection based on leveraging significant associations between a genetic marker and observable trait variation, and has been widely applied for the improvement of agricultural species (Collard, Jahufer, Brouwer, & Pang, 2005; Francia et al., 2005; Mohan et al., 1997; Xu & Crouch, 2008). The main idea of MAS is that

using Linkage Disequilibrium (LD) – the nonrandom association of alleles at distinct loci in a population – between markers and quantitative trait loci (QTLs), provides an indirect means of capturing desirable genetic attributes. Applying MAS to plants requires genotyping and testing large numbers of progeny to identify favorable alleles, and assumes that a small number of such alleles are of sufficient effect size as to be tractable during the selection phase (Collard & Mackill, 2008). Other methods of breeding that use markers and do not depend on LD include fingerprinting and pedigree verification or reconstruction (Muranty et al., 2014). Although MAS has been effective for selection using large-effect alleles with known association to a marker, it has been less successful when many alleles have a small impact on a trait. For this reason marker-assisted selection (MAS) has failed to significantly improve polygenic traits (Hospital, 2009; Xu & Crouch, 2008).

Genomic selection (GS) is an extension of marker-assisted selection (MAS) whereby genome-wide marker effects are first genotyped in a large training population with both phenotypic and genotypic data available, and prediction in subsequent populations relies on a model estimated in the training population (Jannink, Lorenz, & Iwata, 2010). Because GS captures the complete haplotype space of a population, it is suitable for improving quantitative traits in a way that MAS is not (Desta & Ortiz, 2014). The aim of GS is to provide breeders with a method of making selections in breeding programs and accelerating the breeding cycle (Lorenz et al., 2011). GS enables very accurate selection decisions through estimation of genomic estimated breeding values (GEBVs) (Nakaya & Isobe, 2012), which has become possible due to the large number of single nucleotide polymorphisms (SNP) that can be detected and scored by next generation sequencing (Jannink et al., 2010).

While the goal of GS is to predict phenotype from genotype, it does not explicitly identify the underlying genes that have important roles in specific traits. Association genetics is a complementary approach to GS that enables identification of specific genes or even specific sequence variants as the causative agents underlying phenotypic variation (McKown et al., 2014). Association genetic approaches have also been applied in some breeding programs, but their primary use has been in human biomedical genetics to elucidate the genetic basis of disease (Filiault & Maloof, 2012). Association mapping (AM), also called linkage disequilibrium (LD) mapping or association analysis, is a population based method used to detect and map QTLs based on the strength of the correlation between traits of interest and a marker (Jannink & Walsh, 2002). AM is based on historical LD to show trait –marker relations. Many economically important traits of forest trees are regulated by multiple genes, and GWAS approaches have been used to identify these genes (Harfouche et al., 2012).

Populus trichocarpa (black cottonwood) is one of the most suitable species among forest trees for association analysis due to its rapid growth, obligate outbreeding, and extensive variation in economically and ecologically important traits. Furthermore, *P. trichocarpa* has a complete genome sequence, which facilitates resequencing and assembly of data from many genotypes. Another important target species for association genetic studies is *Pinus taeda* (loblolly pine), which is among the most economically important tree species worldwide. Conifers in general and *P. taeda* in particular have very large genomes, and the sequence of *P. taeda* was recently reported (Zimin et al., 2014) at more than seven times that of the human genome (~21.15 billion pairs and ~21.6 Gb) (O'Brien, Smith, Gardner, & Murray, 1996).

The overarching objective of this study is to test genomic selection for *P. trichocarpa* and loblolly pine. Specifically, I estimated GEBVs accuracies for *P.trichocarpa* using approximately

1.7 million SNPs and five different traits for 369 individuals, and for *Pinus taeda* with approximately 60,000 SNPs and three different traits for 758 individuals using a ridge regression–best linear unbiased prediction (rrBLUP) model. As a second objective, I applied several GWAS models to identify significant markers related to the same traits for *Pinus taeda*.

II. Literature Review

II.1. Genome-Wide Association Study (GWAS)

GWAS is a method for analysing many common genetic variants among unrelated individuals to identify SNPs that have an association with a trait of interest. Association genetic research has revealed the causes of many diseases, as well as revealing the genetic basis of agriculturally important traits in both animal and plant species. In plants, QTL mapping is a major approach in bi-parental crosses. However this method has restrictions, including low allelic diversity and limited genomic resolution. GWAS has the potential to overcome these limitations. Uchiyama et al. (2013) identified three main advantages of GWAS over bi-parental QTL mapping: *i) GWAS does not require the development of a specific segregating population to detect QTL, (ii) GWAS can explore QTL controlling variations in a much larger and more representative gene-pool without any prior information about candidate genes, and (iii) GWAS is considered to provide much higher resolution than bi-parental QTL mapping, resulting in narrow confidence intervals for the loci detected*. Among the first GWAS in plants was a study in *Arabidopsis*, in which 107 phenotypes were examined, including shade avoidance (Filiault & Maloof, 2012), flowering time (Li, Huang, Bergelson, Nordborg, & Borevitz, 2010), heavy metal (Chao et al., 2012) and salt tolerance (Baxter et al., 2010). GWAS also was applied to 14 agronomic traits in rice using a genotyping-by-sequencing approach, and identified loci that represented almost 36% of the phenotypic variance, with six of these loci co-locating with previously identified genes (Huang et al., 2010). GWAS in tree species is a fairly new research area, but a few studies have been reported. One of the recent GWAS for *P. trichocarpa* (McKown et al. (2014) studied 40 different traits, including biomass, ecophysiology and phenology, and 29,355 SNPs, and identified 275 unique genes, in trait associations.

II.2. Marker Assisted Selection

It has been foreseen for over two decades that using molecular markers in plant and tree breeding programs should make possible accelerated gains from selection. In theory, MAS should be more efficient than traditional selection with strong enough LD (Hospital, 2009). Most breeders believe that MAS could help overcome drawbacks of the traditional forest tree breeding cycle. Compared with traditional breeding, MAS for forest trees is particularly preferable for four main reasons: "*i) the possibility of early selection, ii) the limitation of phenotyping costs, iii) the possibility of evaluating the genotypes of the breeding population more accurately by combining marker and phenotype information, iv) the possibility offered only by the use of molecular markers of monitoring inbreeding at the individual level and managing explicitly genetic diversity at the population level"* (Muranty et al., 2014). Although MAS has potential advantages over conventional breeding, markers will not necessarily be useful or more efficient for every trait (Neale et al., 1992). For most traits, effective phenotypic selection methods already exist, and these will often be less costly for selection in large populations. However, with whole-genome scans, many traits can be selected on the basis of genotype if the genetic control of the trait is understood. Developing MAS breeding methods have been proposed for pines, poplars, and eucalyptus (Muranty et al., 2014).

II.3. Genomic Selection

Fundamentally, GS is simply marker-assisted selection on a genome-wide scale. One early and visionary publication by Haley and Visscher (1998) outlined the possibility of making selections subject to genotypes from very dense genetic markers across the entire genome. The main idea that led these authors to consider new ways of using marker data was the realization that the

number of genetic markers would not always be limiting. It is important to note that the technological capability to discover and genotype high-density markers did not exist for crop, livestock, or forest tree species when this paper was published, but they had the foresight to see that the high-density marker genotyping technology could be developed and would have an important effect on the utility of markers for breeding programs. GS is a different approach from association analysis (mapping) since it does not seek to map the impact of specific genes, but instead has the goal of predicting phenotypes from multi-locus genotypes. GS uses random effect models to estimate the effect of all the markers without significance testing of individual variants to capture QTLs that explain small proportions of genetic variation that would not pass conventional significance thresholds (Heffner, Jannink, & Sorrells, 2011).

Estimating GEBVs of an individual from markers is the ideal selection criterion for GS, although GEBVs do not infer the function of the underlying genes. GEBVs based only on individuals' genotype perform well in simulation studies as well as in pilot studies (Kirst, Resende, Munoz, & Neves, 2011; Nakaya & Isobe, 2012), and their accuracy has held up in empirical studies of dairy cattle, mice, maize, oats, barley, and *Arabidopsis* (Jannink et al., 2010). In previous studies, GS for annual plants was analyzed by using different numbers of markers, training population ratio, and population size for different traits and species (Nakaya & Isobe, 2012). In addition, different analytical models have been proposed and used for prediction of GEBVs. For example, rrBLUP assumes that all marker effects are normally distributed and the marker effects have the same variance. Bayes A model assumes that marker effects are normally distributed with their own variance. Bayes B assumes the variance of most marker effects are equal to zero with probability π, and the remaining $(1-\pi)$ of all markers have non-zero effect with individual variance. Bayes C uses both shrinkage and variable selection methods and is characterized by a Gaussian

distribution. Finally, Bayes LASSO assumes that the regression coefficients have double exponential prior distributions (Habier, Fernando, Kizilkaya, & Garrick, 2011; Kärkkäinen & Sillanpää, 2012; M. F. Resende et al., 2012). Spindel et al. (2015) implemented GS on rice breeding lines with three traits and 73k SNPs using rrBLUP, and the mean accuracies of GEBVs were 0.29, 0.44, and 0.27 for grain yield, flowering time, and height, respectively. In the same study, using Bayes LASSO, the mean accuracies of GEBVs were found 0.11, 0.38, and 0.14 for the same traits. Previous studies suggested that larger training population size increases the accuracy of GEBVs (Lorenz et al., 2011; Solberg, Sonesson, & Woolliams, 2008). In addition, the accuracy of GEBVs depends on the marker types, heritability of traits, and number of QTLs (Lorenz et al., 2011; Nakaya & Isobe, 2012).

Several studies have demonstrated the effectiveness of GEBVs for prediction. Lorenzana and Bernardo (2009) obtained GEBV accuracies between 0.48 and 0.73 with 1339 SSRs on 8 morphological traits for maize (Table 1). In biparental population of wheat, Heffner et al. (2011) estimated the accuracy of GEBVs at 0.41 and 0.73 using the ridge regression model on 8-grain quality traits. In the same study, the avarage ratio of GS accuracy was 0.66 for training population size of 96 while it was 0.42 for training population sizes of 24. Another example from animal science by Mujibi et al. (2011) has shown that the accuracies of GEBVs for dairy cattle ranged between -0.07 and 0.48 (Table 1).

Interestingly, relativelty lower numbers of markers were used in plant empirical studies. The differences in accuracies may be because of the lower genetic diversity in animal species owing to the smaller number of parental lines (Nakaya & Isobe, 2012).

II.4. GS in Forest Trees

Interest in applying GS to tree breeding has grown in recent years. Early simulation studies (Grattapaglia & Resende, 2011) suggested that a marker density of 2 markers/cM would be adequate for small (Ne < 30) populations, and 10 or more markers/cM, to obtain enough power to mimic conventional phenotypic evaluation. Lorenzana and Bernardo (2009) suggested that, the loblolly pine breeding cycle can be shortened from 15 years to half that by employing GS. The authors attributed the success of SNP markers to tracing familial linkages rather than historical LD between SNP and trait QTL. They concluded that marker-QTL phase would likely not hold in the next generation because of recombinations, and because of the small population used to develop marker-trait associations. In the same study, empirical genotype data (3406 SNPs) was used to estimate GEBVs accuracies, which varied between 0.61 and 0.83 for wood lignin and cellulose content, while for growth traits such as height and volume, GEBVs ranged between 0.30 and 0.68 respectively. Resende et al. (2012) evaluated the efficiency of GS on lobloly pine using 4825 SNPs and GEBVs accuracies ranged from 0.65 and 0.75 for diameter and from 0.63 and 0.75 for height. In addition, Resende et al. (2012) analysed predictive ability of two different statistical methods (ridge regression, Bayes A, Bayes C, and Bayes LASSO) on 17 traits. They concluded that rrBLUP and Bayes A had lower predictive ability than Bayes C and Bayes LASSO for fusiform rust resistance due to the fact that few genes of large effect may control fusiform rust. In the same study, the accuracies of GEBVs were estimated as 0.38 and 0.46 for height and dbh respectively, and there was no statistical difference in the methods. In addition, GEBVs for growth architecture traits (crown width across the planting beds (CWAC), crown width along the planting beds (CWAL), branch angle average (BA), and average branch diameter (BD)) were very similar and not statistically different in the four statistical methods.

Table 1. Ranges of GEBVs accuracy investigated in empirical plant GS studies (Nakaya & Isobe, 2012)

Species	Training population ratio*	No. of genotyped markers†	Accuracy of GEBVs‡	Models for GEBV prediction	Traits	Reference
Maize	0.43, 0.65, 0.80	1339 SSRs and RFLPs	0.48–0.73	BLUP	8 morphological traits, 3 chemical components, grain moisture	Lorenzana and Bernardo (2009)
Maize	0.20, 0.40, 0.60, 0.80	1106 SNPs	0.26–0.57	RRBLUP	Three flowering traits	Guo et al. (2011)
Wheat	0.14, 0.28, 0.55	574 DArTs	0.41–0.73	RRBLUP	8-grain quality	Heffner et al. (2011)
Beef Cattle	not shown	37959 SNPs	-0.07-0.48	RRBLUP	Average daily gain, dry matter intake, residual feed intake	Mujibi et al. (2011)
Loblolly pine	not shown	3938 SNPs	0.64–0.77	BLUP	Diameter at breast height, total height	Resende et al. (2011)
Loblolly pine	Not shown	3406 SNPs	0.3–0.83	Pedigree model	Growth and quality traits	Isik et al. (2011)
Eucalyptus	783	3120 DArTs	0.53–0.69	BLUP	Height, diameter at breast height, wood density, pulp yield, lignin content	Grattapaglia et al. (2011)
Eucalyptus	0.9	3564 DArTs	0.54–0.62	BLUP	Puccinia rust resistance	

Resende et al. (2012) also tested the efficiency of GS in two eucalypt breeding populations. The accuracies of GEBVs ranged from 0.55- 0.88 (based on 3000 markers). Accuracies depend on the heritability of the trait studied, but these values for pine compare favorably with those obtained for crop species (Table 1).

II.5. Study species - *Populus trichocarpa*

Populus trichocarpa (black cottonwood) grows primarily on the wet sites in the western US and Canada, and is the largest of the North American poplars. *P. trichocarpa* is harvested and used for lumber, veneer, and fiber products. Importantly, *P. trichocarpa* is a preeminent lignocellulosic feedstock for bioenergy and has seen the intense development to this end in recent years. Many kinds of wildlife also have different uses such as foliage, twigs, and buds for food. Furthermore, black cottonwood is planted for shade and in windbreaks and shelterbelts. The range of *P. trichocarpa* extends from Kodiak Island along Cook Inlet in Alaska, through coastal British Columbia, Washington, and Oregon, to the mountains in southern California. While also found on the interior side of coastal mountain ranges, these populations are admixed with a sister species, *P. balsamifera* (Figure 1). The sequencing of 16 *P. trichocarpa* genomes revealed widespread patterns of LD and population structure (Slavov et al., 2012), and fine scale genecological studies have shown substantial adaptive phenotypic variation in growth (Pauley & Perry, 1954).

Figure 1.The range of *P. trichocarpa* (*http://esp.cr.usgs.gov/data/little/)* (Little Jr, 1971)

II.6. *Pinus taeda*

The range of *Pinus taeda* (loblolly pine) extends through 14 states from southern New Jersey, through east Texas and south to central Florida. *P. taeda* is the most commercially important forestry species in the southeast US, with about 11.7 million ha under cultivation. This constitutes over one-half of the standing pine volume in the southern United States. *P. taeda* is a medium-lived tree species, responds well to silvicultural treatments and can be cultivated as either even-aged or uneven-aged natural stands. The *P. taeda* genome was recently sequenced (Neale et al., 2014; Zimin et al., 2014) , and comprised of 20.15 billion base pairs. It is one of the largest genome of any species assembled to date (Main, 2014). Conifer genomes are known to

hold large amounts of repetitive DNA, and these elements constitute 82% of the genome in loblolly pine. The estimated number of genes is 50,172, and 15,653 of these have been experimentally validated (Wegrzyn et al., 2014).

Figure 2.Ranges of *Pinus taeda* *(http://esp.cr.usgs.gov/data/little)* (Little Jr, 1971)

III. Methods

III. 1. Plant material

III.1.1. *Populus trichocarpa*

Branch cuttings from 789 poplar clones were collected from a native range of the species, from Alaska to California, in 2010 to produce plantlets in a greenhouse (Oubida et al., 2015). After 6 months in the greenhouse, the plantlets were planted (May 2011) in a randomized block design with 4 blocks in a common garden at the Reynolds Homestead Forest Resource Research Center located in Patrick County Virginia (36° 37' N and 80° 09' W, altitude 359 m).The average of climate parameters between 1981 to 2009 for nearest weather station of Patrick County (VA) indicated a mean annual precipitation (MAP) as 1275 mm, a mean annual temperature (MAT) as 14.1 °C, a mean coldest month temperature as 4.2 °C, and a mean warmest month temperature as 24.4 °C (weatherbase.com). The common garden had drip irrigation and was controlled for weeds.

In the current study, 369 individuals were genotyped using sequence capture. For this, genomic DNA was extracted from poplar leaf tissues using a Qiagen DNeasy Plant Mini Kit (Qiagen, Inc, Valencia, CA). Sequence capture baits were designed for most exons in the poplar genome and synthesized by Agilent (Santa Clara, CA). These baits were hybridized in solution, and the resulting libraries were sequenced on an Illumina HiSeq instrument. After adapter sequence and low quality reads trimmed, short reads aligned to the poplar genome with the burrows-wheeler aligner, and variant calling is performed using SAMtools, which yielded ~1.7 million SNP markers.

Figure 3. Origins of *Populus trichocarpa* accessions used in this study (red points) (Oubida et al., 2015)

Five phenotypic traits were measured. Height (H) and stem diameter (22cm above ground level) were measured when the trees were dormant (March 2013) using a telescopic pole and digital caliper, respectively. Bud flush and bud set were first measured in April 2012 and September 2012, respectively, for apical buds, which were scored on a weekly basis over one month. Bud flush was defined as the date when a leaf had appeared 1 cm from the bud, and bud set was defined as a fully developed bud and covered by dark brown scales. The week at which the bud event (flushing or setting) occurred was converted to the Julian day (the number of days from the 1st January). The final trait, cold injury, was measured by electrolytic leakage on branch section obtained in October 2012 (Hannerz, Aitken, King, & Budge, 1999). Branch discs (1 - 2mm) were placed in solutions of 500 *u*l of distillated water and a trace of silver iodide (Agl) to aid ice nucleation. Control samples for all genotypes were kept at 4°C, and the temperature of the test samples were lowered at a rate of 4°C/hour until they reached -20°C, and then held there for 2

hrs. Following an initial conductivity measurement, all samples were then heat-killed at 95°C for 4 hours, and the conductivity measured once more. Freezing damage was estimated as an index of injury scale with the method of Flint, Boyce, and Beattie (1967):

$$I_t = 100(R_t - R_0) / (1 - R_0); Rt = L_t / L_k; R_0 = L_0 / L_d$$

where I_t denotes the index of injury, R_t and R_0 stand for relative conductivities for 20°C and 4°C respectively, L_t is the specific conductivity of leachate from sample frozen at -20°C, L_k is the specific conductivity of leachate from sample frozen at -20°C and then heat-killed, L_0 is the specific conductivity of leachate from non-frozen control samples (-4°), and L_d is the specific conductivity of leachate from non-frozen (-4°C) heat-killed samples (Flint et al., 1967).

III.1.2. *Pinus taeda*

The pine data were obtained from the Plantation Selection Seed Source Study (PSSSS) (McKeand & Bridgwater, 1998), which was established to study the performance, adaptability, and patterns of variation of loblolly families from various geographic sources, planted in different regions of the southeastern US. The Virginia test site, which is in the minimum cold hardiness zone for the species, was established on the Fleshman Tract in Appomattox County (Figure 3). In the current study, we sampled 758 individuals from this trial, which were genotyped for 60,902 SNP markers using genotyping-by-sequencing. To do this, DNA was extracted as above and digested with PstI/MspI restriction enzymes. Barcoded adapters were then ligated to the fragments and the resulting libraries were sequenced as above. The data were aligned with BWA to the loblolly pine reference genome and variants were called using the Genome Analysis Toolkit (GATK). The traits were provided by collaborators including height,

stem straightness, and dbh (diameter at breast height). The measurements were taken at the age of eight.

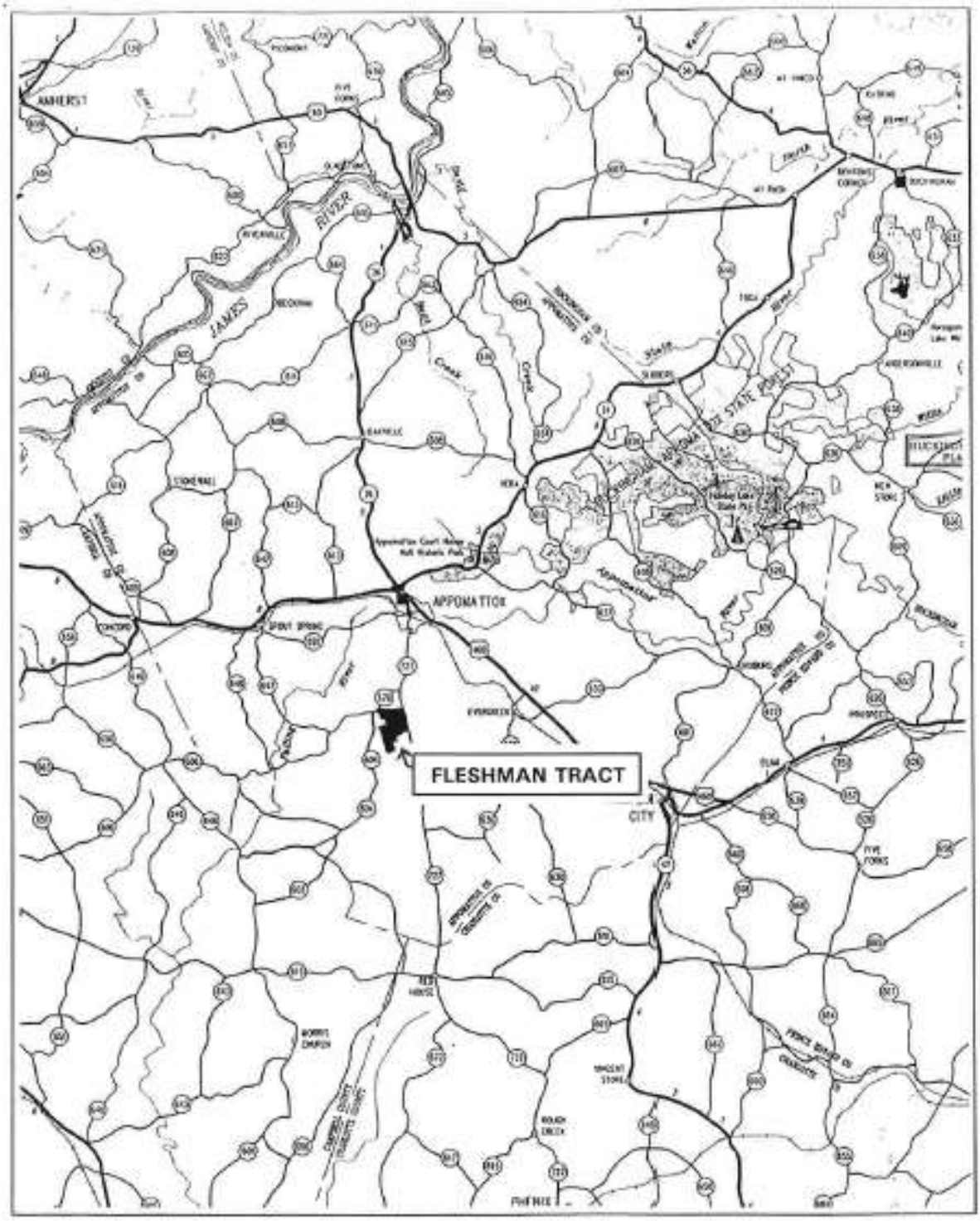


Figure 4. Location of the PSSSS trial on the Fleshmen Tract, Appomattox County, Virginia

III.2. Data Analysis

III.2.1. Heritability

Broad-sense heritability (H^2) was estimated for P. *trichocarpa* using a mixed model with block as a fixed factor and genotype and population as random factors to obtain the variance components of genotypes within and between populations. The linear model to calculate heritability is as follows:

$$Y_{jkl} = \mu + b_j + g_k + e_{jkl} \ ,$$

where Y_{jkl} is the phenotype of the l_{th} clone in the k^{th} block from the j^{th} genotype, μ is the grand mean, and e_{jkl} are the residuals. The broad-sense heritability for *P.trichocarpa* was calculated from estimates of total genetic variation (σ^2_G) and total phenotypic variation (σ^2_P) using the following formula: $H^2 = \sigma^2_G / \sigma^2_p$

Where σ^2_G is the genotype variance component and σ^2_p the total variance component.

Heritabilities for *P.taeda* were estimated with the following linear model:

$$y_{(ijk)} = F_{(i)} + b_{(j)} + Fb_{(ij)} + e_{(ijk)},$$

where F_i is the random general combining ability of the i^{th} family, b_j is the fixed j^{th} block effect, and $Fb_{(ij)}$ is the random i^{th} family by j^{th} site interaction effect since the formula of *P. trichocarpa* heritability calculation is based on clonal replication of trees.

The narrow sense heritability (h^2) (the other type is family heritability) was calculated as

$$h^2 = Va/(Vf+Vfb+Ve) = 4Vf/(Vf+Vfb+Ve)$$

Where Va is additive genetic variance, Vf is family variance, Vb is block variance, Vfb is family x block interaction, and Ve is error.

III.2.2. Genome-wide association study (GWAS)

GWAS was performed using TASSEL (Trait Analysis by aSSociation Evolution and Linkage) software, as well as rrBLUP and FFBSKAT (Fast Family-Based Sequence Kernel Association Test) packages in R for pine only. TASSEL is one of the most widely used software for association analysis. In TASSEL, GWAS was performed using general linear model (GLM) with population membership predictors as covariates to control for population structure (Bradbury et al., 2007). Population structure was estimated using the multidimensional scaling (MDS) (Zhu & Yu, 2009) package in R. MDS is similar to Principal Component Analysis (PCA) (Wang et al., 2009), which is a statistical method commonly used for reducing the dimensionality based on the level of similarity of individuals. In our analysis, the first ten dimensions extracted from MDS were used in the GLM model. In addition, to determine the effect size of SNPs on traits of interest, we estimated r^2 using TASSEL.

rrBLUP performs GWAS by solving the following mixed model

$$y = X\alpha + P\beta + K\gamma + \varepsilon$$

where y is the vector of measured phenoytpic values, X is the vector of SNP marker genoytpes, α is a vector of fixed effects that can model both environmental factors or population structure, β models the genetic background of each line as a random effect with

Var[g] = Qσ2, γ models the additive SNP effect as a fixed effect and ε is residual (Endelman, 2015). When we performed GWAS with rrBLUP, we used P3D (population parameters previously determined), introduced by Zhang et al. (2010). It is suggested that GWAS performs faster when P3D is enabled, however there is a tradeoff between speed and accuracy in terms of

using the previously calculated population parameters, such as genetic variance and residual variance. Finally, we used FFBSKAT, which is another R package for region/gene association analysis of quantitative traits. (Svishcheva, Belonogova, & Axenovich, 2014). FFBSKAT stands for kernel machine-based regression approach to the region-based association analysis, and is useful for identification of rare genetic variants for family-based or genetically related samples. FFBSKAT performs a score based variance component test to reveal the association between an SNP and traits of interest with additional covariates (population structure and kinship). After the models were fit, P-values were adjusted for multiple testing via the false discovery rate using the *qvalue* library in R. The number of associations were reported at using more liberal *(q<0.1)* and stringent *(q<0.001)* cutoffs in the result section.

III.2.3. Genomic Selection

The genomic prediction was performed using the R package ridge regression–best linear unbiased prediction (rrBLUP) (Endelman, 2011). rrBLUP assumes that all marker effects are normally distributed, and the marker effects have the same variance. The data for both species had nearly normal distributions (Figure S1 and Figure S2).

The basic rrBLUP model is as follows:

$$Y = \mu + Xg + e,$$

where Y is Nx1 vector of phenotypic means, μ is overall mean of the training set, X is the marker matrix, g is marker effects matrix, and e is Nx1 vector of residual effects (Endelman, 2011).

In rrBLUP, it is necessary to first designate the training and validation populations. The training data is used for fitting the model, which is subsequently used to predict phenotypic values in the validation population. For genomic selection we used different percentage of the training and validation populations since the percentage of training population affects on GEBVs.

We used 1.73 million and 60k SNP markers in poplar and loblolly pine, respectively. To test imputation accuracy, the pine SNP markers were imputed using both the A.mat function in rrBLUP, which uses the mean of each marker for imputation, as well as with Beagle software, which is a haplotype-based imputation method. For poplar, only Beagle was used for imputation. As the performance of any given run depends on the randomly selected training samples used, the mixed.solve function was repeated 10, 25, 75, and 100 times for poplar, and 100 times for pine. GEBVs were compared with the validation set via Pearson correlation. The correlation values were used as a measure of the efficiency of genomic selection for each trait in both species and called GEBVs accuracy.

IV. Results

IV.1. Heritability

To determine the extent to which the phenotypic variation is genetically controlled and predictive through GS, we initially estimated to broad sense heritability (H^2) for all traits in *P. trichocarpa* (Oubida et al., 2015), and narrow sense heritability (h^2) for all traits in *P. taeda.* Heritabilities ranged from 0.34 to 0.56 for *P. trichocarpa* and 0.14 to 0.37 for *P.taeda* (Table 2).

Table 2. The broad sense heritability (H^2) in *P. trichocarpa* and the narrow sense heritability (h^2) in *P.taeda* for studied traits

Populus trichocarpa	
Traits	H^2
Cold Hardiness	0.40
Dbh	0.34
Bud Flush	0.40
Height	0.42
Bud Set	0.56
Pinus taeda	h^2
Height	0.28
Stem Strt	0.14
Dbh	0.37

IV.2. Genome-Wide Association Study (GWAS)

GWAS was performed for all traits in both imputed and non-imputed data for *P.taeda* using TASSEL, rrBLUP, and FFBSKAT. The SNPs effects on the traits were small in general. For example, the mean effect size was highest for height using non-imputed data (0.0061), while it

was the lowest on when structure effects were included (0.0044) (Table 3, Figure S3 and Figure S4**)**. When the imputed data was used, the SNPs effect was lower than non-imputed data. Overall, the SNPs effect size reduced when the structure covariates was included in both imputed and non-imputed data.

Table 3.The mean estimated SNP effect size on traits in both data by using TASSEL

Imputed		
Traits	**M**	**M+S**
Height	0.0039	0.0031
Strt	0.0035	0.0031
Dbh	0.0032	0.0029
Non-Imputed		
Height	0.0061	0.0044
Strt	0.0058	0.0046
Dbh	0.0056	0.0052

For the imputed data, 3244 associations were found for dbh, 4077 associations for stem straightness, and 5280 SNPs for height *($p \leq 0.05$)* using the TASSEL reduced model (marker data only), whereas 2729, 3272 and 3531 associations were found in the full model, for which we also included population structure. FFBSKAT showed a similar number of SNPs associations (2989, 3046 and 3058). Interestingly, there was an inflation of SNPs association (~20k) found in rrBLUP for imputed data (Figure 5).

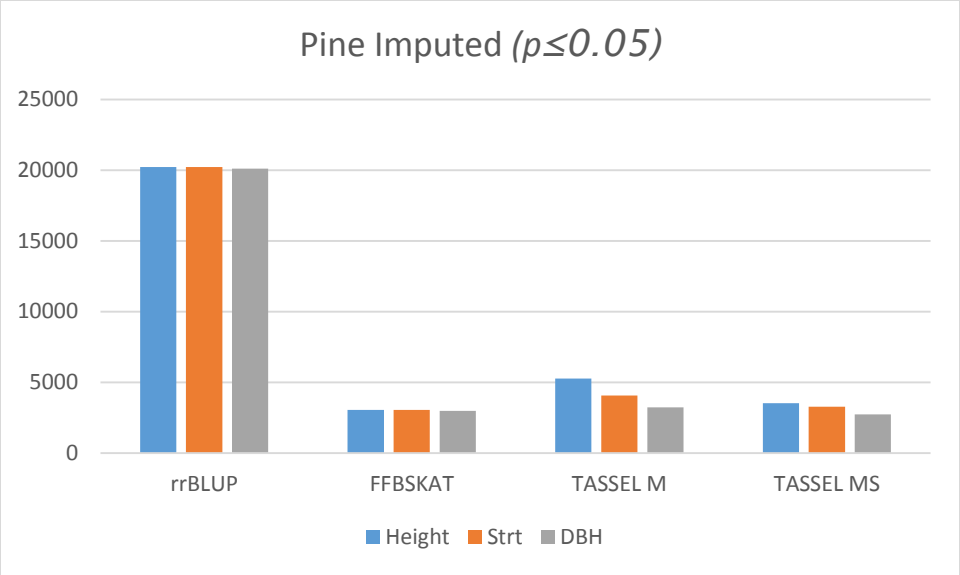


Figure 5. The Number of SNPs Significantly Associated with Traits for Pine Imputed Data (p≤0.05)

For non-imputed data, we found 3977 associations for dbh, 4349 associations for stem straightness, and 5025 associations for height statistically significant for reduced model whereas 3567, 3349, and 3362 SNPs were significant in full model in TASSEL respectively. In FFBSKAT, there were 2970 associations for dbh, 3081 associations for stem straightness, and 2889 associations for height significant. We discovered 6888 SNPs for dbh, 6900 SNPs for stem straightness, and 7043 SNPs for height significant associations (Figure 6).

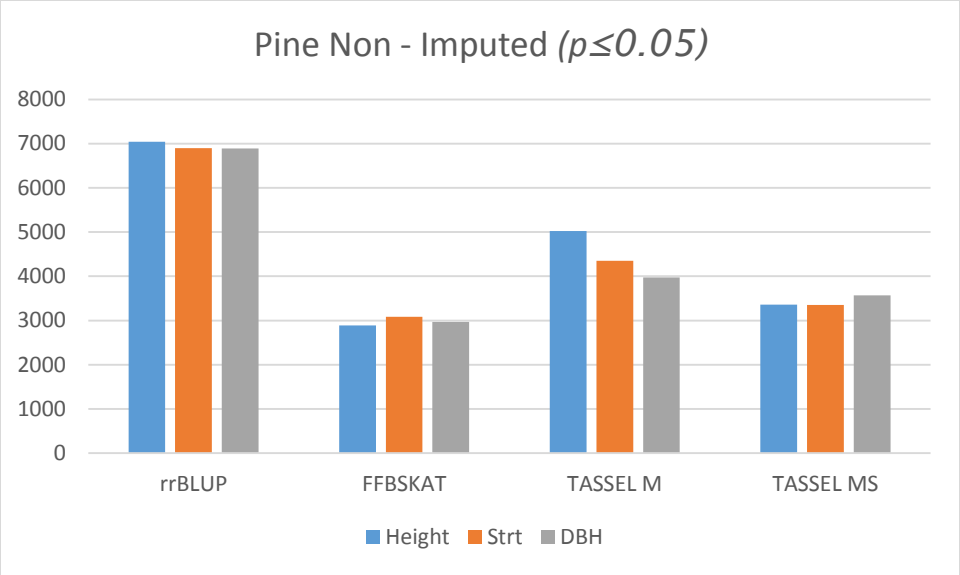


Figure 6. The Number of SNPs Significantly Associated with Traits for Pine Non-Imputed Data (p≤0.05)

After P-values were adjusted for multiple testing, we found 30 and 261 ($q \leq 0.1$) associations for stem straightness and height, respectively, using the reduced model in TASSEL. However, only four were statistically significant ($q \leq 0.1$) for stem straightness when we included population structure into markers (M+S), and there were two significant associations for height. In FFBSKAT, there was only one statistically significant marker in height by using imputed data. Using non-imputed data, we identified two markers that were statistically different. In rrBLUP, there was no statistically different association. The number of associations at different cut-offs as given in Figure S7, Figure S8 Figure S9, and Figure S10. Totally, we identified 329 association with all models in both imputed and non-imputed data and 22 of these are in common (Table S1).

The quantile-quantile plots showed that the FFBSKAT package performed slightly better than other models in controlling type I error for all traits in both imputed and non-imputed data (Figure 7 and 8) . In spite of the fact that TASSEL reduced model performed poorly for all traits

in both imputed and non-imputed data for controlling type I error for height and stem straightness traits in both data. Type I error of TASSEL reduced model decreased with imputed data, but it was still the worst model for height and stem straightness. On the other hand, all methods performed well for dbh.

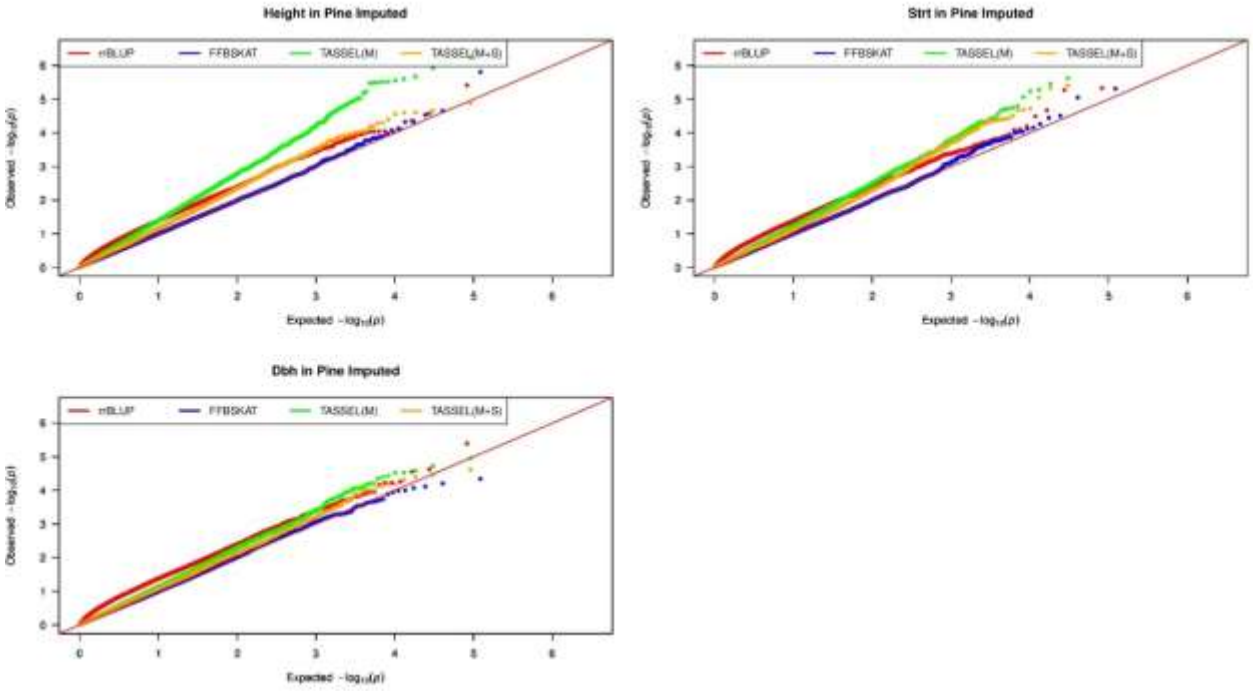


Figure 7. QQ Plots for GWAS results for imputed data

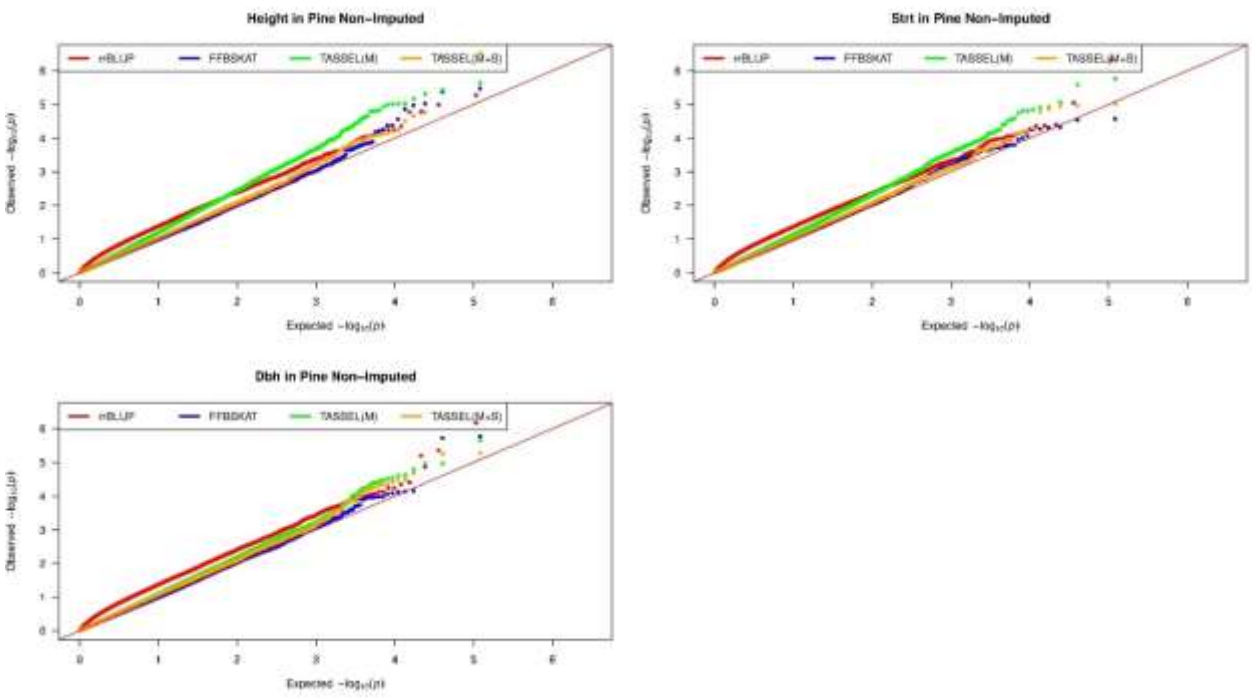


Figure 8. QQ Plots for GWAS results for non-imputed data

IV.3. Genomic Selection

Using 1.73 million SNP markers for five different traits in *Populus trichocarpa*, we compared the prediction accuracy estimated from rrBLUP model with different numbers of SNPs. Testing the effect of repetation on the accuracy of GEBV for poplar showed that there was no significant difference between the number of cycles (Table 4). Thus, 100 cycles were selected and used for detecting accuracy of GEBVs.

Table 4. Accuracy of GEBVs for different cycles in %60 training population

Cycles	Traits				
	Cold Hardiness	DBH	Bud Flush	Height	Bud Set
10	0.1033	0.1434	0.1769	0.2086	0.0615
25	0.0968	0.1443	0.1276	0.2357	0.0775
50	0.1014	0.1685	0.1462	0.2188	0.0901
75	0.0809	0.1547	0.144	0.2185	0.093
100	0.1033	0.1495	0.1543	0.2246	0.0864

GEBVs accuracies were estimated for each trait with 100 cycles. Height had the highest accuracy for all training population sizes, ranging from 0.08 to 0.23. GEBVs accuracies of bud set were the lowest up to %50 training population and ranged between 0.03 and 0.07, but increased with greater training population sizes. The accuracies of dbh and bud flush were similar and varied between 0.05 and 0.21. Finally, the accuracies of cold hardiness ranged from 0.04 to 0.11 (Figure 9).

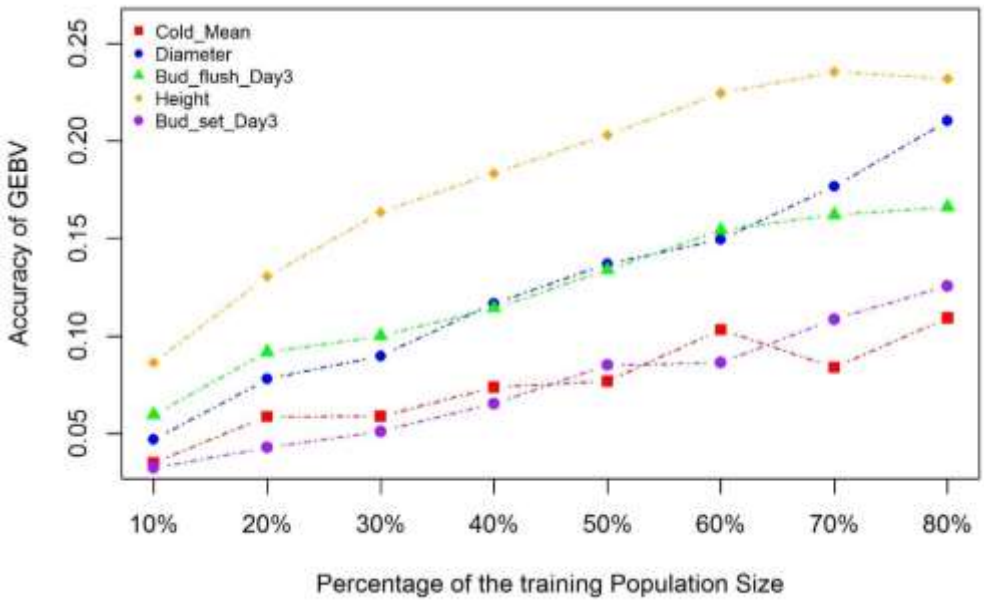


Figure 9. The Accuracy of GEBVs in Different Training Population Size with 100 Cycles

Overall, larger training population sizes resulted in higher accuracies. However, the accuracies of height and bud flush remained steady even though the accuracies of dbh and bud set showed a sharper increase after %60 training population. For the smallest training population, the accuracies were 0.03, 0.04, 0.05, 0.06, and 0.08 for bud set, cold hardiness, diameter, bud flush, and height, whereas they were 0.09 0.10, 0.15, 0.15, and 0.22, respectively, for the largest training population.

For three traits measured in *Pinus taeda* with 60k SNP markers both using mean imputation and Beagle software imputation, GEBVs showed a strong relationship between training population size and the accuracy of GEBVs. The accuracy of height using mean imputation varied between 0.09 and 0.24. The accuracy of dbh and stem straightness with mean imputation ranged from 0.04 to 0.08 and 0.07 to 0.16, respectively (Figure 10).

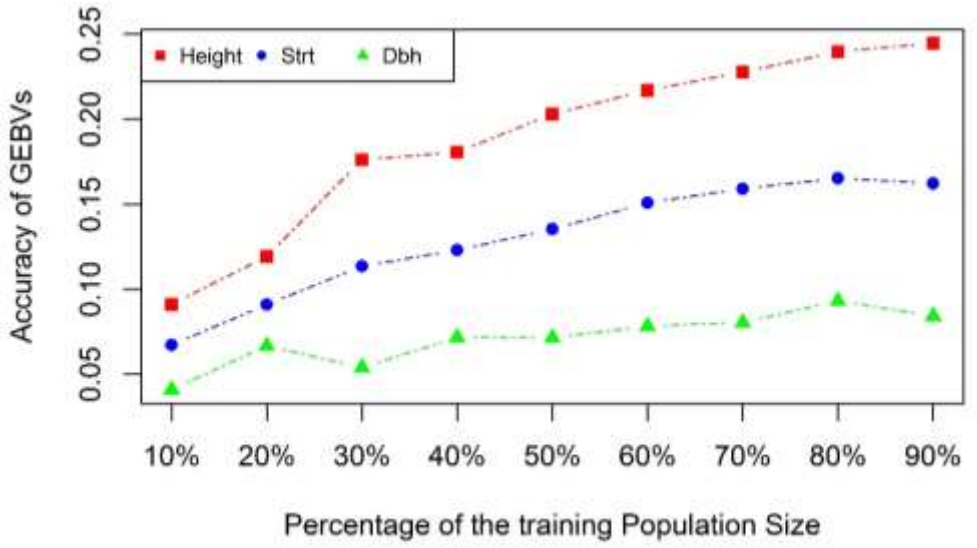


Figure 10. The accuracy of GEBVs in the Different Population Size with 100 Cycles for the Mean Imputation

The accuracy of GEBVs was higher using the Beagle imputation approach at 60% training population (Figure 11). However, testing population size effect was similar with mean imputation. Using Beagle imputation, the accuracy of GEBVs was between 0.12 and 0.26 for height, and between 0.11 and 0.22 for stem straightness. The accuracy of GEBVs for dbh ranged from 0.05 to 0.11. The highest accuracy of GEBVs with Beagle imputation was obtained at 80 percent training populations for both height and dbh although it was acquired at 70 percent for stem straightness (Figure 12).

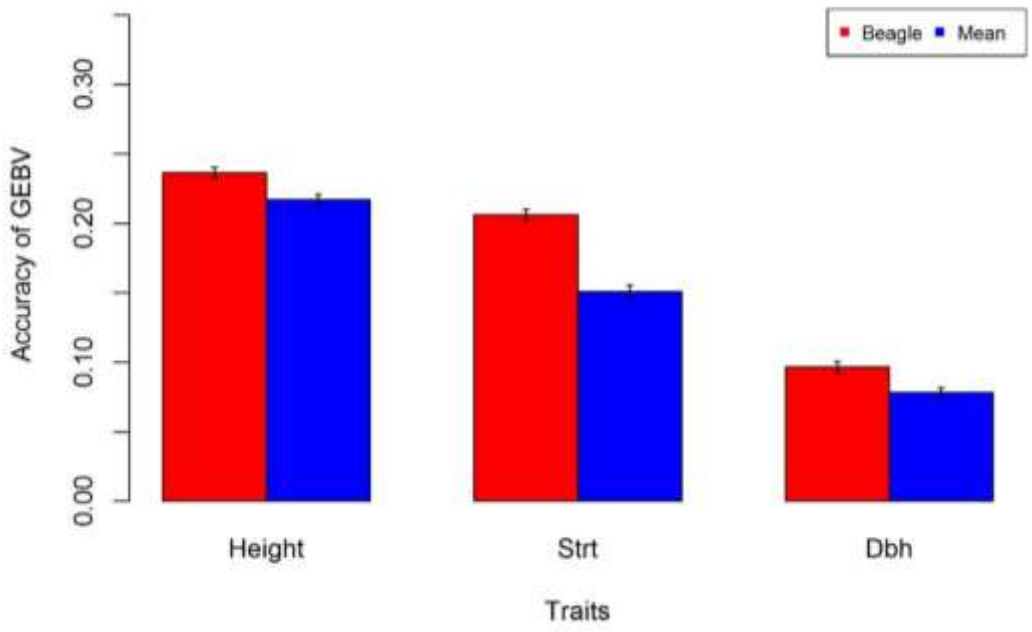


Figure 11. Accuracy of GEBVs in the Different Imputation Methods with 60% Training Population

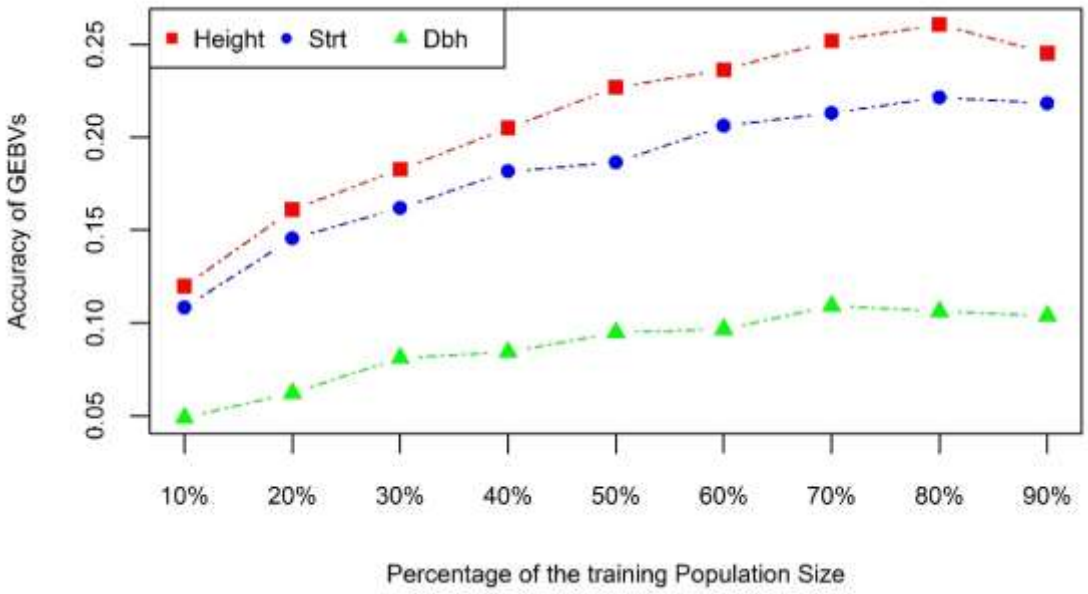


Figure 12. Accuracy of GEBVs in the Different Population Size with 100 Cycles for the Beagle Imputation

IV. Discussion

Completion of the genome sequences of major forest tree species has enabled truly genome-scale approaches to identify the genetic determinants of quantitative traits. *P.trichocarpa* was the first forest tree genome to be sequenced (Tuskan et al., 2006), while the *P. taeda* genome has only recently been released in draft form (http://pinegenome.org/pinerefseq). Sequencing of these and other tree genomes enables discovery of genes controlling economically or ecologically important traits through GWAS and the development of genomic tools to assist with selection in breeding programs.

The application of GS to tree species, if successful, will have a great impact on tree breeding, which has historically been logistically difficult and time-consuming. Tree improvement can only be effective where variation in a trait is genetically controlled, as opposed to being governed by the environment. We therefore first estimated heritability for *P. trichocarpa* phenology traits with 369 individuals – bud set, bud flush, and cold hardiness – and found a high proportion of variation – 0.56, 0.40, and 0.40, respectively – can be explained by genetics. Similarly, the heritability of growth traits (height and dbh) were 0.42 and 0.34. These values are of a similar range, although somewhat lower, to those reported by McKown et al. (2014), who found the mean broad sense heritability of *P. trichocarpa* in different years for bud set, bud flush and height was 0.74, 0.88, and 0.65, respectively. This difference may be attributable to the different populations and environment from which the heritability were estimated. We also estimated heritability for *P.taeda* height and dbh as 0.28 and 0.37 with 758 individiuals. While lower, which likely partially reflects the more narrow genetic base sampled in this population, this suggests potential for improving these traits. Xiang, Li, and Isik (2003) found similar values for heritability of these traits in *P.taeda* (0.22 and 0.19 for height and dbh at age 8). The

difference may be due to the sample size used, and Hill (2013) suggested that larger sample sizes improve heritability estimates since the standard error is decreased in proportion to family size.

For association studies of quantitative traits, markers effect sizes are frequently very low, which supports the polygenic inheritance of such traits (Stringer, Wray, Kahn, & Derks, 2011). Our results are consistent with this. We found mean SNPs effect sizes for height, stem straightness, and dbh varied between ~3-4%. When population structure was included, these effect sizes decreased somewhat. The reason of this reduction is that some of the variance explained by the simple model (marker only) may be due to population structure. On the other hand, including population structure as a covariate may result in overcorrection.

In all models, dbh had the lowest number of significant SNP associations *($p \leq 0.05$)* even though height was the highest. In FFBSKAT, surprisingly height had the lowest number of SNP association with 2889 SNPs. Among the models, rrBLUP had the most number of significant SNPs in both data. After we adjusted P-values for multiple testing the number of associations in all models sharply decreased, especially in rrBLUP, which suggests a high false positive rate for the simple model. However, our results suggest that using population structure is not enough to overcome type I error. Cappa et al. (2011) reported that using kinship had a more sufficient effect on the marker-trait associations. In addition, the imputation resulted in having more statistical significant association than non- imputed data after p –values adjusted. Cappa et al. (2013) found 962 associations (2364 DArTs markers versus six traits) before multiple testing corrections *($p<0.05$)* in *Eucalyptus globulus*. After correcting the rate of false discoveries, the number of markers – traits associations was reduced 18 (16 for dbh).

Given ample genetic variation, we sought to test genomic selection as a tool to predict growth and adaptability traits in *Pinus taeda* and *Populus trichocarpa*. M. F. Resende et al. (2012) reported that the number of SNP markers had a large effect on GEBVs accuracies. We found that increasing the size of training population size also improved GEBVs accuracy. With the smallest training population for *P. trichocarpa*, the accuracies were ~3-8%, whereas they were ~9-22% for the largest training population. The highest accuracy of GEBVs was obtained at 80 percent training populations for both height and dbh in *P. taeda*, although it was acquired at 70 percent for stem straightness. However, above approximately %60 training population size, accuracy plateaued for most of traits. We also analyzed the effect of the number of bootstrap samples, and found that there was no significant difference in the sample size (10, 25, 75, 100), which is consistent with M. F. Resende et al. (2012), who tested the effect of cross-validation using 10-fold and leave-one-out (N-fold) and found that the GEBVs accuracies were not significantly different. Poland et al. (2012) tested the effects of four different imputation methods, the marker mean value (mean), missing genotypes as heterozygotes, Random Forest regression, and a multivariate normal (MVN)-expectation maximization (EM) algorithm, for GS in wheat. They found that the (EM) algorithm gave more accurate imputation than heterozygous or mean imputation at the marker level, although there was no significant difference on the accuracy of GEBVs among imputation methods. We used two different imputation methods, the marker mean value and Beagle, for *P.taeda* and found that they were not significantly different. Interestingly, height had the best accuracy in both species and imputation methods using rrBLUP. Furthermore, dbh for both species had nearly the same accuracies. This may be due to our ability to accurately measure these traits compared with bud phenology, for example, which reduces sampling error and hence error in genomic prediction. Finally, GEBVs accuracies are

affected by the method of discovering and calling SNPs (Beaulieu, Doerksen, MacKay, Rainville, & Bousquet, 2014). When SNPs are pre-selected by sequencing and subsequently genotyped using SNP arrays, an ascertainment may be introduced, particularly where the discovery and prediction populations are different. We used sequence capture for *P.trichocarpa* and genotyping-by-sequencing (GBS) for *P.taeda,* which should mitigate such biases.

V. Conclusion

The results of this study are promising for the implementation of GS in forest trees breeding. For all traits of interest, increasing the size of training population size increased the GEBV accuracy. While these GEBVs did not quite match the heritabilities of the individual traits, they came close. Our GWAS corroborates the likely benefit of genome-wide selection as opposed to marker aided selection – effect size of individual SNPs was generally small, and corrections for multiple testing meant that many SNPs did not pass the significance threshold.

Future studies would benefit from additional sequencing depth to reduce the need for imputation (which was necessary because missing data is not tolerated for many GS algorithms). In addition, GS and GWAS should be implemented with more traits in multiple locations to get a sense for genotype x environment interactions that may impact the expression of the traits of interest.

References

Baxter, I., Brazelton, J. N., Yu, D., Huang, Y. S., Lahner, B., Yakubova, E., . . . Nordborg, M. (2010). A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1; 1. *PLoS Genet, 6*(11), e1001193.

Beaulieu, J., Doerksen, T. K., MacKay, J., Rainville, A., & Bousquet, J. (2014). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC genomics, 15*(1), 1.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics, 23*(19), 2633-2635.

Cappa, E. P.,El-Kassaby, Y. A., Garcia, M. N., Acuña, C., Borralho, N. M., Grattapaglia, D., & Poltri, S. N. M. (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in Eucalyptus globulus. *PloS one, 8*(11), e81267.

Cappa, E. P., Martínez, M. C., Garcia, M. N., Villalba, P. V., & Poltri, S. N. M. (2011). *Effect of population structure and kinship relationships on the results of association mapping tests of growth and wood quality traits in four Eucalyptus populations.* Paper presented at the BMC Proceedings.

Chao, D.-Y., Silva, A., Baxter, I., Huang, Y. S., Nordborg, M., Danku, J., . . . Salt, D. E. (2012). Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in Arabidopsis thaliana. *PLoS Genet, 8*(9), e1002923.

Collard, Jahufer, M., Brouwer, J., & Pang, E. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica, 142*(1-2), 169-196.

Collard, B. C.,& Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1491), 557-572.

Desta, Z. A., & Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science, 19*(9), 592-601.

Endelman, J., (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome, 4*(3), 250-255.

Endelman, J., (2015). Package 'rrBLUP'.

Filiault, D. L., & Maloof, J. N. (2012). A genome-wide association study identifies variants underlying the Arabidopsis thaliana shade avoidance response. *PLoS Genet, 8*(3), e1002589.

Flint, H., Boyce, B., & Beattie, D. (1967). Index of injury-a useful expression of freezing injury to plant tissues as determined by the electrolytic method. *Canadian Journal of Plant Science, 47*(2), 229-230.

Francia, E., Tacconi, G., Crosatti, C., Barabaschi, D., Bulgarelli, D., Dall'Aglio, E., & Vale, G. (2005). Marker assisted selection in crop plants. *Plant Cell, Tissue and Organ Culture, 82*(3), 317-342.

Grattapaglia, D., & Resende, M. D. (2011). Genomic selection in forest tree breeding. *Tree Genetics & Genomes, 7*(2), 241-255.

Guo, Z., Tucker, D. M., Lu, J., Kishore, V., & Gay, G. (2012). Evaluation of genome-wide selection efficiency in maize nested association mapping populations. *Theoretical and Applied Genetics, 124*(2), 261-275.

Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC bioinformatics, 12*(1), 1.

Haley, C., & Visscher, P. (1998). Strategies to utilize marker-quantitative trait loci associations. *Journal of dairy science, 81*, 85-97.

Hannerz, M., Aitken, S. N., King, J. N., & Budge, S. (1999). Effects of genetic selection for growth on frost hardiness in western hemlock. *Canadian Journal of Forest Research, 29*(4), 509-516.

Harfouche, A., Meilan, R., Kirst, M., Morgante, M., Boerjan, W., Sabatti, M., & Mugnozza, G. S. (2012). Accelerating the domestication of forest trees in a changing world. *Trends in plant science, 17*(2), 64-72.

Heffner, E. L., Jannink, J.-L., Iwata, H., Souza, E., & Sorrells, M. E. (2011). Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science, 51*(6), 2597-2606.

Heffner, E. L., Jannink, J.-L., & Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *The Plant Genome, 4*(1), 65-75.

Hill, W. G. (2013). On estimation of genetic variance within families using genome-wide identity-by-descent sharing. *Genetics Selection Evolution, 45*(1), 1.

Hospital, F. (2009). Challenges for effective marker-assisted selection in plants. *Genetica, 136*, 303-310.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., . . . Zhang, Z. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature genetics, 42*(11), 961-967.

Isik, F., Whetten, R., Zapata-Valenzuela, J., Ogut, F., & McKeand, S. (2011). *Genomic selection in loblolly pine-from lab to field.* Paper presented at the BMC Proceedings.

Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics, 9*(2), 166-177.

Jannink, J.-L., & Walsh, B. (2002). Association mapping in plant populations. *Quantitative genetics, genomics and plant breeding*, 59-68.

Kärkkäinen, H. P., & Sillanpää, M. J. (2012). Back to basics for Bayesian model building in genomic selection. *Genetics, 191*(3), 969-987.

Kirst, M., Resende, M., Munoz, P., & Neves, L. (2011). *Capturing and genotyping the genome-wide genetic diversity of trees for association mapping and genomic selection.* Paper presented at the BMC Proceedings.

Li, Y., Huang, Y., Bergelson, J., Nordborg, M., & Borevitz, J. O. (2010). Association mapping of local climate-sensitive quantitative trait loci in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences, 107*(49), 21199-21204.

Libby, W., Stettler, R. F., & Seitz, F. (1969). Forest genetics and forest-tree breeding. *Annual Review of Genetics, 3*(1), 469-494.

Little Jr, E. L. (1971). Atlas of United States trees. Volume 1. Conifers and important hardwoods. Miscellaneous publication 1146. *US Department of Agriculture, Forest Service, Washington, DC*.

Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., . . . Jannink, J.-L. (2011). 2 Genomic Selection in Plant Breeding: Knowledge and Prospects. *Advances in agronomy, 110*, 77.

Lorenzana, R. E., & Bernardo, R. (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theoretical and Applied Genetics, 120*(1), 151-161.

Main, D. (2014). SCIENTISTS SEQUENCE THE LARGEST GENOME TO DATE. *Popular Science*.

McKeand, S. E., & Bridgwater, F., E. (1998). A strategy for the third breeding cycle of loblolly pine in the southeastern US. *Silvae Genetica, 47*(4), 223-234.

McKown, A. D., Klápště, J., Guy, R. D., Geraldes, A., Porth, I., Hannemann, J., . . . Ehlting, J. (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. *New Phytologist, 203*(2), 535-553.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T., Yano, M., Bhatia, C., & Sasaki, T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular breeding, 3*(2), 87-103.

Mujibi, F., Nkrumah, J., Durunna, O., Stothard, P., Mah, J., Wang, Z., . . . Moore, S. (2011). Accuracy of genomic breeding values for residual feed intake in crossbred beef cattle. *Journal of animal science, 89*(11), 3353-3361.

Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., & Sanchez, L. (2014). Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. *Tree Genetics & Genomes, 10*(6), 1491-1510.

Nakaya, A., & Isobe, S. N. (2012). Will genomic selection be a practical method for plant breeding? *Annals of botany*, mcs109.

Namkoong, G., Kang, H.-C., & Brouard, J. S. (1988). *Tree breeding: principles and strategies*: Springer-Verlag.

Neale, D. B., Devey, M., Jermstad, K., Ahuja, M., Alosi, M., & Marshall, K. (1992). Use of DNA markers in forest tree improvement research *Population Genetics of Forest Trees* (pp. 391-407): Springer.

Neale, D. B., Wegrzyn, J. L., Stevens, K. A., Zimin, A. V., Puiu, D., Crepeau, M. W., . . . Liechty, J. D. (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome biology, 15*(3), R59.

O'Brien, I. E., Smith, D. R., Gardner, R. C., & Murray, B. G. (1996). Flow cytometric determination of genome size in Pinus. *Plant Science, 115*(1), 91-99.

Oubida, R. W., Gantulga, D., Zhang, M., Zhou, L., Bawa, R., & Holliday, J. A. (2015). Partitioning of multivariate phenotypes using regression trees reveals complex patterns of adaptation to climate across the range of black cottonwood (Populus trichocarpa). *Frontiers in plant science, 6*.

Pauley, S. S., & Perry, T. O. (1954). *Ecotypic variation of the photoperiodic response in Populus*.

Plomion, C., Durel, C., & Verhaegen, D. (1996). Marker-assisted selection in forest tree breeding programs as illustrated by two examples: Maritime pine and eucalyptus. *Annales des Sciences Forestieres (France)*.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., . . . Sorrells, M. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome, 5*(3), 103-113.

Resende, M., Munoz, P., Acosta, J., Peter, G., Davis, J., Grattapaglia, D., . . . Kirst, M. (2012). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist, 193*(3), 617-624.

Resende, M. D., Resende, M. F., Sansaloni, C. P., Petroli, C. D., Missiaggia, A. A., Aguiar, A. M., . . . Faria, D. A. (2012). Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist, 194*(1), 116-128.

Resende, M. F., Muñoz, P., Resende, M. D., Garrick, D. J., Fernando, R. L., Davis, J. M., . . . Kirst, M. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.). *Genetics, 190*(4), 1503-1510.

Slavov, G. T., DiFazio, S. P., Martin, J., Schackwitz, W., Muchero, W., Rodgers-Melnick, E., . . . Pennacchio, L. A. (2012). Genome resequencing reveals multiscale geographic structure and extensive linkage disequilibrium in the forest tree Populus trichocarpa. *New Phytologist, 196*(3), 713-725.

Solberg, T., Sonesson, A., & Woolliams, J. (2008). Genomic selection using different marker types and densities. *Journal of animal science, 86*(10), 2447-2454.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., . . . McCouch, S. R. (2015). Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet, 11*(2), e1004982.

Stringer, S., Wray, N. R., Kahn, R. S., & Derks, E. M. (2011). Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PloS one, 6*(11), e27964.

Svishcheva, G. R., Belonogova, N. M., & Axenovich, T. I. (2014). FFBSKAT: fast family-based sequence kernel association test. *PloS one, 9*(6), e99407.

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., . . . Salamov, A. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *science, 313*(5793), 1596-1604.

Uchiyama, K., Iwata, H., Moriguchi, Y., Ujino-Ihara, T., Ueno, S., Taguchi, Y., . . . Watanabe, A. (2013). Demonstration of genome-wide association studies for identifying markers for wood property and male strobili traits in Cryptomeria japonica. *PloS one, 8*(11), e79866.

Varshney, R. K., Hoisington, D. A., & Tyagi, A. K. (2006). Advances in cereal genomics and applications in crop breeding. *Trends in biotechnology, 24*(11), 490-499.

Wang, D., Sun, Y., Stang, P., Berlin, J. A., Wilcox, M. A., & Li, Q. (2009). *Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling.* Paper presented at the BMC proceedings.

Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L.-S., Loopstra, C. A., Vasquez-Gross, H. A., . . . Martínez-García, P. J. (2014). Unique features of the loblolly pine (Pinus taeda L.) megagenome revealed through sequence annotation. *Genetics, 196*(3), 891-909.

Xiang, B., Li, B., & Isik, F. (2003). Time trend of genetic parameters in growth traits of Pinus taeda L. *Silvae Genetica, 52*(3-4), 114-120.

Xu, Y., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Science, 48*(2), 391-407.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., . . . Ordovas, J. M. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature genetics, 42*(4), 355-360.

Zhu, C., & Yu, J. (2009). Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics, 182*(3), 875-888.

Zimin, A., Stevens, K. A., Crepeau, M. W., Holtz-Morris, A., Koriabine, M., Marçais, G., . . . de Jong, P. J. (2014). Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics, 196*(3), 875-890.

Supplementary Materials

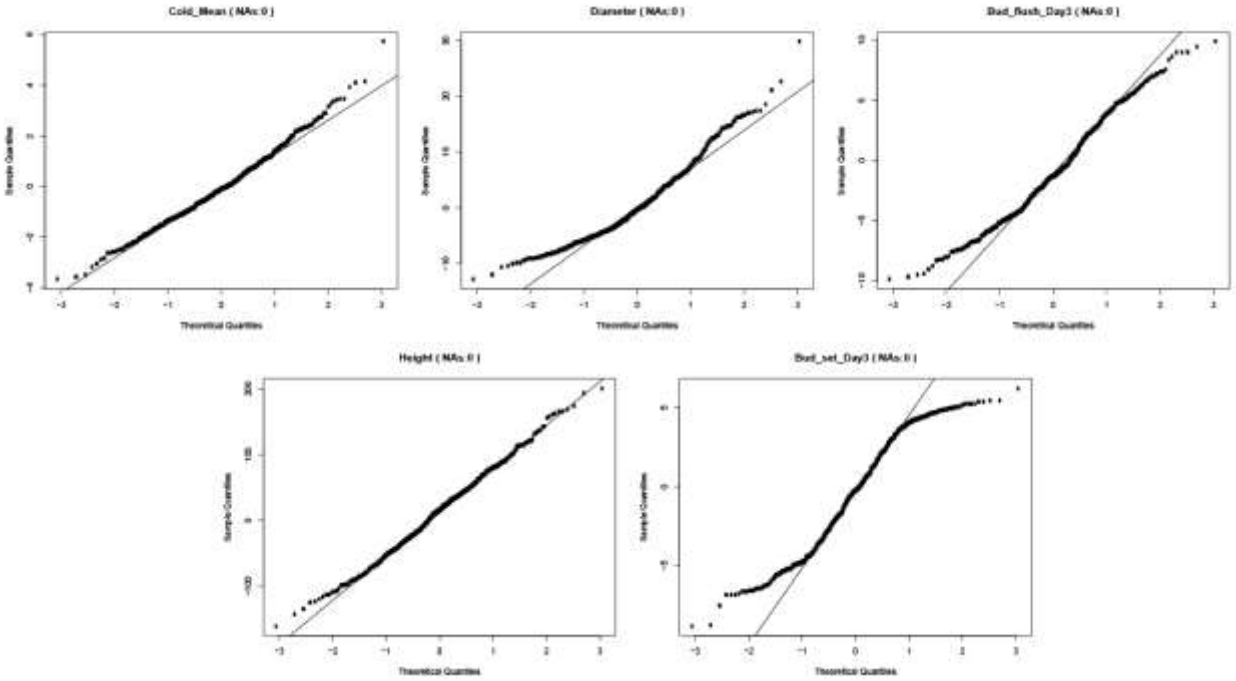


Figure S1. Distributions of the *Populus trichocarpa* phenotypes

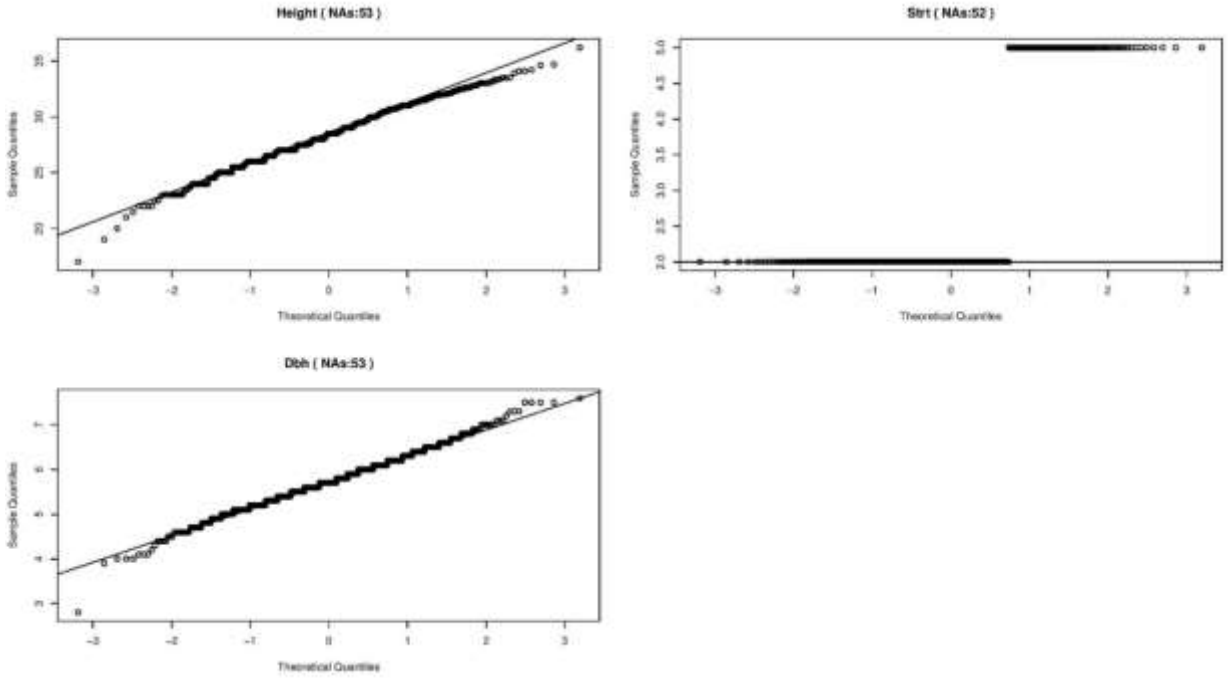


Figure S2. Distributions of the *Pinus taeda* Phenotypes

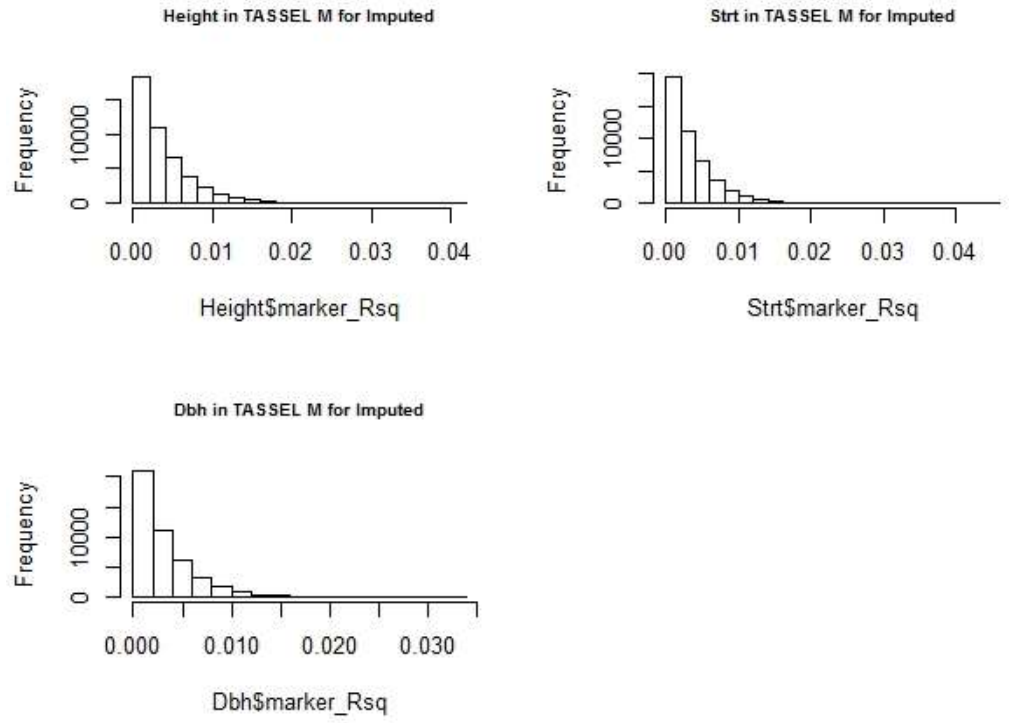


Figure S3. Distribution of SNPs effects on traits for imputed data (only markers)

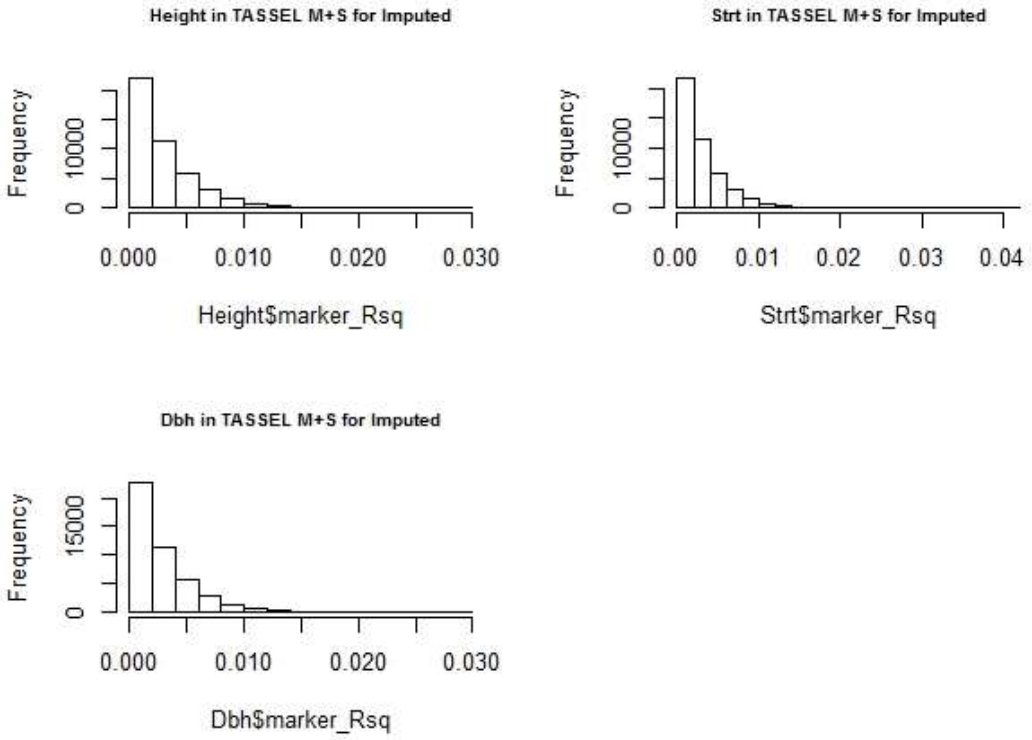


Figure S4. Distribution of SNPs effects on traits for imputed data (markers+ structure)

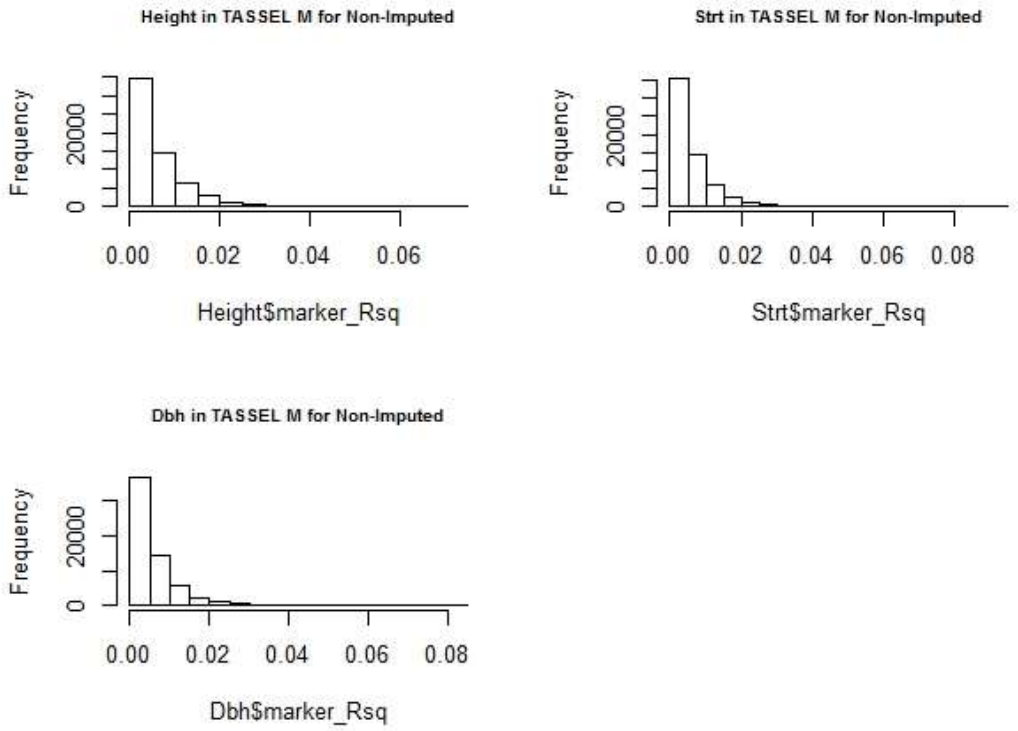


Figure S 5. Distribution of SNPs effects on traits for non-imputed data (only markers)

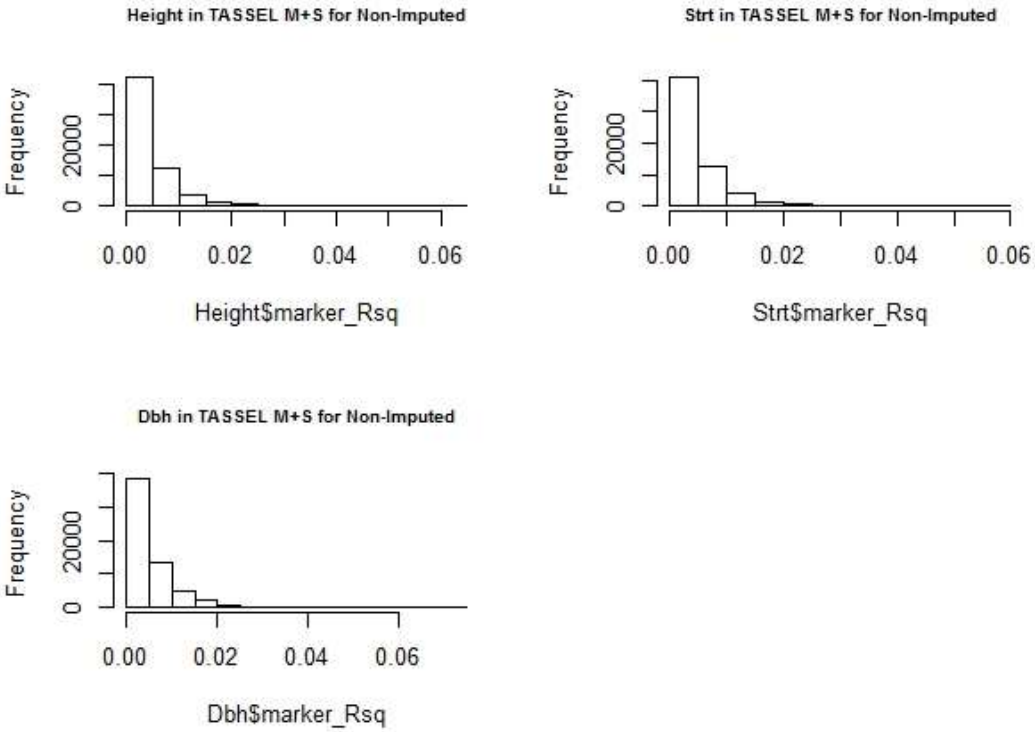
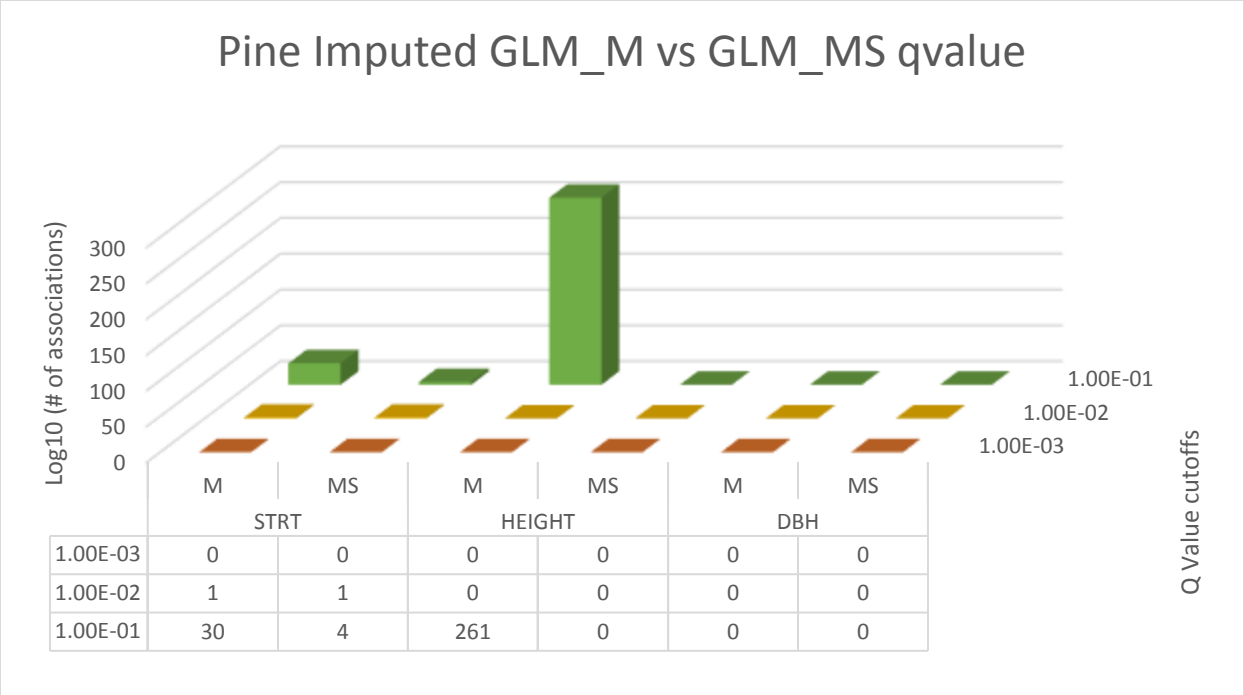
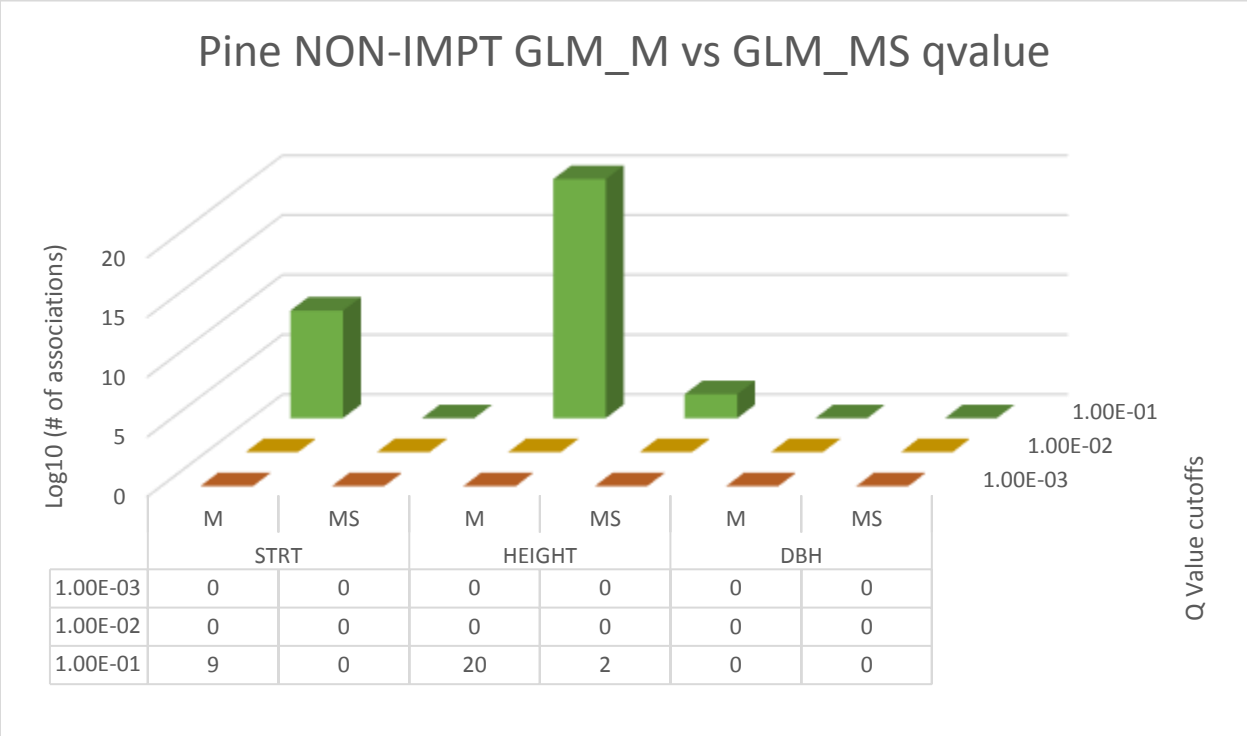


Figure S6. Distribution of SNPs effects on traits for non-imputed data (markers+ structure)



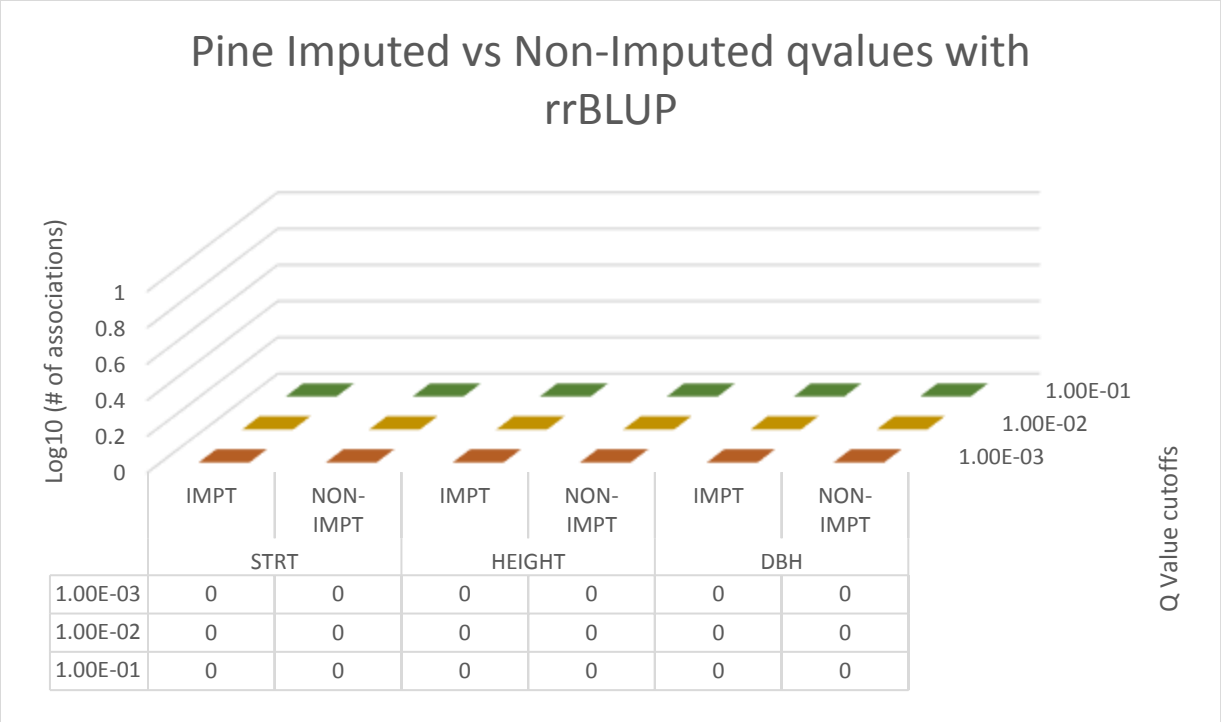
	STRT		HEIGHT		DBH	
	M	MS	M	MS	M	MS
1.00E-03	0	0	0	0	0	0
1.00E-02	1	1	0	0	0	0
1.00E-01	30	4	261	0	0	0

Figure S7. Adjusted the number of associations at different cut-offs in TASSEL (M and M+S) for imputed data



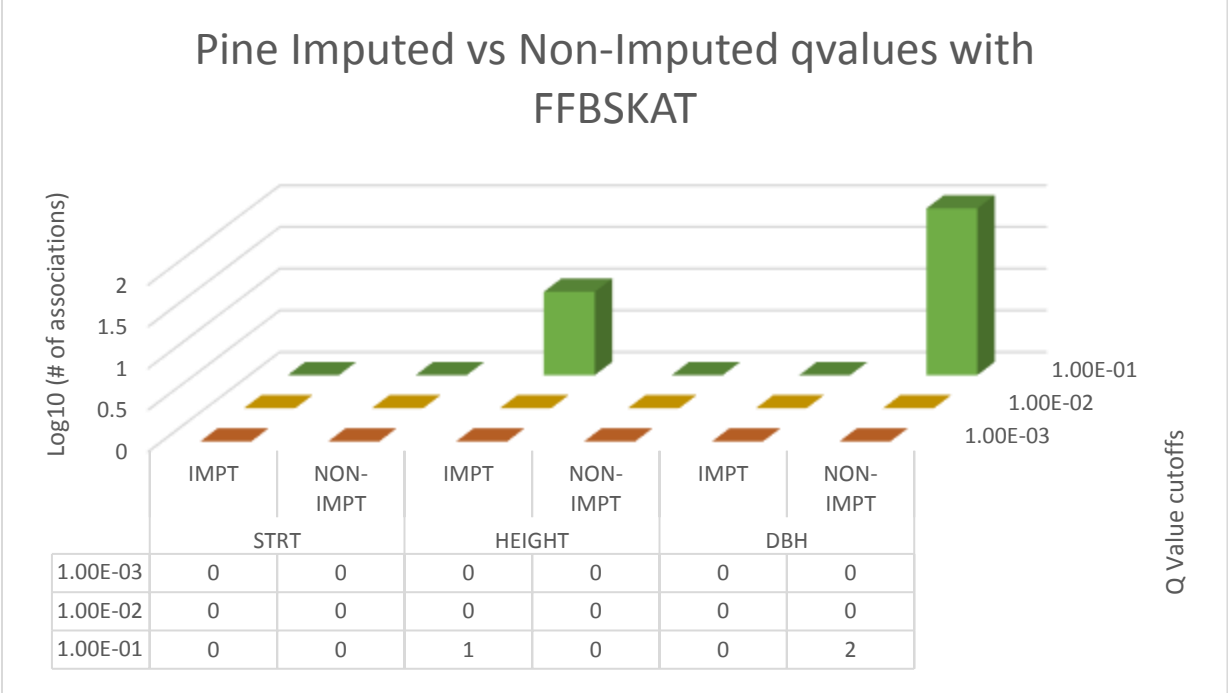
	STRT		HEIGHT		DBH	
	M	MS	M	MS	M	MS
1.00E-03	0	0	0	0	0	0
1.00E-02	0	0	0	0	0	0
1.00E-01	9	0	20	2	0	0

Figure S8. Adjusted the number of associations at different cut-offs in TASSEL (M and M+S) for non-imputed data



	STRT		HEIGHT		DBH	
	IMPT	NON-IMPT	IMPT	NON-IMPT	IMPT	NON-IMPT
1.00E-03	0	0	0	0	0	0
1.00E-02	0	0	0	0	0	0
1.00E-01	0	0	0	0	0	0

Figure S9. Adjusted the number of associations at different cut-offs in rrBLUP for both imputed and non-imputed data



	STRT		HEIGHT		DBH	
	IMPT	NON-IMPT	IMPT	NON-IMPT	IMPT	NON-IMPT
1.00E-03	0	0	0	0	0	0
1.00E-02	0	0	0	0	0	0
1.00E-01	0	0	1	0	0	2

Figure S10. Adjusted the number of associations at different cut-offs in FFBSKAT for both imputed and non-imputed data

Table S1. Idenditified SNP markers *(q<0.1)* names with p and q values in both data with all models (Highlighted rows show common associations)

Model_DATA_Trait	***SNP markers***	***p value***	***q value***
FFBSKAT_NON_DBH	tscaffold7173__92974	1.67E-06	0.05331171
FFBSKAT_NON_DBH	scaffold568043__65258	1.89E-06	0.05331171
FFBSKAT_IMP_HEIGT	tscaffold653__47572	1.54E-06	0.08852707
TASSEL_M+S_NON_IMP_height	scaffold553817.2__142879	3.74E-07	0.01110121
TASSEL_M+S_NON_IMP_height	scaffold553817.2__142880	3.01E-07	0.01110121
TASSEL_M_NON_IMP_strt	scaffold562589__32067	1.72E-06	0.065723357
TASSEL_M_NON_IMP_strt	scaffold291992__42889	2.62E-06	0.065723357
TASSEL_M_NON_IMP_strt	scaffold370988__66069	8.53E-06	0.097134869
TASSEL_M_NON_IMP_strt	tscaffold4642__476614	1.23E-05	0.097134869
TASSEL_M_NON_IMP_strt	tscaffold4916__816347	1.43E-05	0.097134869
TASSEL_M_NON_IMP_strt	C32572950__430936	1.21E-05	0.097134869
TASSEL_M_NON_IMP_strt	scaffold670722__11887	1.55E-05	0.097134869
TASSEL_M_NON_IMP_strt	scaffold670722__11935	1.55E-05	0.097134869
TASSEL_M_NON_IMP_strt	tscaffold6180__359419	1.79E-05	0.099805753
TASSEL_M_NON_IMP_Height	C32557778__122208	1.48E-05	0.062306914
TASSEL_M_NON_IMP_Height	C32557778__122219	9.53E-06	0.062306914
TASSEL_M_NON_IMP_Height	scaffold19427.2__75817	1.54E-05	0.062306914
TASSEL_M_NON_IMP_Height	scaffold743948__76098	9.61E-06	0.062306914
TASSEL_M_NON_IMP_Height	scaffold746166__122697	4.85E-06	0.062306914
TASSEL_M_NON_IMP_Height	scaffold746166__122709	2.30E-06	0.062306914
TASSEL_M_NON_IMP_Height	scaffold765982__34665	1.60E-05	0.062306914
TASSEL_M_NON_IMP_Height	scaffold81469.2__63412	1.60E-05	0.062306914
TASSEL_M_NON_IMP_Height	scaffold862033.1__28873	1.01E-05	0.062306914
TASSEL_M_NON_IMP_Height	tscaffold2678__968650	3.70E-06	0.062306914
TASSEL_M_NON_IMP_Height	tscaffold4177__107912	1.15E-05	0.062306914
TASSEL_M_NON_IMP_Height	tscaffold6646__280924	6.81E-06	0.062306914
TASSEL_M_NON_IMP_Height	tscaffold7592__221537	9.73E-06	0.062306914
TASSEL_M_NON_IMP_Height	scaffold315690.1__5044	2.35E-05	0.079533838
TASSEL_M_NON_IMP_Height	scaffold903385__24349	2.22E-05	0.079533838
TASSEL_M_NON_IMP_Height	tscaffold4175__301894	2.62E-05	0.082966042
TASSEL_M_NON_IMP_Height	scaffold20961.2__9292	2.80E-05	0.083435737
TASSEL_M_NON_IMP_Height	scaffold862033.1__28834	3.48E-05	0.092880021
TASSEL_M_NON_IMP_Height	tscaffold105__300758	3.36E-05	0.092880021
TASSEL_M_NON_IMP_Height	scaffold835161__33537	3.74E-05	0.094846941
TASSEL_M_imp_Height	scaffold321944__21191	3.20E-06	0.011103516
TASSEL_M_imp_Height	scaffold379417.2__10982	2.12E-06	0.011103516
TASSEL_M_imp_Height	scaffold647395__31641	1.18E-06	0.011103516
TASSEL_M_imp_Height	scaffold746166__122709	2.52E-06	0.011103516

TASSEL_M_imp_Height	scaffold81469.2__63412	3.02E-06	0.011103516
TASSEL_M_imp_Height	scaffold878148.1__59648	2.74E-06	0.011103516
TASSEL_M_imp_Height	scaffold901781__48779	3.15E-06	0.011103516
TASSEL_M_imp_Height	scaffold901781__48831	3.35E-06	0.011103516
TASSEL_M_imp_Height	tscaffold2678__968650	3.05E-06	0.011103516
TASSEL_M_imp_Height	tscaffold4182__30157	5.75E-07	0.011103516
TASSEL_M_imp_Height	scaffold743948__76098	5.98E-06	0.017990987
TASSEL_M_imp_Height	tscaffold8196__332340	6.56E-06	0.018097297
TASSEL_M_imp_Height	C32557778__122208	9.45E-06	0.022337368
TASSEL_M_imp_Height	scaffold223269__258115	9.37E-06	0.022337368
TASSEL_M_imp_Height	scaffold901781__49124	1.04E-05	0.022870663
TASSEL_M_imp_Height	scaffold117157__8550	1.22E-05	0.023663738
TASSEL_M_imp_Height	scaffold746166__122697	1.17E-05	0.023663738
TASSEL_M_imp_Height	tscaffold3026__45799	1.29E-05	0.023702684
TASSEL_M_imp_Height	tscaffold348__274748	1.48E-05	0.024474015
TASSEL_M_imp_Height	tscaffold8669__510933	1.48E-05	0.024474015
TASSEL_M_imp_Height	scaffold228381.2__6720	1.66E-05	0.025627248
TASSEL_M_imp_Height	tscaffold1962__1055993	1.70E-05	0.025627248
TASSEL_M_imp_Height	tscaffold6646__280924	1.83E-05	0.02628482
TASSEL_M_imp_Height	scaffold827872__224501	1.94E-05	0.026800696
TASSEL_M_imp_Height	scaffold182300.3__22709	2.18E-05	0.027868923
TASSEL_M_imp_Height	scaffold291660.1__25582	2.27E-05	0.027868923
TASSEL_M_imp_Height	scaffold729135__104305	2.23E-05	0.027868923
TASSEL_M_imp_Height	scaffold642151__62387	2.45E-05	0.029001738
TASSEL_M_imp_Height	scaffold102788__200176	2.71E-05	0.029850154
TASSEL_M_imp_Height	tscaffold1808__89538	2.66E-05	0.029850154
TASSEL_M_imp_Height	C32557778__122219	2.84E-05	0.030347039
TASSEL_M_imp_Height	scaffold550806__16077	3.24E-05	0.032757214
TASSEL_M_imp_Height	tscaffold5828__219537	3.27E-05	0.032757214
TASSEL_M_imp_Height	scaffold765982__34665	3.77E-05	0.03564861
TASSEL_M_imp_Height	tscaffold7918__10569	3.76E-05	0.03564861
TASSEL_M_imp_Height	C32567992__5490	4.79E-05	0.038944523
TASSEL_M_imp_Height	scaffold111891__119622	4.96E-05	0.038944523
TASSEL_M_imp_Height	scaffold315690.1__5040	4.58E-05	0.038944523
TASSEL_M_imp_Height	scaffold362231.2__75268	4.96E-05	0.038944523
TASSEL_M_imp_Height	scaffold844251.2__59153	5.06E-05	0.038944523
TASSEL_M_imp_Height	tscaffold1954__163064	4.74E-05	0.038944523
TASSEL_M_imp_Height	tscaffold4177__107912	4.78E-05	0.038944523
TASSEL_M_imp_Height	tscaffold5828__219656	4.83E-05	0.038944523
TASSEL_M_imp_Height	tscaffold4022__724911	5.33E-05	0.040096088
TASSEL_M_imp_Height	tscaffold5218__36581	5.93E-05	0.043649133

TASSEL_M_imp_Height	tscaffold4352__3481274	6.20E-05	0.043653312
TASSEL_M_imp_Height	tscaffold633__92603	6.20E-05	0.043653312
TASSEL_M_imp_Height	tscaffold7303__251994	6.65E-05	0.045834293
TASSEL_M_imp_Height	tscaffold1954__163066	6.79E-05	0.045871085
TASSEL_M_imp_Height	tscaffold8673__21932	7.11E-05	0.047053801
TASSEL_M_imp_Height	tscaffold4179__119881	7.40E-05	0.047093424
TASSEL_M_imp_Height	tscaffold7288__94479	7.38E-05	0.047093424
TASSEL_M_imp_Height	scaffold831169__192813	7.92E-05	0.049468454
TASSEL_M_imp_Height	scaffold474610__165169	8.21E-05	0.049699196
TASSEL_M_imp_Height	scaffold746001.1__69680	8.26E-05	0.049699196
TASSEL_M_imp_Height	scaffold536159.3__75146	8.46E-05	0.04999519
TASSEL_M_imp_Height	scaffold903385__24349	9.53E-05	0.054394373
TASSEL_M_imp_Height	tscaffold1954__163091	9.43E-05	0.054394373
TASSEL_M_imp_Height	tscaffold7592__221537	9.72E-05	0.054550288
TASSEL_M_imp_Height	scaffold827872__240408	9.99E-05	0.055120879
TASSEL_M_imp_Height	scaffold699751__97974	0.00010515	0.055252666
TASSEL_M_imp_Height	tscaffold5091__391360	0.00010438	0.055252666
TASSEL_M_imp_Height	tscaffold5218__36578	0.0001035	0.055252666
TASSEL_M_imp_Height	tscaffold1744__666313	0.00010811	0.055920417
TASSEL_M_imp_Height	tscaffold7197__10025	0.00010995	0.05594628
TASSEL_M_imp_Height	tscaffold8906__121745	0.00011154	0.05594628
TASSEL_M_imp_Height	scaffold858168__149188	0.00011338	0.056020394
TASSEL_M_imp_Height	scaffold13784__109460	0.00012313	0.057928348
TASSEL_M_imp_Height	scaffold13784__109465	0.00013718	0.057928348
TASSEL_M_imp_Height	scaffold181259__53885	0.00012542	0.057928348
TASSEL_M_imp_Height	scaffold474610__165122	0.00013265	0.057928348
TASSEL_M_imp_Height	scaffold901781__48984	0.00013676	0.057928348
TASSEL_M_imp_Height	scaffold903836__29632	0.00013696	0.057928348
TASSEL_M_imp_Height	tscaffold1808__89510	0.00013742	0.057928348
TASSEL_M_imp_Height	tscaffold2587__292574	0.00012761	0.057928348
TASSEL_M_imp_Height	tscaffold4175__301894	0.00012237	0.057928348
TASSEL_M_imp_Height	tscaffold4352__3481258	0.00012298	0.057928348
TASSEL_M_imp_Height	tscaffold5218__36552	0.00013496	0.057928348
TASSEL_M_imp_Height	tscaffold6431__80673	0.00013824	0.057928348
TASSEL_M_imp_Height	C32505780__1867	0.00013999	0.0579284
TASSEL_M_imp_Height	scaffold362231.2__75282	0.00014639	0.058439074
TASSEL_M_imp_Height	scaffold710503__257071	0.00014526	0.058439074
TASSEL_M_imp_Height	tscaffold3410__49408	0.00014652	0.058439074
TASSEL_M_imp_Height	scaffold13784__109459	0.00015496	0.059488821
TASSEL_M_imp_Height	scaffold844251.2__59194	0.00015316	0.059488821
TASSEL_M_imp_Height	tscaffold5532__108557	0.00015384	0.059488821

TASSEL_M_imp_Height	tscaffold8663__620205	0.00015634	0.059488821
TASSEL_M_imp_Height	scaffold95336.2__75455	0.00015923	0.059899988
TASSEL_M_imp_Height	scaffold844090.1__184425	0.00016739	0.060893737
TASSEL_M_imp_Height	scaffold899056__39652	0.00016658	0.060893737
TASSEL_M_imp_Height	scaffold901781__48997	0.0001639	0.060893737
TASSEL_M_imp_Height	scaffold431635__76771	0.00017291	0.060894317
TASSEL_M_imp_Height	tscaffold4022__724915	0.00017183	0.060894317
TASSEL_M_imp_Height	tscaffold761__1163470	0.00017245	0.060894317
TASSEL_M_imp_Height	scaffold897334.2__4414	0.00018459	0.064323412
TASSEL_M_imp_Height	tscaffold4104__356364	0.00018707	0.064508571
TASSEL_M_imp_Height	scaffold127397.1__91935	0.00018936	0.064625069
TASSEL_M_imp_Height	tscaffold2554__1904416	0.0001943	0.064971384
TASSEL_M_imp_Height	tscaffold4175__301859	0.00019328	0.064971384
TASSEL_M_imp_Height	C32565652__131668	0.00023269	0.066301986
TASSEL_M_imp_Height	scaffold327550__51623	0.00023433	0.066301986
TASSEL_M_imp_Height	scaffold358484.1__8891	0.0002287	0.066301986
TASSEL_M_imp_Height	scaffold412521.2__161935	0.00020904	0.066301986
TASSEL_M_imp_Height	scaffold457074.1__42888	0.0002196	0.066301986
TASSEL_M_imp_Height	scaffold567144__130840	0.00023015	0.066301986
TASSEL_M_imp_Height	scaffold579331__141196	0.00021664	0.066301986
TASSEL_M_imp_Height	scaffold893600__70198	0.00022367	0.066301986
TASSEL_M_imp_Height	scaffold897334.2__4413	0.00021347	0.066301986
TASSEL_M_imp_Height	scaffold904737__249829	0.00022881	0.066301986
TASSEL_M_imp_Height	tscaffold1679__917286	0.00022556	0.066301986
TASSEL_M_imp_Height	tscaffold258__441015	0.00022532	0.066301986
TASSEL_M_imp_Height	tscaffold2942__1905138	0.00021125	0.066301986
TASSEL_M_imp_Height	tscaffold3026__45888	0.00022359	0.066301986
TASSEL_M_imp_Height	tscaffold5419__37847	0.00021821	0.066301986
TASSEL_M_imp_Height	tscaffold5537__8929	0.00023315	0.066301986
TASSEL_M_imp_Height	tscaffold686__33947	0.00021309	0.066301986
TASSEL_M_imp_Height	tscaffold7698__193004	0.00022321	0.066301986
TASSEL_M_imp_Height	C32564004__101204	0.00025411	0.067283192
TASSEL_M_imp_Height	scaffold328900.1__51913	0.00025609	0.067283192
TASSEL_M_imp_Height	scaffold486613__69125	0.00025483	0.067283192
TASSEL_M_imp_Height	scaffold698029.2__9125	0.00024586	0.067283192
TASSEL_M_imp_Height	scaffold895030__2673	0.00025067	0.067283192
TASSEL_M_imp_Height	scaffold901044__138781	0.00024024	0.067283192
TASSEL_M_imp_Height	tscaffold3974__46559	0.000242	0.067283192
TASSEL_M_imp_Height	tscaffold4179__119899	0.00025495	0.067283192
TASSEL_M_imp_Height	tscaffold7622__440039	0.00025391	0.067283192
TASSEL_M_imp_Height	scaffold594585__51118	0.0002624	0.068398192

TASSEL_M_imp_Height	C32567992__5471	0.00027614	0.069781859
TASSEL_M_imp_Height	scaffold328900.1__51947	0.00027414	0.069781859
TASSEL_M_imp_Height	scaffold447661__147477	0.00027543	0.069781859
TASSEL_M_imp_Height	tscaffold5999__454313	0.00027523	0.069781859
TASSEL_M_imp_Height	scaffold848651__23719	0.00028394	0.071129793
TASSEL_M_imp_Height	tscaffold1122__1567178	0.00028782	0.071129793
TASSEL_M_imp_Height	tscaffold5826__281596	0.00028792	0.071129793
TASSEL_M_imp_Height	tscaffold193__100350	0.00029512	0.071899635
TASSEL_M_imp_Height	tscaffold8727__65628	0.00029538	0.071899635
TASSEL_M_imp_Height	scaffold517514__27824	0.00029967	0.071980279
TASSEL_M_imp_Height	scaffold826177__4966	0.00030006	0.071980279
TASSEL_M_imp_Height	C32565652__131723	0.00030958	0.072683912
TASSEL_M_imp_Height	scaffold699751__68991	0.00030575	0.072683912
TASSEL_M_imp_Height	tscaffold6892__179143	0.00030776	0.072683912
TASSEL_M_imp_Height	scaffold874532__80243	0.00032138	0.074922975
TASSEL_M_imp_Height	scaffold767838__31149	0.00032396	0.074996304
TASSEL_M_imp_Height	scaffold433454.1__60238	0.00035109	0.077191841
TASSEL_M_imp_Height	scaffold581869__165667	0.00034967	0.077191841
TASSEL_M_imp_Height	scaffold885421.1__57243	0.00033806	0.077191841
TASSEL_M_imp_Height	tscaffold1353__139586	0.00035443	0.077191841
TASSEL_M_imp_Height	tscaffold2383__27769	0.00033843	0.077191841
TASSEL_M_imp_Height	tscaffold3026__45883	0.0003544	0.077191841
TASSEL_M_imp_Height	tscaffold530__165150	0.00034699	0.077191841
TASSEL_M_imp_Height	tscaffold7780__171134	0.00034995	0.077191841
TASSEL_M_imp_Height	tscaffold8746__1418378	0.00035358	0.077191841
TASSEL_M_imp_Height	tscaffold1870__11122	0.00036408	0.07877527
TASSEL_M_imp_Height	scaffold797596__77532	0.00037472	0.079011759
TASSEL_M_imp_Height	scaffold901044__138755	0.00037223	0.079011759
TASSEL_M_imp_Height	scaffold901044__138808	0.00037223	0.079011759
TASSEL_M_imp_Height	tscaffold1382__26864	0.00037375	0.079011759
TASSEL_M_imp_Height	scaffold143827__19058	0.00039199	0.079929849
TASSEL_M_imp_Height	scaffold409447__9788	0.00038182	0.079929849
TASSEL_M_imp_Height	scaffold419899.2__65851	0.0003867	0.079929849
TASSEL_M_imp_Height	scaffold51451__153914	0.00039203	0.079929849
TASSEL_M_imp_Height	scaffold665687__158642	0.00039839	0.079929849
TASSEL_M_imp_Height	scaffold793454__66876	0.00039795	0.079929849
TASSEL_M_imp_Height	scaffold891136.2__32121	0.00038765	0.079929849
TASSEL_M_imp_Height	tscaffold2231__89254	0.00039509	0.079929849
TASSEL_M_imp_Height	scaffold419899.2__65790	0.00040217	0.080202165
TASSEL_M_imp_Height	tscaffold1308__87140	0.00041079	0.08143065
TASSEL_M_imp_Height	scaffold101238__11541	0.0004228	0.082332361

TASSEL_M_imp_Height	scaffold791166.2__37947	0.00042048	0.082332361
TASSEL_M_imp_Height	tscaffold619__182806	0.00042068	0.082332361
TASSEL_M_imp_Height	tscaffold1808__89634	0.00043754	0.084287777
TASSEL_M_imp_Height	tscaffold6646__281080	0.00044048	0.084287777
TASSEL_M_imp_Height	tscaffold7177__116209	0.00043913	0.084287777
TASSEL_M_imp_Height	tscaffold4352__3666760	0.00044414	0.084499697
TASSEL_M_imp_Height	tscaffold856__134909	0.00044703	0.084563535
TASSEL_M_imp_Height	scaffold232176__49407	0.00045217	0.085049857
TASSEL_M_imp_Height	scaffold746001.1__69648	0.00046901	0.087718934
TASSEL_M_imp_Height	scaffold231872__96470	0.0004771	0.088235001
TASSEL_M_imp_Height	scaffold231872__96471	0.0004771	0.088235001
TASSEL_M_imp_Height	tscaffold7490__473981	0.00048155	0.088563219
TASSEL_M_imp_Height	scaffold111891__119623	0.00049807	0.089831418
TASSEL_M_imp_Height	scaffold323141__88497	0.00049524	0.089831418
TASSEL_M_imp_Height	scaffold699751__98040	0.00049527	0.089831418
TASSEL_M_imp_Height	tscaffold5330__61637	0.0004993	0.089831418
TASSEL_M_imp_Height	scaffold463697__28613	0.00050515	0.090392654
TASSEL_M_imp_Height	scaffold693733__116685	0.00051285	0.090789006
TASSEL_M_imp_Height	scaffold777287__165141	0.0005116	0.090789006
TASSEL_M_imp_Height	tscaffold761__84376	0.00051906	0.091399585
TASSEL_M_imp_Height	scaffold215683__106058	0.00053118	0.091585136
TASSEL_M_imp_Height	scaffold262518__234461	0.00052922	0.091585136
TASSEL_M_imp_Height	scaffold760607__80285	0.00052949	0.091585136
TASSEL_M_imp_Height	tscaffold158__446920	0.0005243	0.091585136
TASSEL_M_imp_Height	tscaffold3822__589164	0.00053716	0.092136321
TASSEL_M_imp_Height	C32349178__33771	0.0005649	0.092422987
TASSEL_M_imp_Height	scaffold387130__226849	0.00054569	0.092422987
TASSEL_M_imp_Height	scaffold435552__41830	0.00056608	0.092422987
TASSEL_M_imp_Height	scaffold441117.1__57795	0.00056456	0.092422987
TASSEL_M_imp_Height	scaffold508668.1__19835	0.00056675	0.092422987
TASSEL_M_imp_Height	scaffold887087__44858	0.00056193	0.092422987
TASSEL_M_imp_Height	scaffold890666__83146	0.00056644	0.092422987
TASSEL_M_imp_Height	tscaffold1967__79716	0.00054628	0.092422987
TASSEL_M_imp_Height	tscaffold4749__245018	0.0005534	0.092422987
TASSEL_M_imp_Height	tscaffold6718__66280	0.00055265	0.092422987
TASSEL_M_imp_Height	scaffold779918.1__44087	0.00057873	0.092552927
TASSEL_M_imp_Height	scaffold890666__83030	0.00057528	0.092552927
TASSEL_M_imp_Height	tscaffold3065__287659	0.00057724	0.092552927
TASSEL_M_imp_Height	tscaffold8908__84203	0.00057173	0.092552927
TASSEL_M_imp_Height	C32565652__131711	0.00058155	0.092556779
TASSEL_M_imp_Height	scaffold475926__435904	0.00058987	0.092986847

TASSEL_M_imp_Height	scaffold848651__23877	0.00058972	0.092986847
TASSEL_M_imp_Height	tscaffold2354__404126	0.00060136	0.094348846
TASSEL_M_imp_Height	scaffold102861.2__113636	0.00062695	0.09465625
TASSEL_M_imp_Height	scaffold231795__21982	0.00061628	0.09465625
TASSEL_M_imp_Height	scaffold563497__171064	0.00063513	0.09465625
TASSEL_M_imp_Height	scaffold637208__12383	0.00062535	0.09465625
TASSEL_M_imp_Height	scaffold824875__6433	0.00063616	0.09465625
TASSEL_M_imp_Height	scaffold871133__131055	0.00063278	0.09465625
TASSEL_M_imp_Height	scaffold891136.2__32102	0.00060739	0.09465625
TASSEL_M_imp_Height	scaffold904019__4459	0.00061605	0.09465625
TASSEL_M_imp_Height	tscaffold1967__120571	0.00064122	0.09465625
TASSEL_M_imp_Height	tscaffold3822__589122	0.00062267	0.09465625
TASSEL_M_imp_Height	tscaffold3822__589128	0.00062525	0.09465625
TASSEL_M_imp_Height	tscaffold4352__3666768	0.00064335	0.09465625
TASSEL_M_imp_Height	tscaffold6718__66236	0.00063703	0.09465625
TASSEL_M_imp_Height	tscaffold8727__65621	0.00063967	0.09465625
TASSEL_M_imp_Height	scaffold771685.2__14857	0.00065066	0.095308181
TASSEL_M_imp_Height	tscaffold8749__445391	0.00066329	0.096730204
TASSEL_M_imp_Height	C32492822__60104	0.00067866	0.097257876
TASSEL_M_imp_Height	scaffold116089__125172	0.00067593	0.097257876
TASSEL_M_imp_Height	scaffold365144__40411	0.00067432	0.097257876
TASSEL_M_imp_Height	tscaffold686__33982	0.00067428	0.097257876
TASSEL_M_imp_Height	scaffold177904.1__111914	0.00069729	0.097508249
TASSEL_M_imp_Height	scaffold296557__15363	0.00069151	0.097508249
TASSEL_M_imp_Height	scaffold517514__27784	0.00069808	0.097508249
TASSEL_M_imp_Height	scaffold517514__27797	0.00069521	0.097508249
TASSEL_M_imp_Height	scaffold844333__106310	0.00068853	0.097508249
TASSEL_M_imp_Height	scaffold877603__114938	0.00068729	0.097508249
TASSEL_M_imp_Height	scaffold231795__21850	0.00070558	0.09759446
TASSEL_M_imp_Height	scaffold309571__127490	0.00072332	0.09759446
TASSEL_M_imp_Height	scaffold8934__114773	0.00072257	0.09759446
TASSEL_M_imp_Height	scaffold902240__16600	0.0007057	0.09759446
TASSEL_M_imp_Height	tscaffold1782__108349	0.00072095	0.09759446
TASSEL_M_imp_Height	tscaffold2568__170017	0.00072495	0.09759446
TASSEL_M_imp_Height	tscaffold2767__1281072	0.00072523	0.09759446
TASSEL_M_imp_Height	tscaffold4386__1158245	0.00072393	0.09759446
TASSEL_M_imp_Height	tscaffold9052__44173	0.00070898	0.09759446
TASSEL_M_imp_Height	C32304344__9005	0.00073118	0.097689744
TASSEL_M_imp_Height	scaffold221300__63706	0.00073184	0.097689744
TASSEL_M_imp_Height	scaffold110155__34623	0.00073556	0.097791986
TASSEL_M_imp_Height	tscaffold7945__466885	0.00074413	0.098535634

TASSEL_M_imp_Height	C32492822__60094	0.00075147	0.098796653
TASSEL_M_imp_Height	tscaffold761__84347	0.00075207	0.098796653
TASSEL_M_imp_Height	C32565652__131619	0.00076552	0.098992225
TASSEL_M_imp_Height	scaffold291742.3__13372	0.00076505	0.098992225
TASSEL_M_imp_Height	scaffold446753__69589	0.00076458	0.098992225
TASSEL_M_imp_Height	tscaffold6603__41044	0.00075763	0.098992225
TASSEL_M_imp_Height	tscaffold997__240823	0.00077262	0.099521596
TASSEL_M_imp_Height	scaffold109629__50369	0.00078329	0.09988115
TASSEL_M_imp_Height	tscaffold2847__58253	0.00078748	0.09988115
TASSEL_M_imp_Height	tscaffold326__422265	0.00078017	0.09988115
TASSEL_M_imp_Height	tscaffold616__1864735	0.00078653	0.09988115
TASSEL_M_imp_STRT	tscaffold4866__23071	1.09E-07	0.003776553
TASSEL_M_imp_STRT	scaffold197254__38524	2.41E-06	0.039647412
TASSEL_M_imp_STRT	scaffold13426__257381	5.23E-06	0.039647412
TASSEL_M_imp_STRT	tscaffold1121__166000	5.73E-06	0.039647412
TASSEL_M_imp_STRT	tscaffold1877__65610	3.49E-06	0.039647412
TASSEL_M_imp_STRT	scaffold706759__11139	8.56E-06	0.049377938
TASSEL_M_imp_STRT	tscaffold2423__399094	1.63E-05	0.066752541
TASSEL_M_imp_STRT	tscaffold4916__816347	1.93E-05	0.066752541
TASSEL_M_imp_STRT	tscaffold8548__181062	2.12E-05	0.066752541
TASSEL_M_imp_STRT	tscaffold5482__90040	1.77E-05	0.066752541
TASSEL_M_imp_STRT	tscaffold1706__3306234	1.99E-05	0.066752541
TASSEL_M_imp_STRT	tscaffold5694__3741825	2.75E-05	0.079143766
TASSEL_M_imp_STRT	scaffold32313__19704	3.41E-05	0.082418601
TASSEL_M_imp_STRT	tscaffold5591__150810	4.69E-05	0.082418601
TASSEL_M_imp_STRT	tscaffold3624__48069	4.75E-05	0.082418601
TASSEL_M_imp_STRT	tscaffold4741__20472	4.53E-05	0.082418601
TASSEL_M_imp_STRT	scaffold291992__42889	3.84E-05	0.082418601
TASSEL_M_imp_STRT	tscaffold6790__514079	3.70E-05	0.082418601
TASSEL_M_imp_STRT	tscaffold6180__359419	4.77E-05	0.082418601
TASSEL_M_imp_STRT	tscaffold7689__109641	4.67E-05	0.082418601
TASSEL_M_imp_STRT	scaffold653336__58299	5.14E-05	0.084643281
TASSEL_M_imp_STRT	scaffold871133__131055	6.25E-05	0.087732828
TASSEL_M_imp_STRT	tscaffold8851__103167	6.34E-05	0.087732828
TASSEL_M_imp_STRT	tscaffold5581__391416	5.87E-05	0.087732828
TASSEL_M_imp_STRT	tscaffold4904__371931	5.92E-05	0.087732828
TASSEL_M_imp_STRT	tscaffold5883__237632	6.94E-05	0.092308301
TASSEL_M_imp_STRT	scaffold48162__335168	7.46E-05	0.093206777
TASSEL_M_imp_STRT	tscaffold6466__2776624	7.54E-05	0.093206777
TASSEL_M_imp_STRT	tscaffold2554__1904416	8.57E-05	0.098920771
TASSEL_M_imp_STRT	tscaffold380__474466	8.58E-05	0.098920771

TASSEL_MS_imp_STRT	tscaffold4866__23071	1.70E-07	0.006756057
TASSEL_MS_imp_STRT	scaffold197254__38524	3.93E-06	0.062722299
TASSEL_MS_imp_STRT	tscaffold5482__90040	4.73E-06	0.062722299
TASSEL_MS_imp_STRT	scaffold13426__257381	8.81E-06	0.087671247