

Genomic selection in forest tree breeding: the concept and an outlook to the future

Fikret Isik

New Forests

International Journal on the Biology,
Biotechnology, and Management of
Afforestation and Reforestation

ISSN 0169-4286

New Forests
DOI 10.1007/s11056-014-9422-z



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Genomic selection in forest tree breeding: the concept and an outlook to the future

Fikret Isik

Received: 25 September 2013 / Accepted: 17 March 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Using large numbers of DNA markers to predict genetic merit [genomic selection (GS)] is a new frontier in plant and animal breeding programs. GS is now routinely used to select superior bulls in dairy cattle breeding. In forest trees, a few empirical proof of-concept studies suggest that GS could be successful. However, application of GS in forest tree breeding is still in its infancy. The major hurdle is lack of high throughput genotyping platforms for trees, and the high genotyping costs, though, the cost of genotyping will likely decrease in the future. There has been a growing interest in GS among tree breeders, forest geneticists, and tree improvement managers. A broad overview of pedigree reconstruction and GS is presented. Underlying reasons for failures of marker-assisted selection were summarized and compared with GS. Challenges of GS in forest tree breeding and the outlook for the future are discussed, and a GS plan for a cloned loblolly pine breeding population is presented. This review is intended for tree breeders, forest managers, scientist and students who are not necessarily familiar with genomic or quantitative genetics jargon.

Keywords Genomic selection · Forest tree breeding · Kinship analysis · Pedigree reconstruction · Marker assisted selection

Genomic selection (GS)

Introduction

Populating the genome with dense marker coverage and exploiting the associations between markers and phenotype for selection is called genomic selection (GS) (Meuwissen

F. Isik (✉)
Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC,
USA
e-mail: fisik@ncsu.edu

et al. 2001). GS has revolutionized cattle breeding in the last few years (Goddard et al. 2011). Crop and tree breeders have enthusiastically embraced the idea since its successful application in several dairy cattle breeding programs around the world. There has been an increasing interest in GS among forest geneticists, tree breeders and forest managers (Grattapaglia and Resende 2011; Resende et al. 2011, 2012a, b; Zapata-Valenzuela et al. 2012, 2013).

Pedigree reconstruction and kinship analyses have also become popular. In recent years, we have seen an increasing numbers of publications on fingerprinting, pedigree reconstruction and error detection in tree breeding (El-Kassaby and Lstibůrek 2009; El-Kassaby et al. 2011; Hansen and McKinney 2010; Hansen and Nielsen 2010).

The objectives of this paper were (1) to introduce the concept of GS and pedigree analysis to forestry and tree breeding communities, (2) discuss the challenges and opportunities of implementation of GS, and (3) present a GS implementation plan for loblolly pine (*Pinus taeda* L.) breeding, one of the most intensively studied tree species.

This review was written for readers who are not intimately familiar with quantitative genetics, molecular genomics, and statistical applications. Where necessary, the technical terms are defined. Acronyms are avoided to make the paper easier to read. The discussions were limited to GS and pedigree reconstruction in forest trees to keep the article short and help with the flow.

In contrast to *marker assisted selection* (MAS) which utilizes markers to track small numbers of loci with large effects, GS relies on the joint analysis of large numbers of DNA markers to predict *breeding values* (BV) of individuals (Meuwissen et al. 2001). In GS, there is no defined small subset of significant markers used for prediction. Instead, all the loci causing phenotypic variation among individuals are traced with dense marker coverage (Hayes and Goddard 2010). The sum of the marker effects is used to predict BV of individuals (Goddard et al. 2011). In GS the prediction of BV is based on markers only without known phenotype. In MAS, depending on the link between marker and *quantitative trait loci* (commonly called QTL), phenotype might be needed to make selection decisions.

Estimated Breeding Value (EBV): Genetic merit of an individual. An individual passes on a half of its breeding value (genes) to progeny. We never know the actual breeding value of an individual but we estimate it by measuring the relatives, particularly the progeny of the individual. EBVs are expressed in the same measurement unit as the recorded trait (e.g. centimeter for height) and they relate to the population mean.

Marker Assisted Selection (MAS): The use of DNA markers to study the inheritance of quantitative trait loci (QTL) with large effects, and use the marker-linked QTL to make selection decisions in breeding.

Quantitative trait loci (QTL): A locus (gene) that has a large or small effect on a quantitative trait, such as height.

GS in animal breeding

GS is the new frontier and a paradigm shift for plant and animal breeding programs. Many dairy cattle breeding programs routinely use *single nucleotide polymorphism* (SNP) markers to predict the genetic merit of bulls (Hayes and Goddard 2010). Since 2009,

artificial insemination companies have shifted towards of marketing *genomic bulls*. Young bulls selected based on SNP markers but without milk yield data from their daughters are referred to as 'genomic bulls' (Scheffers and Weigel 2012).

Single nucleotide polymorphism (SNP): Nucleotides A, C, T, G are building blocks of DNA. A difference in a single nucleotide (one of the letters) between individuals is called *Single Nucleotide Polymorphism* or SNP (pronounced as *snip*).

With GS, the potential to increase genetic gain was estimated to be twofold, and savings in logistical costs were estimated to be 92 % in cattle breeding (Schaeffer 2006). In Europe, the EuroGenomics consortium has been coordinating the genotyping. About 65,000 animals were genotyped using the Labogena *SNP chip* between 2008 and 2012, and the number was expected to increase rapidly in 2012 and beyond (Boichard et al. 2012). High-density *SNP genotyping arrays* are available for chicken (Kranis et al. 2013) and for other major animal breeding programs (Scheffers and Weigel 2012).

SNP chip: A small (~ 1 cm square) silicon glass with a large number of synthetic, single-stranded DNA nucleotides. DNA of individuals is hybridized on the chip to identify presence of specific DNA sequences that differ by a single nucleotide. Most SNP chips are designed to identify 50,000 SNPs (referred as 50K SNP chip). Several SNP chips together on a slide are called *SNP arrays*.

In order to understand how GS could be so efficient, we need to take a look at the response to selection R , using a modification of the traditional breeder's equation.

$$R = h\sigma_A i/T \quad (1)$$

where h is the square root of heritability (proportion of variation in phenotypic variance explained by the genetic factors) or the correlation between true and predicted BV, σ_A is the additive genetic standard deviation, i is the selection intensity, and T is the time per cycle spent to achieve the response. It is clear that the time is a critical component on the response. High response (genetic gain) from GS comes mainly from reducing the lengths of breeding cycle (time) by eliminating costly progeny testing. For example, in dairy cattle, with GS, genomic bulls are selected at 21 months instead of 54 months using traditional progeny testing (Scheffers and Weigel 2012).

High-throughput genotyping allows higher selection intensities (i) to be applied. In dairy cattle, about 1 bull is selected for every 20 bulls progeny tested (Scheffers and Weigel 2012) and 30–40 daughters are typically used to evaluate a bull for milk yield. With GS, very large number of newborn bulls can be screened. For example, in 2012, about 1,000 new-born bulls were genotyped per month, and about 33,000 bulls have been screened for selection since 2009 (Scheffers and Weigel 2012). This is a dramatic increase in selection intensity.

GS could also increase the reliability (h) of predictions of BV and thus may increase the response to selection. When averaged across 11 traits in dairy cattle, reliabilities of BV based on genomic data were 23 % greater (50 vs 27 %) than reliabilities based on parental averages (VanRaden et al. 2009). The increase in reliability of BV was the equivalent of having data from 11 additional daughters for each bull.

GS in crop breeding

In crop species, breeding cycles are usually much shorter. For example, in maize breeding it takes about 2 years to finish a cycle of breeding and testing in the field (Heffner et al.

2009). Major crop breeding programs use winter nurseries in the southern hemisphere to reduce breeding and field-testing time in half. Still, for many crop species and grasses, the value of GS is well recognized (Jannink et al. 2010; Heffner et al. 2009). Phenotyping large numbers of lines is costly and time consuming. In some grass and turf species, a breeding cycle can take as long as 5 years (Hayes et al. 2013). It is not surprising to see a growing interest in GS for crop breeding (Zhao et al. 2012; Würschum et al. 2013). Major crop breeding organizations and private companies have been investing in developing genotyping platforms. For rice and maize, for example, a 60 K SNP arrays was developed (Gupta et al. 2008). For wheat, a low density (9 K) SNP array was designed (Ganal et al. 2011). Successful application of GS in biparental populations of maize, and barley was demonstrated (Lorenzana and Bernardo 2009). We should expect a growing number of GS applications in crop species in the near future.

Enabling technologies

The genomic revolution in breeding is mainly attributed to efficiency in high-throughput DNA sequencing technologies. Ever-improving technologies have made DNA sequencing faster, more accurate, and far cheaper. The price to sequence a single nucleotide of DNA has fallen 100 million fold since 1990 (Eggen 2012). SNP markers have been the choice in genotyping efforts. They are abundant in plant and animal genomes and have the potential to produce high throughput data at a moderate cost. Compared to *microsatellite markers* they have low mutation rate. A very high majority of SNP markers are biallelic making them easy to analyze. Next generation sequencing technologies have enabled large panels of SNPs to be discovered (Perkel 2008). Medium density (50 K) SNP arrays are available for many livestock animals (Hayes and Goddard 2010). The discovery of high quality SNPs is still expensive and requires substantial investment and collaboration between stakeholder and scientists (Davey et al. 2011).

Microsatellites or SSR markers: Simple sequence repeats (SSR) of DNA. They are usually short, 2 to 8 nucleotides in length. SSR markers are popular in population genetics, forensic studies, and paternity analyses because they are highly polymorphic or multiallelic (variable).

Looking back: marker assisted selection (MAS)

The idea of using MAS in breeding programs goes back many decades. There has been a substantial number of studies on discovery and utilization of QTL in breeding (Groover et al. 1994; Sewell et al. 2000; Jermstad et al. 2001; Emebiri et al. 1997; Guimarães 2007). QTL mapping played a major role for dissecting the genetic basis of phenotypic variation in forest trees (Chagné et al. 2003; Lerceteau et al. 2000; Pot et al. 2006). We have better understating of the effect of individual QTL on phenotype (variation explained) and their position in the genome (Neale and Savolainen 2004). For example, QTL mapping revealed a simple inheritance of fusiform rust resistance genes in loblolly pine (Wilcox et al. 2001; Isik et al. 2012). The idea of discovering QTL and use in MAS was particularly appealing because forest trees have long (up to 30 years) breeding cycles, and MAS was seen as the tool to reduce it. Following Bernardo (2008), Google Scholar was used to tally the number of published papers having all the following words “QTL, QTLs, forest trees, MAS” that

appear anywhere in a manuscript. For forest trees, there were 560 published manuscripts from 1970 to 2012. The number of papers spiked between 1990 and 2010 (Fig. 1).

Wood specific gravity was considered an ideal trait to develop a MAS scheme in loblolly pine because of its high heritability (~ 0.4) and the long time required to measure the phenotype (Sewell et al. 2000). In *Pinus pinaster*, 1–3 QTL were mapped for various seedling traits, and the percent variance explained by each ranged from 6 to 17 % (Plomion et al. 1996). Kumar and Garrick (2001) reported that MAS provides an opportunity to increase the efficiency of within-family selection in forest tree breeding. O'Malley and McKeand (1994) suggested that MAS for low heritability traits, such as growth, could be valuable in tree breeding if families for QTL mapping be identified from breeding records. Using simulation, Wilcox et al. (2001) suggested that significant financial returns from MAS are possible in *Pinus radiata* even if selection is based on upon DNA markers linked to a few loci each of relatively small effect. Williams and Neale (1992) suggested that the potential of using MAS in interspecific tree breeding programs was higher than in pure species breeding because of large linkage disequilibrium (LD) and large phenotypic variance observed in F2 and backcross populations. LD is an ambiguous term. A broad definition is given below to clarify the terminology. In the same study, small populations with deep pedigrees and high levels of inbreeding were recommended for successful MAS for conifers. In *P. radiata*, at least eight QTL explaining a total of 14 % of the variation in juvenile wood were detected (Devey et al. 2004).

Linkage disequilibrium (LD): LD is the association of alleles. During gamete formation (e.g. sperm or egg) homolog chromosomes exchange DNA segments, a process called recombination. Alleles located on the same homolog chromosome tend to stay together because they are physically linked (they are in LD). Alleles far apart on a chromosome likely segregate (in linkage equilibrium or LE). For example, two loci Aa and Bb. If alleles A and B happen to be close to each other on the same chromosome, then they tend to be associated until recombination occurs. **LD is measured** using the frequency of gametes formed by alleles as $D = \text{freq}(AB) - \text{freq}(a)b - \text{freq}(Ab) \cdot \text{freq}(aB)$. A commonly used statistic to measure LD in a population is the coefficient of determination (r^2). It has a range of 0 to 1. Higher LD implies markers are tightly linked.

Homolog chromosomes: The two chromosomes, one inherited from the mother and one from the father, containing the same genetic loci in the same order.

Despite substantial efforts and resources put into discovering QTL and use for selection, few success stories have been reported in crop and animal breeding programs, mostly for traits with simple inheritance such as disease resistance (Dekkers 2004; Stuber et al. 1999; Xu and Crouch 2008). MAS has not been successful in forest trees. The key factors contributing to the failure of MAS in crop and animal breeding apply to forest trees as well. Limitations of MAS in forest trees were long ago pointed out (Strauss et al. 1992). One of the reasons why MAS failed in tree breeding is that QTL discovered in experimental populations generally explained a small percent ($< 5\%$) of phenotypic variance. For example, in *P. radiata*, QTL explained 0.78–3.8 % of variation in juvenile wood density and diameter (Devey et al. 2004), suggesting a more complex genetic architecture for traits. In fact, the data and large volume of published papers have consistently suggested that most of the traits of interest to improve such as growth and yield have complex inheritance patterns (Hill et al. 2008).

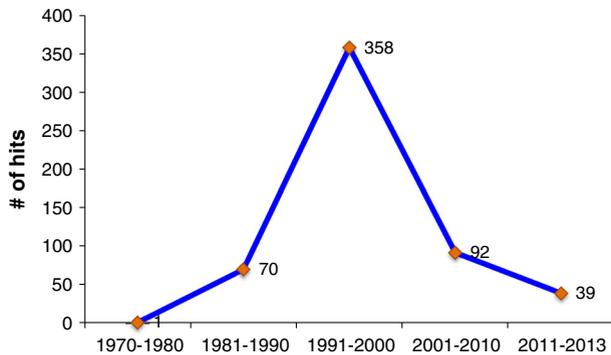


Fig. 1 Number of hits in Google Scholar searching for 'QTL, QTLs, forest trees, MAS' words that appear anywhere in manuscripts since 1970. There was a spike in the number of publications using these five words between 1990 and 2000 suggesting large efforts to discover QTLs for use in tree breeding

QTL-marker associations discovered in bi-parental populations were limited to those specific genetic backgrounds. QTL detection exploits the LD (association) between genetic markers and trait loci in a segregation population by crossing phenotypically different individuals (Sewell et al. 2000; Neale and Savolainen 2004). That is, two individuals with large phenotypic differences are crossed to produce large number of progeny (bi-parental population) for QTL detections. Marker-trait QTL phase (*coupling* vs *repulsion*) was not repeatable in other crosses.

Coupling: Refers to a case where dominant alleles (e.g. QA) are on the same homologue chromosome, and both recessive alleles (e.g. qa) are on the other homologue chromosome. Parental gametes are QA and qa , and the recombinant gametes are Qa and qA . In this phase, the A allele (used as a marker) is associated with the Q allele (the gene affecting phenotype).

Repulsion: Refers to the case where each homologous chromosome has one dominant and one recessive allele from two genes. The parental gametes are qA and Qa , and the recombinant gametes are QA and qa . In repulsion phase the A allele (the marker) is no longer associated with the Q allele (the gene affecting phenotype).

Other potential reasons for inconsistent results across families or populations are; (1) declaring loci as QTL while they are not (false positive); (2) genotype by environment interactions, and (3) overestimation of QTL effects (Beavis 1994; Dekkers 2004). Positioning of QTL in the genome has been far from precise (Holland 2004). For example, in cattle, the approximate confidence intervals for QTL positions are on the order of 20 cM (Dekkers and Hospital 2002). High-resolution QTL mapping (at least <5 cM away from gene) was suggested to develop reliable markers for MAS (Collard et al. 2005).

Centimorgan (cM): A unit used to measure genetic linkage. One cM equals 1 % chance that a marker on a chromosome will become separated from a second marker on the same chromosome due to crossing over in a single generation.

Forest trees are outcrossed species with large undomesticated breeding populations, and such populations have very low LD at the population level (Neale and Savolainen 2004). Low LD coupled with large genome size reduced the power and precision to detect QTLs

in a population of diverse genetic background. Howe et al. (2003) suggested that association studies could be used to detect QTL that are more likely to extend across families.

Genotype by environment interactions further complicated the application of MAS. Genotype by environment interactions in forest trees can be defined as the change in relative rank of families in different environments (Howe et al. 2003). In cases of significant genotype by environment interactions, the effect of QTL may need to be verified in different environments or new QTL need to be identified for several environments (Strauss et al. 1992). The added cost of QTL verification in several environments has been a major obstacle for forest tree species to implement MAS (Strauss et al. 1992; Howe et al. 2003).

Candidate gene approach

In recent years, forest geneticists and breeders have shown interest in candidate gene (CG) approaches. The basic concept of CG is to associate polymorphisms in known (sequenced) genes to phenotypes (Pflieger et al. 2001). In general, the CG approach follows a case-control approach that has been used in epidemiological studies. A population is divided into affected and unaffected groups for a trait. Genes and variants are selected for their possible roles (a priori hypotheses) in the phenotype, and the associations between DNA variants and phenotypes are tested (Tabor et al. 2002). The CG approach has been successfully applied in humans and in several animal species to link DNA polymorphisms to complex diseases (Tabor et al. 2002). With advancement in DNA sequencing technologies, it is possible to saturate the genome with dense markers and examine causal variation in genes (Pflieger et al. 2001; Morgante and Salamini 2003). The CG approach will continue to play a significant role in understanding the genetic basis of phenotypic variation and for fine mapping QTL. The information produced by the CG approach could transfer to plant breeding for traits with simple inheritance mechanisms. Neale and Savolainen (2004) suggested that low LD makes CG approach more feasible than *association genetics* for conifer breeding. However, the CG approach is unlikely to play a major role in forest tree breeding in the near future because of complex inheritance mechanisms observed for polygenic traits.

Genome-wide association study (GWAS) or association genetics: A method to dissect the genetic basis of complex traits and fine-scale genetic mapping. Association genetics aims to capture historical linkage disequilibrium (association between phenotype and markers) occurred throughout the evolutionary history in the undomesticated (natural) population. The method has been popular since availability of large number of markers to overcome the drawback of traditional QTL mapping (Ingvarsson and Street 2011).

Despite challenges in utilizing QTL and candidate genes in MAS, there are reasons to be cautiously optimistic. High throughput genotyping platforms are producing large number of markers at low cost to develop high-density genetic maps. Whole genome sequences of major forest tree species is underway (Sederoff 2013) to develop reference genomes of major forest species. With high density genetic maps, reference genome and increasing computer power; QTL discovery and their individual roles in controlling phenotype will be greatly facilitated in the near future and could be used in MAS.

Pedigree reconstruction in tree breeding programs

Discovery of microsatellite and SNP markers during the last decade opened new opportunities for biologists to study inheritance, evolutionary history, inbreeding, gene flow and conservation of wild and domesticated populations (Blouin 2003; Pemberton 2008). Estimation of relatedness (kinship) and parental assignment have been particularly popular (Blouin 2003). *Kinship* is a measure of the frequency of alleles (genes) shared by relatives. The alleles shared are either assumed to be identical by descent (inherited from the same ancestors) or identical by state (originated from the same locus but not necessarily from the same ancestor) (Falconer and Mackay 1996). *Parentage analysis* or *pedigree reconstruction* involves assigning individuals to a father or/and mother.

Microsatellite markers are preferred for parentage analysis because these markers are informative (*multi-allelic*, meaning that they have more than two forms of a gene), *co-dominant* (heterozygous individuals can be distinguished from homozygous individuals), and reliable (Jones et al. 2010). Although not as informative, *bi-allelic* SNP markers (having two forms of a gene) have become popular in recent years, because they are easy and cheaper to produce.

Based on simulation data, Anderson and Garza (2006) suggested that 60–100 SNP might allow accurate pedigree reconstruction for thousands of individuals. There are large numbers of statistical methods developed for kinship and parentage analysis from DNA markers; including parental exclusion, allocation and parental reconstruction (Jones and Ardren 2003; Jones et al. 2010).

Using DNA markers to correct pedigree errors, reconstruct full pedigree, and evaluate genetic diversity in domesticated and natural populations of forest trees has gained considerable attention in recent years (Moriguchi et al. 2004; Kumar et al. 2007; El-Kassaby and Lstibůrek 2009; Hansen and McKinney 2010; Grattapaglia et al. 2004; Doerksen and Herbinger 2010). I only highlighted a few related to breeding and selection in this section, since methodologies used in most studies are quite similar if not the same.

The concept of pedigree reconstruction to reduce controlled crossing efforts in forest tree breeding was first proposed by Lambeth et al. (2001). In the study, individuals are mated using mixed pollen from large number of trees, and DNA markers are used to construct the full pedigree. Authors called this system as *polymix breeding with parental analysis*. The feasibility of the system was evaluated in a loblolly pine third-generation breeding population with 45 selections. It was concluded that polymix breeding with paternity analysis could be a viable alternative to full-sib breeding and testing systems in loblolly pine. The potential key advantages listed were; simpler and less costly breeding and testing, correction of pedigree errors and as a result, greater genetic gain.

Microsatellite markers were used to identify superior offspring from a hybrid seed orchard of one *Eucalyptus grandis* seed parent and six *E. urophylla* pollen parents (Grattapaglia et al. 2004). A total of 14 microsatellite markers and 256 progeny from a 6 year-old stand were used for analyses. The authors reported 29 % pollen contamination. Based on full pedigree construction, and analysis of phenotype of 256 trees, pollen parents with poor general combining ability were culled. Offspring of the remaining selections in the seed orchard were used to establish new plantations. About 24 % more genetic gain in volume was reported in new stands at age 2–4, compared to before culling the poor pollen parents.

El-Kassaby and colleagues further developed the idea using real data in several species, and called the concept 'Breeding without Breeding' (El-Kassaby and Lstibůrek 2009; El-Kassaby et al. 2010, 2011). Using wind-pollinated seedlings instead of controlled crosses

for breeding, El-Kassaby and Lstibůrek (2009) reported the utility of full pedigree reconstruction in a large-scale operational progeny test of Douglas fir (*Pseudotsuga menziesii*), with 150 crosses tested on three different locations. They suggested that with pedigree reconstruction, the breeding cycle could be shortened for 4 years, and 75–85 % of the genetic response to selection could be captured compared to conventional breeding. In western larch (*Larix occidentalis* Nutt.), the effect of full pedigree reconstruction on variance components, and on accuracy of BV were evaluated using wind-pollinated progeny of 15 parents in a seed orchard with 41 selections (El-Kassaby et al. 2011). The authors concluded that, additive genetic variance and thus heritability are inflated upward in wind-pollinated progeny tests because a proportion of progeny may not be true half-sibs.

Pedigree of 596 individuals from a *Abies nordmanniana* open-pollinated progeny test were constructed using six microsatellite markers (Hansen and McKinney 2010). Trees originated from a seed orchard with 23 clones. Authors reported that, on average about 84 % of trees were true half-sibs, and the rest were either full-sibs, or partially inbred, with additive genetic relationship of 0.29. However, assuming the entire individual as half-sibs (the default assumption) did not affect the variance components or accuracy of BV noticeably. In *Picea rubens*, pedigree analysis was carried out to determine the male fertility in open-pollinated families and to correct pedigree errors (Doerksen and Herbinger 2010). Authors observed an increase in heritability with construction of full pedigree but did not see a significant effect on variance components, and on the accuracy of parental BV.

Parentage analysis and fingerprinting are not alternative tools to MAS or GS, though, we should expect that they would likely play a role in managing tree breeding programs. Forest tree breeders may use small numbers (<200) of SNP markers to correct pedigree errors, to estimate the amount of pollen contamination and measure genetic diversity. A more common application would be using mixed pollen collected from a number of trees to cross mother trees, and reconstruct full pedigree in the progeny tests (e.g. Lambeth et al. 2001, El-Kassaby and Lstibůrek 2009). Such a strategy could reduce breeding efforts and may improve genetic gain per breeding cycle.

Genomic selection in forest trees

In this section, characteristics and current status of major forest tree breeding programs are covered and a summary of literature on GS applications in forest trees is presented next. Challenges and prospects of implementation of GS in forest trees are outlined and are compared with animal breeding.

Characteristics of major forest tree breeding programs

Before we delve into GS in forest trees, we need to remind ourselves about the characteristics and current status of major forest tree breeding programs. This will help us to understand the challenges, and opportunities we face in implementation of GS. Compared to crop and animal breeding programs, forest tree breeding is still at its infancy. The most advanced programs are in their third or fourth cycle of breeding, with very little differentiation from natural populations. Breeding populations usually have large number of founders, with shallow pedigree. Progeny tests in breeding populations were not well connected in order to compare selections across different cycles or different years. Connections often relied on check seedlots composed of mixed unimproved seedlings with

unknown parents. Forest trees in general, and conifers in particular take many years to breed (5–15 years) and progeny test (3–15 years). Breeding programs are logistically more complex and costly. Because of large physical size of trees, it is difficult and expensive to cross trees. Progeny tests are expensive to establish, manage over many years, and measure. Large tracts of land are needed to test progeny. A vast majority of traits measured are surrogate traits, measured at early ages (3–15) with the expectations that they represent actual traits in mature ages (20–50 years).

In some tree breeding programs, genotype by environment interactions could be important (Wu and Colin 2005) and this needs to be addressed in development of genetic stock. Forest trees have large genetic variation for almost any trait; providing an opportunity to increase genetic gain considerably. Given the above major characteristics, if successful, the impact of GS on forest tree breeding would be far greater than in any other crop or animal breeding programs. GS will potentially impact all aspects of breeding and testing. However, genetic gain can be realized mainly from by reducing the progeny testing efforts and time, as we shall see later in an example.

Examples of GS in forest trees

There are relatively few studies on GS in forest trees, but we should expect a large number in the near future as genotyping platforms become available and the costs decrease. In the absence of real data, simulations can be helpful to understand various factors that might affect the outcomes. One of the early studies on GS in forest trees was based on simulations. A deterministic simulation approach was used to examine the effect of *effective population size* (N_e), marker density and heritability on the accuracy of *genomic estimated breeding values* (GEBV) (Grattapaglia and Resende 2011). It was suggested that a marker density of 2 markers/cM would be sufficient for a small population ($N_e < 30$) to achieve accuracy similar to *traditional genetic evaluation* based on progeny testing. For larger populations, a density of 10 or more markers/cM was suggested. The authors concluded that the effects of heritability and the number of QTLs (50 vs 100) on accuracy of prediction accuracies were negligible.

Genomic estimated breeding values (GEBV): Breeding value or genetic merit of an individual based on DNA markers without phenotype.

Effective population size (N_e): The number of individuals in an ideal population that would match the inbreeding rate (or loss of heterozygosity) of a given study population. Sometimes referred to as *inbreeding effective population size*. One way to approximate this statistic is through knowing the change in average inbreeding of the population (ΔF) from one generation to the next $\Delta F = 1/2N_e$ (Falconer and MacKay 1996).

Traditional genetic evaluation (BLUP): Breeders typically use progeny data and relatedness among individuals to predict breeding values. Linear mixed models are solved for predictions. The statistical method developed by Henderson in 1970s is called *best linear unbiased prediction* or BLUP (Henderson 1975).

Since 2011, a few empirical studies on GS in forest trees have been published, and they are encouraging. It is worth mentioning the contribution of the Conifer Translational Genomic Network project in the United States. The project was an Integrated Coordinated Agricultural Project among universities, the US Forest Service, and forest industry funded by a US Department of Agriculture grant in 2008. The goal was to bring genomic tools to

conifer breeding programs in the USA. SNP markers were identified through the resequencing of 7,535 *amplicons* for loblolly pine (Eckert et al. 2010). The project used Infinium SNP genotyping platform to genotype several thousand loblolly pine trees from three breeding programs in the USA. The first empirical studies on GS in forest trees came from the project collaborators.

Amplicons: A piece of DNA or RNA that is artificially amplified using PCR (polymerase chain reactions) to detect genetic variation among individuals.

Zapata-Valenzuela et al. (2012) used 3,406 SNP markers simultaneously as predictors in a cloned population of loblolly pine (534 cloned progeny from 13 crosses). In this study, only 149 clones with genotype data were available. The accuracy of GEBV ranged from 0.61 to 0.83 for cellulose and lignin content. These accuracies were comparable with the accuracy of BV based on pedigree and phenotype. However, for growth traits, the accuracy of predictions from markers was slightly lower. The authors suggested that with GS, the breeding of loblolly pine could be reduced from 15 years to about half that. Given a small pedigreed population, 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.

Resende et al. (2011) used 800 cloned progeny of a loblolly pine population and 4,825 SNP markers to test the efficiency of GS. In the study, accuracies of GEBV varied between 0.55 and 0.88, matching the accuracies achieved by conventional genetic evaluation. The authors suggested that with the application of GS, 53–112 % more genetic gain could be achieved in loblolly pine by reducing breeding cycles in half. The prediction models developed for ages 1–4 had low predictive ability at age 6 of the same population. Also, a model developed at one environment did not perform well in another environment, suggesting the effect of genotype by environment interactions and different marker-trait loci phase at different ages. Resende et al. (2012a, b) used the same population and marker genotypes to examine the effects of different statistical methods (regression and Bayesian methods) on 17 different traits. The predictive ability of models differed marginally for most traits, except for fusiform rust disease. For this trait, Bayesian statistical methods performed better than random regression method. The disease in loblolly pine is sometimes controlled by a few genes of large effect (Wilcox et al. 1996).

Resende et al. (2012a, b) examined the effectiveness of GS in two unrelated eucalypt breeding populations with different effective population sizes ($N_e = 11$ and 51). Cloned trees were genotyped with about 3,000 DArT markers. Accuracies of GEBV varied between 0.55 and 0.88, and were similar to accuracies obtained by traditional genetic evaluation. The models developed for one population were poor predictors for the other. The authors attributed poor performance of models across populations to different patterns of LD, inconsistent allelic effects, and genotype by environment interactions.

DArT markers: A method called *Diversity Array Technology* to produce markers. A fraction of genome (0.1 to 10 %) is sequenced to reduce complexity. The presence and absence of DNA segments are scored in individuals.

In the above applications, markers were fit simultaneously in statistical models and the effects of markers were summed to estimate GEBV. Another way to obtain GEBV is to estimate *realized genomic relationships* (G matrix) using the frequency of alleles shared by individuals. Then, *the additive genetic relationship matrix* (A matrix) derived from pedigree can be substituted with the G matrix in mixed models to predict GEBV (VanRaden

2008). This method is sometimes referred to as genomic BLUP or GBLUP (VanRaden et al. 2009).

Additive Genetic Relationships: Relatives resemble to each other because they share alleles (genes) that can be traced back to a common ancestor. Traditionally, geneticists have used pedigree to come up with the degree of relationships (measured as covariances) between relatives. For example, on average, the sibs share about 50 % of genes. The expected genetic relationship (covariance) between sibs would be 0.5.

Realized Genomic Relationships: The additive genetic relationship estimates based on pedigree are approximations. We know that some sibs may share more than 50 % of genes and some share less (not all sibs are alike). DNA markers allow geneticists to estimate genetic relationship coefficients (covariances) more accurately. Genetic relationships (sometimes called kinships coefficients) based on DNA markers are called *realized genomic relationships*.

The GBLUP method developed by cattle breeders is appealing for forest trees. Zapata-Valenzuela et al. (2013) used allele frequencies and regression methods suggested by Forni et al. (2011) to obtain realized genomic relationships matrices for a cloned population of loblolly pine to predict GEBV ($N = 165$ clones). The authors reported comparable accuracy values of GEBV from GBLUP to traditional genetic evaluation. The study showed that in the absence of phenotype, markers could capture the Mendelian segregation effect in full-sib families (Fig. 2). In contrast, no segregation was observed among full-sibs of a cross (x-axis) when pedigree based average additive genetic relationship matrix was used. In other words, with the A matrix, full-sibs progeny had the same mid-parent BV. Conversely, when the genomic relationship matrices were used, full-sibs segregated so each had a different breeding value, and a selection could be made within a cross without phenotype.

GBLUP could be a powerful tool in forest tree breeding to carry out within-family selection to optimize genetic gain and co-ancestry. Analysis of GBLUP is not different from traditional BLUP. Standard software available for linear mixed models can be used to solve mixed model equations while accounting for experimental design factors and genotype by environment interactions.

In forest tree breeding, a selection can be represented by large number of progeny because once breeding is done; hundreds of seeds can be harvested from the mother tree. In a breeding-testing cycle, tens of thousands of progeny might be available for assessment. Even if genotyping per individual cost goes down substantially, tree breeders may not be able to genotype the whole population due to logistical issues. A more realistic scenario would be genotyping a subset of the population to obtain *realized genomic relationships*. This matrix then could be added to the additive genetic relationship matrix of the whole population derived from pedigree. The hybrid genetic relationship matrix could be used in a BLUP approach to predict BV as suggested by animal breeders (Legarra et al. 2009; Forni et al. 2011; Chen et al. 2011; Aguilar et al. 2010).

Like GBLUP, predictions of BV based on hybrid genetic relationship matrices are expected to be popular in forest tree breeding. Using a cloned loblolly pine population ($N = 354$) and 3,461 SNP markers, Ogut (2012) obtained genomic relationship matrices for a subset of the population ($N_2 = 166$) and blended this with the additive genetic relationship matrix derived from pedigree of the whole population. Also, a simulation population mimicking loblolly pine ($N = 1,000$) was created to examine the effect of

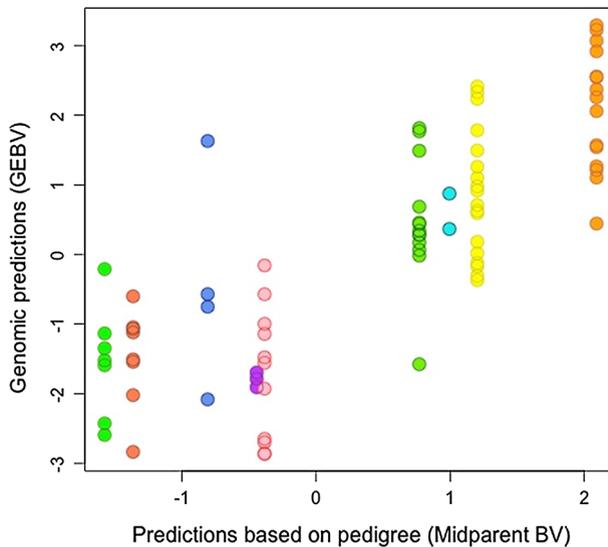


Fig. 2 Predicted breeding values of full-sibs from nine different crosses in loblolly pine using the pedigree (x-axis) and genomic predictions (y-axis). In the absence of phenotype, the expected breeding value of sibs is the same, which is mid-parent breeding value. However, sibs without phenotype can have different breeding values when markers are used (data from Zapata-Valenzuela et al. 2012)

training population size on accuracy of predictions. The accuracies of predictions from hybrid genetic relationship matrices were higher than traditional BLUP based on pedigree for both empirical and simulated data.

Although the above empirical GS studies are encouraging, they are proof of concept. GS can provide higher accuracies than traditional genetic evaluation if markers are in LD with QTL controlling phenotypic variation (Calus and Veerkamp 2007). Given the small population size and low marker density coverage in the above empirical studies, it is likely that markers are tracing large chromosome haplotypes segregating in these populations. In other words, markers are reconstructing the pedigrees among individuals coming from a handful of crosses. In the above studies, the same populations were split for model development and validation. The marker-trait loci phase does not change by splitting the populations to validate the predictive power of markers. In order to test the efficiency of GS in forest trees, the population needs to go through breeding to see the effect of recombinations on marker-trait phase. In other words, a model developed for the current generation should be tested in the subsequent generation to test the predictive ability of markers.

Factors affecting the accuracy GEBV and comparisons with animal breeding

Hayes et al. (2009) listed four factors critical on accuracy of predictions from GS: (1) The level of LD between markers and QTL, (2) effective population size (N_e), particularly the number of individuals with phenotype and genotype used to develop prediction model, (3) marker density, and (4) heritability of the trait. The effects of genotype by environment interactions on GS accuracy have not been covered as much but it is worth examination. The population size and marker density are two factors that can be controlled by breeders

(Hayes et al. (2009). Factors listed above are not necessarily independent from each other, but they are connected and affect the predictions accuracies together. Comparisons of factors between tree breeding and animal populations would be helpful to draw some conclusions about the challenges and opportunities tree breeders may face in implementation of GS.

The extent of **Linkage disequilibrium** (LD) between markers and QTL is an important factor in GS. In Holstein–Friesian (black and white) cattle, LD of $r^2 = 0.2$ occurs at approximately 100 kb (Hayes and Goddard 2010). For GS to be successful, the marker should explain substantial variance in adjacent QTL. In other words, the level of LD should be greater than 0.2 (Hayes et al. 2009). In a simulation using single SNP markers, the accuracy of GEBV increased from 0.68 to 0.82 when the LD among markers increased from 0.1 to 0.2 (Goddard 2009). In forest trees, genome-wide average LD is low, and declines rapidly between adjacent loci (Neale and Savolainen 2004). For example, in loblolly pine, LD decays to less than $r^2 = 0.2$ within 1,500 base pairs (Neale and Savolainen 2004). Markers-QTL linkages may not hold in such populations after breeding (because of frequent recombinations). Tree breeders will likely need more markers and larger training populations to develop stable GS prediction models.

The past **effective population size** (N_e) of a breeding population largely determines the extent of genome-wide LD (Hayes and Goddard 2010). Animal breeding populations typically have $N_e < 100$, and N_e of 30–60 are common (Mc Parland et al. 2007). Tree breeding programs are conservative, and maintain >200 individuals (status number), with a primary motivation of conserving genetic diversity (White et al. 2007). For example, the *P. radiata* breeding population in New Zealand had a target N_e of 400 (Jayawickrama and Carson 2000). The N_e for hybrid tree breeding populations (e.g. Eucalyptus) are smaller, with a range of 10–50 (White et al. 2007). For species with large N_e , more markers would be necessary to establish linkages between marker and QTL, and maintain the predictive power of markers across populations and environments (Goddard et al. 2011). The size of the training population had a relatively minor effect on accuracy of GEBV in simulation studies (Grattapaglia and Resende 2011). When the training population size was increased from 500 to 1,000, correlations between true BV and GEBV increased by only 0.08 (Meuwissen et al. 2001). However, forest tree breeders will likely need a larger training population to overcome disadvantages of large N_e , and loose relatedness in their breeding populations.

Higher marker density increased the prediction accuracies substantially. For example, increasing the spacing of markers (from 2 to 4 cM) reduced prediction accuracies by 9–13 %, because the LD is reduced (Meuwissen et al. 2001). In a more recent simulation study, the accuracy of predictions increased from 0.36 at 2 marker/cM density to 0.80 with 20 marker/CM density (Grattapaglia and Resende 2011). In the above scenarios, the N_e (100) and the heritability (0.2) were kept constant. Tree breeders will likely need a lot more markers than cattle breeders to cover the large genome of trees and compensate for low LD, so that the GEBV would be comparable with predictions from progeny testing.

Genetic linkage: The close association of genes or other DNA sequences on the same chromosome. The closer two genes (or loci) are to each other the more likely they will be inherited (segregate) together.

Heritability of the trait considered has effect on the accuracy of GEBV, provided that the markers are dense enough (Goddard 2009; Grattapaglia and Resende 2011). Based on a deterministic approach, Goddard (2009) demonstrated that heritability affected the accuracy of GEBV, particularly for low heritability traits. Traits with low heritabilities required

a larger training population than did traits with high heritabilities. For example, in order to obtain an accuracy of 0.6, almost 3,000 phenotypic records were needed for a heritability of 0.3. This number decreased to about 1,500 for a heritability of 0.4 (Goddard 2009). In these scenarios, the effective population size was $N_e = 100$. Many traits of importance to tree breeders have low heritability (0.10–0.30) (White et al. 2007). This is also the case for traits important in animal breeding. The average heritability of 27 traits important in cattle breeding was 0.25 (VanRaden et al. 2009). Using stochastic simulations, Grattapaglia and Resende (2011) suggested that the impact of heritability would be small on the predicted accuracies. Tree breeders still need to use the best experimental practices and data analysis models to increase the heritability to eliminate noise from phenotypic data. This in turn would help to develop better prediction models.

Genotype by environment interactions (GxE) affects the structure of breeding programs. In animal breeding, if correlations between two environments are <0.8 , different animals tend to be selected for two environments (Goddard and Hayes 2007). In cattle, a GS model fit in one population was not successful in another unrelated population (Hayes et al. 2009). In forest trees, the scales of genotype by environment interactions are generally larger when the individuals are planted on contrasting sites (White et al. 2007). However, GxE does not appear to be a problem for forest trees as long as environmental extremes are avoided, e.g. in loblolly pine in the southern USA (McKeand et al. 2006). In a cloned population of loblolly pine planted across four climatically different sites in the southern USA, GS prediction models developed for individual sites performed poorly across the different sites (Resende et al. 2012a, b). In *Eucalyptus* sp. hybrid populations in Brazil, when a model developed for one population was tested in another population, accuracies decreased substantially (Resende et al. 2012a, b). The authors of both studies attributed the decline in accuracies of predictions to different marker-QTL phase and GxE. Although, the data on performance of GS models across different environments and populations are very limited, it is likely that the GS models in forest trees will be population specific.

Outlook for the future

The genomic era is bringing new opportunities to forest tree breeders. Whole genome sequencing of major forest tree species has been completed or is underway. *Populus trichocarpa* was the first forest tree whose genome was sequenced (Tuskan et al. 2006). The *E. grandis* genome sequence was released in 2011 (Myburg et al. 2011). The draft assembly of Norway spruce (*Picea abies* L.) was completed in May 2013 (Nystedt et al. 2013); the first available genome of any conifer species. The initial draft assembly of the *P. taeda* genome was completed in 2014 (<http://pinegenome.org/pinerefseq/>). White spruce (*Picea glauca*) genome sequence was completed in 2013 (Birol et al. 2013). Reference sequenced genomes of forest trees will serve as invaluable resources for gene discovery, SNP marker detection, construction of high-density genetic maps, and development of high-throughput genotyping platforms. Such tools will be invaluable for tree improvement in the future. It is an exciting time to be a quantitative geneticist and tree breeder.

Routine application of GS in forest trees is yet to be realized. Forest tree breeders need large number of markers to populate the genome and to explore the LD between markers and many QTL with small effects. Despite advances in efficiency of genotyping technologies, genotyping is still costly for forest trees. Current commercially available genotyping platforms (Illumina, SNP genotyping platform) cost about \$150 per sample, though

the cost is rapidly decreasing. Elshire et al. (2011) suggested a procedure called ‘genotyping by sequencing’ or GbS to construct libraries based on reducing genome complexity with DNA restriction enzymes. The method is simple, quick and considered significantly cheaper compared to SNP chips.

Several labs have been exploring this technology for forest trees. The method employs methylation-sensitive restriction enzymes to enrich gene regions in the pine genome (Chen et al. 2013). To reduce costs, the method uses “barcoded” adapters and an index primer set to allow combining individual samples in a single sequencing lane (Ross Whetten, personal communication, NC State University). Whetten’s lab also modified a DNA extraction protocol originally developed by Ivanova et al. (2008) to reduce the cost. With this new method, the estimated cost for consumables for DNA extraction is about \$0.5 per sample. This is a significant reduction in cost compared to commercially available kits (Ross Whetten, personal communication, NC State University). GbS and newly developed genotyping platforms are yet to be tested for their reliability and repeatability. A chip designed by a consortium to share the cost and to be used by multiple organizations could be a solution, and was developed for eucalypt species (Dario Grattapaglia, Emprapa, Brazil, personal communication, July 2013).

The challenge remains to mine very large amounts of data and convert them to knowledge using statistical algorithms and computing power. Routine application of GS will require new computing skills to manage and process data. Well-structured (with known pedigrees) and precisely phenotyped populations are needed to implement GS. In the meantime, tree breeders should explore strategies about how to utilize large genomic data for decision-making; from prediction of BV to new breeding, testing, and selection strategies. A GS selection cycle is presented in Fig. 3. While a single occurrence of model training is shown, training can be performed iteratively as new phenotype and marker data accumulate. The only difference from traditional recurrent selection is that progeny testing is given a minor role for predictions of BV. Progeny testing is still expected to be important in the first few cycles but should diminish as more trees are genotyped and as the population goes through multiple breeding cycles. By that time we should have inexpensive whole genome sequencing available for major forest trees.

A GS plan for loblolly pine

It is always dangerous to predict the future, particularly when it comes to predicting genotyping costs that are changing so quickly. A possible GS plan for a loblolly pine elite breeding population in the North Carolina State University Cooperative Tree Improvement Program is given (Fig. 4). The elite breeding populations are subsets of the mainline population, and are intensively managed for short-term genetic gain.

BASE Population: GS processes start from a base population, and this population should be precisely phenotyped and should have some history of breeding to build LD. The base population consists of 21 parents and 50 crosses. Approximately 2,000 progeny were cloned via rooted cuttings and were field tested in eight locations in 2010. Trees were measured in the autumn of 2013 for growth and stem quality traits at age four. Since each progeny is cloned and tested on multiple locations, we expect more reliable phenotypes compared to seedling progeny. The parents and 2,000 progeny will be genotyped in 2014 using GbS. A GS model (M1) will be developed and validated by drawing random samples from the population. This will be a proof-of-concept. The top ~50 individuals, with a constraint on coancestry, will be selected to breed for the next cycle.

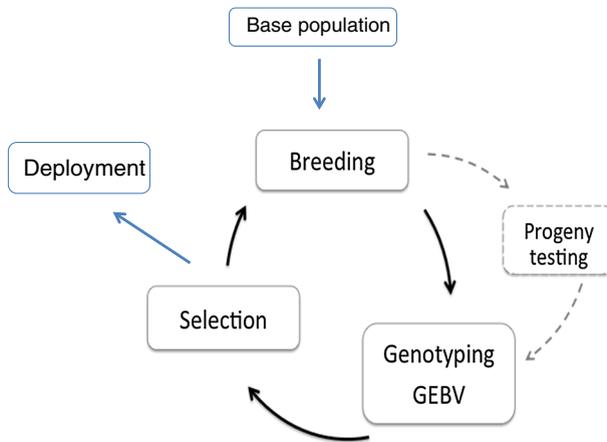


Fig. 3 A schematic presentation of GS cycle in forest trees. GS starts with a base population, composed of ~50 individuals with known pedigree and breeding values. Individuals are crossed (breeding), progeny are genotyped, and selections are made based on markers for next round of breeding. The cycle is repeated (recurrent selection). The cycle is not different from the recurrent selection well known by breeders, except that selection is based on markers (GS). GS will have a growing role as the program moves forward. The *dashed lines* connecting the progeny testing to the cycle suggest a decreasing role of field progeny testing. GS increases genetic gain per time by circumventing the progeny testing

Cycle 1 (2014–2025): Top grafting of ~50 selections in a breeding center will be completed in February–March 2014. Development of male and female strobili will take at least 2 years. Breeding could be completed in the winter of 2016 but will likely take until 2017. Cones will be harvested in October 2018, and seedlings will be raised in 2019. About 7,000 full-sib progeny will be genotyped using >50 K DNA markers (in 2019 and 2020), and a subset (~2,500) will be field-tested in the winter of 2020. Cloning of progeny is not envisioned after the base population to save time (2 years). The loss in accuracy in phenotypes could be compensated by using a higher marker density (>50 K) and by genotyping a larger number of progeny (7,000). Progeny will be phenotyped in the autumn of 2023 when they are 4 years old. The model developed for the base population will be retrained using additional markers and phenotypes (Model 2). The new model will be used to select ~50 individuals to breed in 2025.

Cycle 2 (2026–2032): The top ~50 selections will be top-grafted in winter of 2026. They will be bred 3 years later (winter of 2029) when enough male and female strobili are available. Cones will be harvested in October of 2031. In subsequent months, seeds will be germinated to extract DNA for genotyping. No progeny testing is envisioned. Depending on the genotyping cost, a larger number (~10,000) of progeny could be genotyped using higher density markers (>100 K). Model 2 will be retrained using additional genotypes. A new round of selections (~50) will be made in the winter of 2032 for the next cycle of breeding. The breeding cycle is expected to be 6 years; a substantial decrease compared to Cycle 1 (11 years or longer).

The plan envisions increasing marker density substantially (more than twofold) after each breeding cycle commensurate decreases in genotyping costs. After Cycle 1, no further progeny testing is considered for the future generations, but breeders may progeny test a subset of individuals to further train the prediction models because as the population goes through multiple breeding cycles (recombination), the LD between marker and QTL will

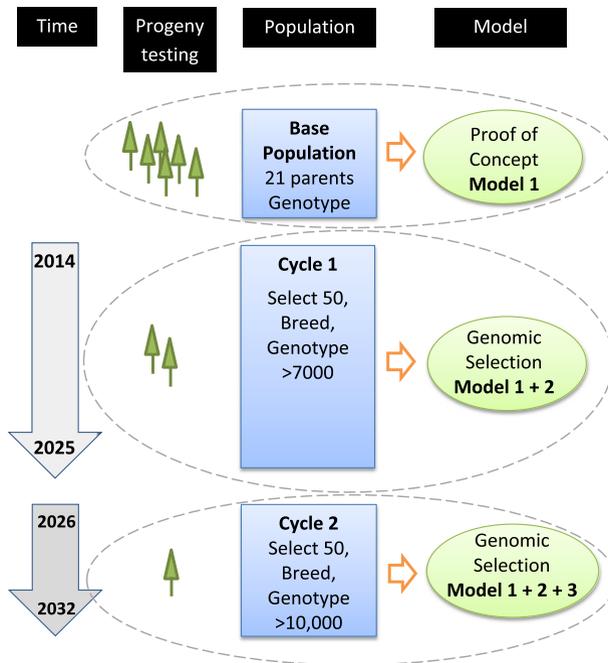


Fig. 4 Genomic selection plan for a loblolly pine breeding population. Base population has 2,000 cloned progeny from 50 crosses. Population was measured at age four in 2014, and will be genotyped. Progeny testing efforts are expected to decrease as the markers-based model is retrained. In cycle 2, selections will be based on GS without phenotype. A small subset of progeny will be field tested to later retrain the model. Number of trees to be genotyped given in Cycles 1 and 2 are conservative. We should expect larger number of trees to be genotyped as genotyping platforms become mature with complete genome sequencing of *Pinus taeda*

degrade and the predictive power of model will diminish. With GS, the breeding cycle time drops almost in half, because selection decisions could be made based on markers without phenotype as soon as seeds are available. This could increase genetic gain twofold per year.

Conclusions

There are many reasons to be optimistic about GS prospects in forest trees. Abundant and cost efficient DNA markers will likely be available to tree breeders, and forest geneticists in the near future. The impact of GS approaches on forest tree breeding will be much higher than any crop species because of long cycles of breeding. Still there are hurdles to overcome. A well organized system, including an efficient database, computing and quantitative skills will be needed to manage and process large data sets. Forest geneticists should avoid over-promising GS to the community until all the necessary ingredients are in place.

Acknowledgments This work was supported by the Conifer Translational Genomics Network Coordinated Agricultural Project (USDA NRI #2007-02781 and AFRI #2009-01879), the NC State University Cooperative Tree Improvement Program, and the Department of Forestry and Environmental Resources at NCSU. Funding was also provided by the Southeastern Partnership for Integrated Biomass Supply Systems

(IBSS), a coordinated agricultural project funded by the USDA Agriculture and Food Research Initiative (AFRI) Sustainable Bioenergy Program; and by the Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP) A Coordinated Agricultural Project funded by the USDA National Institute of Food and Agriculture. I thank my colleagues Drs. Steve McKeand, Ross Whetten John Frampton (North Carolina State University, USA) and Christophe Plomion (INRA Pierroton, France) for valuable discussions on the manuscript. I am thankful to two anonymous reviewers for their excellent suggestions and editorial edits (reviewer 1) to improve the manuscript. I used several web resources for definitions of some technical terms. The following are worth to mention <http://ghr.nlm.nih.gov/glossary> and <http://www.ndsu.edu/pubweb/~mcclean/plsc431/linkage/linkage2.htm>.

References

- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci* 93(2):743–752. doi:[10.3168/jds.2009-2730](https://doi.org/10.3168/jds.2009-2730)
- Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172(4):2567–2582. doi:[10.1534/genetics.105.048074](https://doi.org/10.1534/genetics.105.048074)
- Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Proceedings of the forty-ninth annual corn and sorghum industry research conference. American Seed Trade Association, Washington, DC, pp 250–266
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci* 48(5):1649. doi:[10.2135/cropsci2008.03.0131](https://doi.org/10.2135/cropsci2008.03.0131)
- Biról I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Saint Yuen MM et al (2013) Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics*. doi:[10.1093/bioinformatics/btt178](https://doi.org/10.1093/bioinformatics/btt178)
- Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends Ecol Evol* 18(10):503–511. doi:[10.1016/S0169-5347\(03\)00225-8](https://doi.org/10.1016/S0169-5347(03)00225-8)
- Boichard D, Guillaume F, Baur A, Croiseau P, Rossignol MN, Boscher MY, Druet T et al (2012) Genomic selection in French dairy cattle. *Anim Prod Sci* 52(3):115. doi:[10.1071/AN11119](https://doi.org/10.1071/AN11119)
- Calus MPL, Veerkamp RF (2007) Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *J Anim Breed Genet* 124(6):362–368. doi:[10.1111/j.1439-0388.2007.00691.x](https://doi.org/10.1111/j.1439-0388.2007.00691.x)
- Chagné D, Brown G, Lalanne C, Madur D, Pot D, Neale D, Plomion C (2003) Comparative genome and QTL mapping between maritime and loblolly pines. *Mol Breeding* 12(3):185–195. doi:[10.1023/A:1026318327911](https://doi.org/10.1023/A:1026318327911)
- Chen CY, Misztal I, Aguilar I, Legarra A, Muir WM (2011) Effect of different genomic relationship matrices on accuracy and scale. *J Anim Sci* 89(9):2673–2679. doi:[10.2527/jas.2010-3555](https://doi.org/10.2527/jas.2010-3555)
- Chen C, Mitchell SE, Elshire RJ, Buckler ES, El-Kassaby YA (2013) Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genet Genomes* 9(6):1537–1544. doi:[10.1007/s11295-013-0657-1](https://doi.org/10.1007/s11295-013-0657-1)
- Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (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
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12(7):499–510. doi:[10.1038/nrg3012](https://doi.org/10.1038/nrg3012)
- Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J Anim Sci* 82(13 suppl):E313–E328
- Dekkers JCM, Hospital F (2002) Multifactorial genetics the use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet* 3(1):22–32. doi:[10.1038/nrg701](https://doi.org/10.1038/nrg701)
- Devey ME, Carson SD, Nolan MF, Matheson AC, Te Riini C, Hohepa J (2004) QTL associations for density and diameter in *Pinus radiata* and the potential for marker-aided selection. *Theor Appl Genet* 108(3):516–524. doi:[10.1007/s00122-003-1446-2](https://doi.org/10.1007/s00122-003-1446-2)
- Doerksen TK, Herburger CM (2010) Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. *Tree Genet Genomes* 6(4):591–600. doi:[10.1007/s11295-010-0274-1](https://doi.org/10.1007/s11295-010-0274-1)
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185(3):969–982. doi:[10.1534/genetics.110.115543](https://doi.org/10.1534/genetics.110.115543)

- Eggen A (2012) The development and application of genomic selection as a new breeding paradigm. *Anim Front* 2(1):10–15. doi:[10.2527/af.2011-0027](https://doi.org/10.2527/af.2011-0027)
- El-Kassaby Y, Lstibůrek M (2009) Breeding without breeding. *Genet Res* 91(02):111–120. doi:[10.1017/S001667230900007X](https://doi.org/10.1017/S001667230900007X)
- El-Kassaby YA, Funda T, Lai BSK (2010) Female reproductive success variation in a *Pseudotsuga menziesii* seed orchard as revealed by pedigree reconstruction from a bulk seed collection. *J Hered* 101(2):164–168. doi:[10.1093/jhered/esp126](https://doi.org/10.1093/jhered/esp126)
- El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klápště J, Lstibůrek M (2011) Breeding without breeding: is a complete pedigree necessary for efficient breeding? *PLoS One* 6(10):e25737
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6(5):e19379. doi:[10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379)
- Emebiri LC, Devey ME, Matheson AC, Slee MU (1997) Linkage of RAPD markers to NESTUR, a stem growth index in radiata pine seedlings. *Theor Appl Genet* 95(1–2):119–124. doi:[10.1007/s001220050539](https://doi.org/10.1007/s001220050539)
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Technical, Essex
- Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol* 43(1). <http://www.biomedcentral.com/content/pdf/1297-9686-43-1.pdf>
- Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD et al (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One* 6(12):e28334. doi:[10.1371/journal.pone.0028334](https://doi.org/10.1371/journal.pone.0028334)
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136(2):245–257. doi:[10.1007/s10709-008-9308-0](https://doi.org/10.1007/s10709-008-9308-0)
- Goddard Me, Hayes BJ (2007) Genomic selection. *J Anim Breed Genet* 124(6):323–330. doi:[10.1111/j.1439-0388.2007.00702.x](https://doi.org/10.1111/j.1439-0388.2007.00702.x)
- Goddard ME, Hayes BJ, Meuwissen THE (2011) Genomic selection in livestock populations. *Genet Res* 92(5):413
- Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. *Tree Genet Genomes* 7(2):241–255
- Grattapaglia D, Ribeiro VJ, Rezende GDSP (2004) Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for eucalyptus. *Theor Appl Genet* 109(1):192–199. doi:[10.1007/s00122-004-1617-9](https://doi.org/10.1007/s00122-004-1617-9)
- Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchel-Olds T, Sherman B, Vujcic S, Williams C, Neale D (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* 138(4):1293–1300
- Guimarães EP (2007) Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish. Food & Agriculture Org, Rome, Italy
- Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101(1):5–18. doi:[10.1038/hdy.2008.35](https://doi.org/10.1038/hdy.2008.35)
- Hansen OK, McKinney LV (2010) Establishment of a quasi-field trial in *Abies nordmanniana*—test of a new approach to forest tree breeding. *Tree Genet Genomes* 6(2):345–355. doi:[10.1007/s11295-009-0253-6](https://doi.org/10.1007/s11295-009-0253-6)
- Hansen OK, Nielsen UB (2010) Microsatellites used to establish full pedigree in a half-sib trial and correlation between number of male strobili and paternal success. *Ann For Sci* 67(7):703. doi:[10.1051/forest/2010028](https://doi.org/10.1051/forest/2010028)
- Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. *Genome* 53(11):876–883. doi:[10.1139/G10-076](https://doi.org/10.1139/G10-076)
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92(2):433–443. doi:[10.3168/jds.2008-1646](https://doi.org/10.3168/jds.2008-1646)
- Hayes BJ, Cogan NOI, Pemberton LW, Goddard ME, Wang J, Spangenberg GC, Forster JW (2013) Prospects for genomic selection in forage plant species. *Plant Breeding* 132(2):133–143. doi:[10.1111/pbr.12037](https://doi.org/10.1111/pbr.12037)
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. *Crop Sci* 49(1):1–12
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31(2):423. doi:[10.2307/2529430](https://doi.org/10.2307/2529430)
- Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4(2):e1000008. doi:[10.1371/journal.pgen.1000008](https://doi.org/10.1371/journal.pgen.1000008)

- Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities. In: New directions for a diverse planet. Proceedings of the 4th international crop science congress, The Regional Institute Ltd., Gosford, NSW, Australia (www.Cropscience.Org.Au). http://cropscience.org.au/icsc2004/pdf/203_hollandjb.pdf
- Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH (2003) From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Can J Bot* 81(12):1247–1266. doi:[10.1139/B03-141](https://doi.org/10.1139/B03-141)
- Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. *New Phytol* 189(4):909–922. doi:[10.1111/j.1469-8137.2010.03593.x](https://doi.org/10.1111/j.1469-8137.2010.03593.x)
- Isik F, Amerson HV, Whetten RW, Garcia SA, McKeand SE (2012) Interactions of Fr genes and mixed-pathogen inocula in the loblolly pine-fusiform rust pathosystem. *Tree Genet Genomes* 8(1):15–25. doi:[10.1007/s11295-011-0416-0](https://doi.org/10.1007/s11295-011-0416-0)
- Ivanova NV, Fazekas AJ, Hebert PDN (2008) Semi-automated, membrane-based protocol for DNA isolation from plants. *Plant Mol Biol Report* 26(3):186–198. doi:[10.1007/s11105-008-0029-4](https://doi.org/10.1007/s11105-008-0029-4)
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9(2):166–177. doi:[10.1093/bfpg/elq001](https://doi.org/10.1093/bfpg/elq001)
- Jayawickrama KJS, Carson MJ (2000) A breeding strategy for the New Zealand radiata pine breeding cooperative. *Silvae Genetica* 49(2):82–89
- Jermstad KD, Bassoni DL, Jech KS, Wheeler NC, Neale DB (2001) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. I. Timing of vegetative bud flush. *Theor Appl Genet* 102(8):1142–1151. doi:[10.1007/s001220000505](https://doi.org/10.1007/s001220000505)
- Jones AG, Adren WR (2003) Methods of parentage analysis in natural populations. *Mol Ecol* 12(10):2511–2523. doi:[10.1046/j.1365-294X.2003.01928.x](https://doi.org/10.1046/j.1365-294X.2003.01928.x)
- Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. *Mol Ecol Resour* 10(1):6–30. doi:[10.1111/j.1755-0998.2009.02778.x](https://doi.org/10.1111/j.1755-0998.2009.02778.x)
- Kranis A, Gheyas AA, Boschiero C, Turner F, Le Y, Smith S, Talbot R et al (2013) Development of a high density 600K SNP genotyping array for chicken. *BMC Genom* 14(1):59. doi:[10.1186/1471-2164-14-59](https://doi.org/10.1186/1471-2164-14-59)
- Kumar S, Garrick DJ (2001) Genetic response to within-family selection using molecular markers in some radiata pine breeding schemes. *Can J For Res* 31(5):779–785. doi:[10.1139/x01-009](https://doi.org/10.1139/x01-009)
- Kumar S, Gerber S, Richardson TE, Gea L (2007) Testing for unequal paternal contributions using nuclear and chloroplast SSR markers in polycross families of radiata pine. *Tree Genet Genomes* 3(3):207–214. doi:[10.1007/s11295-006-0056-y](https://doi.org/10.1007/s11295-006-0056-y)
- Lambeth C, Lee BC, O'Malley D, Wheeler N (2001) Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet* 103:930–943
- Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. *J Dairy Sci* 92(9):4656–4663. doi:[10.3168/jds.2009-2061](https://doi.org/10.3168/jds.2009-2061)
- Lerceteau E, Plomion C, Andersson B (2000) AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*: a preliminary study. *Mol Breeding* 6(5):451–458. doi:[10.1023/A:1026548716320](https://doi.org/10.1023/A:1026548716320)
- Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet* 120(1):151–161. doi:[10.1007/s00122-009-1166-3](https://doi.org/10.1007/s00122-009-1166-3)
- McParland S, Kearney JF, Rath M, Berry DP (2007) Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *J Anim Sci* 85(2):322–331. doi:[10.2527/jas.2006-367](https://doi.org/10.2527/jas.2006-367)
- McKeand SE, Jokela EJ, Huber DA, Byram TD, Lee Allen H, Bailian L, Mullin TJ (2006) Performance of improved genotypes of loblolly pine across different soils, climates, and silvicultural inputs. *For Ecol Manage* 227(1–2):178–184. doi:[10.1016/j.foreco.2006.02.016](https://doi.org/10.1016/j.foreco.2006.02.016)
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Morgante M, Salamini F (2003) From plant genomics to breeding practice. *Curr Opin Biotechnol* 14(2):214–219. doi:[10.1016/S0958-1669\(03\)00028-4](https://doi.org/10.1016/S0958-1669(03)00028-4)
- Moriguchi Y, Taira H, Tani N, Tsumura Y (2004) Variation of paternal contribution in a seed orchard of *Cryptomeria japonica* determined using microsatellite markers. *Can J For Res* 34(8):1683–1690. doi:[10.1139/x04-029](https://doi.org/10.1139/x04-029)
- Myburg A, Grattapaglia D, Tuskan G, Jenkins J, Schmutz J, Mizrahi E, Hefer C et al (2011) The *Eucalyptus grandis* genome project: genome and transcriptome resources for comparative analysis of woody plant biology. *BMC Proc* 5(Suppl 7):120. doi:[10.1186/1753-6561-5-S7-120](https://doi.org/10.1186/1753-6561-5-S7-120)
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci* 9(7):325–330

- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F et al (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497(7451):579–584. doi:[10.1038/nature12211](https://doi.org/10.1038/nature12211)
- O'Malley DM, McKeand SE (1994) Marker assisted selection for breeding value in forest trees. *Forest Genetics* 1(4):207–218
- Ogut F (2012) Predictions of genetic merit in tree breeding using factor analytic linear mixed models and blended genomic relationship matrices. Thesis, North Carolina State University, Raleigh, NC, USA. <http://www.lib.ncsu.edu/resolver/1840.16/8252>
- Pemberton JM (2008) Wild pedigrees: the way forward. *Proc R Soc B Biol Sci* 275(1635):613–621. doi:[10.1098/rspb.2007.1531](https://doi.org/10.1098/rspb.2007.1531)
- Perkel J (2008) SNP genotyping: six technologies that keyed a revolution. *Nat Methods* 5(5):447–453. doi:[10.1038/nmeth0508-447](https://doi.org/10.1038/nmeth0508-447)
- Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. *Mol Breeding* 7(4):275–291
- Plomion C, Durel C-E, O'Malley DM (1996) Genetic dissection of height in maritime pine seedlings raised under accelerated growth conditions. *Theor Appl Genet* 93(5–6):849–858. doi:[10.1007/BF00224085](https://doi.org/10.1007/BF00224085)
- Pot D, Rodrigues J-C, Rozenberg P, Chantre G, Tibbits J, Cahalan C, Pichavant F, Plomion C (2006) QTLs and candidate genes for wood properties in maritime pine (*Pinus pinaster* Ait.). *Tree Genet Genomes* 2(1):10–24. doi:[10.1007/s11295-005-0026-9](https://doi.org/10.1007/s11295-005-0026-9)
- Resende MFR Jr, Munoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M (2011) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytol*. doi:[10.1111/j.1469-8137.2011.03895.x/full](https://doi.org/10.1111/j.1469-8137.2011.03895.x/full)
- Resende MFR Jr, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012a) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190(4):1503–1510
- Resende MDV, Resende Jr MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA (2012) Genomic selection for growth and wood quality in eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol*. doi:[10.1111/j.1469-8137.2011.04038.x/full](https://doi.org/10.1111/j.1469-8137.2011.04038.x/full)
- Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* 123(4):218–223
- Schefers JM, Weigel KA (2012) Genomic selection in dairy cattle: integration of DNA testing into breeding programs. *Anim Front* 2(1):4–9. doi:[10.2527/af.2011-0032](https://doi.org/10.2527/af.2011-0032)
- Sederoff R (2013) Genomics: a spruce sequence. *Nature* 497(7451):569–570. doi:[10.1038/nature12250](https://doi.org/10.1038/nature12250)
- Sewell MM, Bassoni DL, Megraw RA, Wheeler NC, Neale DB (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. *Theor Appl Genet* 101(8):1273–1281. doi:[10.1007/s001220051607](https://doi.org/10.1007/s001220051607)
- Strauss S, Lande R, Namkoong G (1992) Limitations of molecular-marker-aided selection in forest tree breeding. *Can J For Res* 22:1051–1061
- Stuber CW, Polacco M, Lynn Senior M (1999) Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential. *Crop Sci* 39(6):1571. doi:[10.2135/cropsci1999.3961571x](https://doi.org/10.2135/cropsci1999.3961571x)
- Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 3(5):391–396
- Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793):1596–1604. doi:[10.1126/science.1128691](https://doi.org/10.1126/science.1128691)
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91(11):4414–4423
- VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci* 92(1):16–24. doi:[10.3168/jds.2008-1514](https://doi.org/10.3168/jds.2008-1514)
- White TL, Adams TW, Neale DB (2007) Forest genetics. CABI, Cambridge, MA, USA
- Wilcox PL, Amerson HV, Kuhlman EG, Liu BH, O'Malley DM, Sederoff RR (1996) Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. *Proc Natl Acad Sci USA* 93(9):3859–3864
- Wilcox PL, Carson SD, Richardson TE, Ball RD, Horgan GP, Carter P (2001) Benefit-cost analysis of DNA marker-based selection in progenies of *Pinus radiata* seed orchard parents. *Can J For Res* 31(12):2213–2224. doi:[10.1139/x01-144](https://doi.org/10.1139/x01-144)
- Williams CG, Neale DB (1992) Conifer wood quality and marker-aided selection: a case study. *Can J For Res* 22(7):1009–1017. doi:[10.1139/x92-135](https://doi.org/10.1139/x92-135)

- Wu HX, Colin MA (2005) Genotype by environment interactions in an Australia-wide radiata pine diallel mating experiment: implications for regionalized breeding. *For Sci* 51(1):29–40
- Würschum T, Reif JC, Kraft T, Janssen G, Zhao Y (2013) Genomic selection in sugar beet breeding populations. *BMC Genetics* 14(1):85. doi:[10.1186/1471-2156-14-85](https://doi.org/10.1186/1471-2156-14-85)
- Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci* 48(2):391. doi:[10.2135/cropsci2007.04.0191](https://doi.org/10.2135/cropsci2007.04.0191)
- Zapata-Valenzuela J, Isik F, Maltecca C, Wegrzyn J, Neale D, McKeand S, Whetten R (2012) SNP markers trace familial linkages in a cloned population of *Pinus taeda*—prospects for genomic selection. *Tree Genet Genomes*. doi:[10.1007/s11295-012-0516-5](https://doi.org/10.1007/s11295-012-0516-5)
- Zapata-Valenzuela J, Whetten RW, Neale DB, McKeand SE, Isik F (2013) Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. *G3: Genes|Genomes|Genetics*. doi:[10.1534/g3.113.005975](https://doi.org/10.1534/g3.113.005975). <http://www.g3journal.org/content/early/2013/03/29/g3.113.005975>
- Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012) Accuracy of genomic selection in european maize elite breeding populations. *Theor Appl Genet* 124(4):769–776. doi:[10.1007/s00122-011-1745-y](https://doi.org/10.1007/s00122-011-1745-y)