Quantitative Genetics and Genomics Converge to Accelerate Forest Tree Breeding

Dario Grattapaglia [1,2,3,4*], Orzenil B. Silva-Junior [1,2], Rafael T. Resende [1],
Eduardo P. Cappa [5,6], Bárbara S. F. Müller [1,3], Biyue Tan [7], Fikret Isik [4], Blaise Ratcliffe [8] and
Yousry A. El-Kassaby [8]

[1] EMBRAPA Recursos Genéticos e Biotecnologia, Brasília, Brazil, [2] Programa de Ciências Genômicas e Biotecnologia, Universidade Católica de Brasília, Brasília, Brazil, [3] Departamento de Biologia Celular, Universidade de Brasília, Brasília, Brazil, [4] Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, United States, [5] Centro de Investigación de Recursos Naturales, Instituto de Recursos Biológicos, INTA, Buenos Aires, Argentina, [6] Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina, [7] Biomaterials Division, Stora Enso AB, Stockholm, Sweden, [8] Department of Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, BC, Canada

Forest tree breeding has been successful at delivering genetically improved material for multiple traits based on recurrent cycles of selection, mating, and testing. However, long breeding cycles, late flowering, variable juvenile-mature correlations, emerging pests and diseases, climate, and market changes, all pose formidable challenges. Genetic dissection approaches such as quantitative trait mapping and association genetics have been fruitless to effectively drive operational marker-assisted selection (MAS) in forest trees, largely because of the complex multifactorial inheritance of most, if not all traits of interest. The convergence of high-throughput genomics and quantitative genetics has established two new paradigms that are changing contemporary tree breeding dogmas. Genomic selection (GS) uses large number of genome-wide markers to predict complex phenotypes. It has the potential to accelerate breeding cycles, increase selection intensity and improve the accuracy of breeding values. Realized genomic relationships matrices, on the other hand, provide innovations in genetic parameters' estimation and breeding approaches by tracking the variation arising from random Mendelian segregation in pedigrees. In light of a recent flow of promising experimental results, here we briefly review the main concepts, analytical tools and remaining challenges that currently underlie the application of genomics data to tree breeding. With easy and cost-effective genotyping, we are now at the brink of extensive adoption of GS in tree breeding. Areas for future GS research include optimizing strategies for updating prediction models, adding validated functional genomics data to improve prediction accuracy, and integrating genomic and multi-environment data for forecasting the performance of genetic material in untested sites or under changing climate scenarios. The buildup of phenotypic and genome-wide data across large-scale breeding populations and advances in computational prediction of discrete genomic features should also provide opportunities to enhance the application of genomics to tree breeding.

Keywords: genomic selection (GS), tree breeding, quantitative genetics, whole-genome regression, single nucleotide polymorphisms (SNP), marker assisted selection (MAS), realized genomic relationship

INTRODUCTION

Forest tree breeding encompasses a number of steps to increase the frequency of advantageous alleles for several traits concurrently in a target population. Recurrent cycles of selection ultimately result in genetically improved planting material by maximizing genetic gain per unit time in the most cost-effective way (Namkoong et al., 1988; White et al., 2007). Long breeding cycles, late and poor flowering, weak juvenile-mature correlations and changes in climate, market demands and emerging pest and disease pressures, pose, however, daunting challenges. The advancement and ultimate output of tree breeding programs are therefore highly conditional on the length of a breeding cycle. To maximize genetic gains per unit time, extensive efforts in tree breeding were devoted to the two fundamental means by which the length of a breeding cycle can be decreased, namely, early selection and accelerated breeding. While the former is based on the understanding of juvenile-mature correlations and the practice of selection on juvenile traits (Williams, 1988), the latter involved early flower induction methods with hormone, stress treatments, and top grafting (Greenwood et al., 1991; Hasan and Reid, 1995). In the late 80s, the advent of DNA markers and two seminal papers on the dissection of discrete Mendelian factors underlying quantitative traits (Lander and Botstein, 1989), and marker-assisted selection (MAS) (Lande and Thompson, 1990), were seen as powerful tools to overcome the time challenge of tree breeding (El-Kassaby, 1982; Neale and Williams, 1991; Grattapaglia et al., 1992; Williams and Neale, 1992).

Here we cover the current state of the science on the general theme of optimizing and accelerating tree breeding using genomic technologies. A brief overview of the path across QTL (quantitative trait loci) and association mapping is first presented. It provides a quick historical perspective on how and why we reached the point of convergence between quantitative genetics and genomics. Furthermore, it also serves to substantiate the fact that reductionist "genetic dissection" approaches or attempts to use single candidate genes or diffuse, indirect information from transcriptomics, have not proven useful for breeding practice and are therefore not discussed further. We focus on the factors that affect and the challenges that remain to fully integrate genomic data in tree breeding in light of the recent promising results of whole-genome prediction. Although making predictions is difficult "especially about the future" as Niels Bohr and others once amusingly said, we attempt to look at the near future of tree breeding when genotyping, whole genome sequencing and computational prediction of genomic features for thousands of trees will not be limiting. We anticipate a future where the progressive advances made possible by routine genomic selection (GS) in multiple large populations will provide a more powerful platform to revisit the discovery of discrete genomic elements that may further enhance whole-genome phenotype prediction and eventually allow direct discrete interventions at the DNA sequence level.

THE PATH FROM GENETIC DISSECTION TO GENOMIC SELECTION

The prospects of MAS for forest trees was properly doubted early on, limiting its potential value to specific genetic backgrounds resulting from linkage equilibrium of forest tree populations, QTLs interacting with environments and changes of allele frequencies across generations (Strauss et al., 1992). Notwithstanding those sound advices, a number of QTL mapping experiments in the major conifers and eucalypts advanced, encouraged by the promising results of QTL mapping in inbred crops and model systems. In retrospect it is startling to consider how far removed from real-life tree breeding those bi-parental QTL mapping studies in forest trees were (Grattapaglia, 2017). The motivating hypothesis was that it would be possible to locate and estimate the effects of most individual QTLs underlying complex traits in every population and environment and implement them in tree breeding practice. A substantial number of studies reporting hundreds QTLs in forest trees was reported (reviewed in (Kirst et al., 2004; Grattapaglia et al., 2009; Neale and Kremer, 2011). Although several supposedly "major effect" QTLs were found in those early studies, those proved to be largely overestimated in effect size and underestimated in number. Indeed, subsequent multi-family experiments and larger sample sizes, revealed significantly larger numbers of QTLs with correspondingly smaller effects and inconsistent performance across environments and genetic backgrounds (Ukrainetz et al., 2008; Novaes et al., 2009; Thumma et al., 2010; Gion et al., 2011).

To solve the perceived shortcomings of QTL detection in single mapping families, association genetics was put forward as a way to provide population-wide marker–trait associations applicable to breeding (Neale and Savolainen, 2004). The limitation of methods to interrogate DNA polymorphisms at the time only allowed candidate-gene approaches (Thumma et al., 2005; Gonzalez-Martinez et al., 2007), which were then followed by genome-wide association mapping (GWAS) in several forest tree species (Beaulieu et al., 2011; Cumbie et al., 2011; Cappa et al., 2013; Porth et al., 2013; Mckown et al., 2014). However, irrespective of the marker density used, population size and improved analytical methods to account for low-frequency variants (Fahrenkrog et al., 2017; Müller et al., 2017, 2018; Resende et al., 2017a), only few polymorphisms of very modest effect have been detected, largely still lacking independent validation, the cornerstone for the scientific credibility of GWAS results. In effect, after 25 years of research efforts based on the principle and experimental approaches of genetic dissection of quantitative traits, no translation of such efforts to operational tree breeding was achieved (Grattapaglia et al., 2009; Grattapaglia, 2014; Isik, 2014).

The ineffectiveness of fully dissecting complex traits, and the limitations of MAS has not been exclusive to forest trees, but has also been recognized in crops (Bernardo, 2008) and domestic animals (Dekkers, 2004). This realization has caused a significant shift in the paradigm and technical approach to

plant and animal MAS. These fields have now moved from the *a priori* discovery of discrete marker-trait associations to the capture of the whole-genome effect assisted by DNA marker data, harmonizing with the multifactorial polygenic nature of quantitative genetics, as predicted by Fisher's infinitesimal model (Fisher, 1918). This shift was only possible following the development of improved and accessible genomic technologies that allow interrogating thousands of genome-wide single-nucleotide polymorphisms (SNPs) using cost effective platforms. The concept of using the "total allelic" (Nejati-Javaremi et al., 1997) or "total genomic" (Haley and Visscher, 1998) relationship from marker data to derive estimates of breeding values was later termed "Genomic Selection" (GS) in Meuwissen et al. (2001) seminal paper. It demonstrated that "*the selection on genetic values predicted from markers could substantially increase the rate of genetic gain per unit time in animals and plants, especially if combined with techniques to shorten the generation interval.*"

GS employs a genome-wide panel of markers, typically SNP (single nucleotide polymorphism), whose effects on the phenotype are estimated in a "training" population. In forest trees, such a training set is usually composed by sampling one to a few thousand individuals in progeny trials derived from mating a few dozen parents that constitute the target elite germplasm bred. SNPs are used to build prediction models to be later applied to "selection candidates" for which only genotypes are gathered and phenotypes are predicted by the genomic data. The prediction models are cross-validated against a "validation" population, a set of genetically related individuals to the training set but that did not participate in the estimation of marker effects. A prediction model that delivers a high correlation between the observed and predicted breeding values is subsequently used in the breeding phase to calculate the genomic estimated breeding or total genotypic values of the selection candidates (**Figure 1**). GS fundamentally exploits the genetic relationship between the training population and the prospective selection candidates and to a lesser extent the linkage disequilibrium (LD) between marker data and QTL effects. By precluding prior discrete marker selection derived from rigorous significance tests, and by estimating marker effects in a larger and breeding-representative population of trees, GS captures substantial proportions of the heritability contributed by the large numbers of genomic effects that QTL mapping or GWAS are, on principle, neither able nor intended to capture.

PERSPECTIVES OF GENOMIC SELECTION IN TREE BREEDING

GS can have a substantial impact on the rate of genetic gain. Let's recall Falconer's breeder's equation ($\Delta G = i r \sigma_A / L$) (Falconer, 1989), where i is the selection intensity; r is the accuracy of selection, or heritability in the original Falconer's expression, corresponding to the correlation between the estimated and true breeding values; σ_A is the additive-genetic standard deviation of the trait under inspection; and L is the generation interval. GS can

increase the rate of genetic gain of breeding cycle by increasing (*i*) because the phenotypes of a much larger number of seedlings in the nursery can be predicted with marker data compared to the number of trees that can be tested in conventional field progeny trials. Additionally, the use of realized genomic relationships is associated with increased accuracy in estimating σ_A and breeding values (*r*) (Hayes et al., 2009; El-Dien et al., 2018). Yet, in forest trees, the potentially greatest impact of GS on the rate of genetic progress will originate from decreasing (*L*). Phenotypes of the selection candidates can be predicted at very early ages, for example, when the seedlings are a few weeks old. GS not only could preclude or at least enhance the efficiency of progeny testing but would also optimize clonal testing phases by advancing a smaller number of pre-selected trees to be assessed in multi-site expanded clonal trials (Resende et al., 2012a) (**Figure 1**). In conifers, GS coupled to somatic embryogenesis for clonal propagation of elite genotypes could allow selecting elite zygotic embryos based on their genomic value saving a significant amount of time, and avoiding the costs and uncertainties currently involved in cryopreservation rescue (Resende et al., 2012b). Additionally, GS will allow simultaneous and early selection for multiple trait in large numbers of individuals, an impossible task in conventional tree breeding that currently largely adopts tandem selection. The final impact of GS would therefore be a significant improvement in the general efficiency of a tree breeding program, provided, of course, that genotyping is inexpensive and GS models are accurate.

What makes GS distinctive from what tree breeders have done so far is that instead of relying uniquely on the expected pedigree, frequently prone to errors, DNA data allows one to build additive and non-additive genomic relationship matrices that more accurately specify the relationships among individuals and simultaneously account for contemporary as well as historical pedigree. This procedure not only allows rectifying pedigree inaccuracies, but critically it captures the within family variation resulting from random Mendelian segregation term. Accordingly, the realized genetic covariances are now based on the actual fraction of the genome that is identical by descent or by state between individuals (Vanraden, 2008). It has been shown in a number of studies in forest trees that realized genomic relationships can produce more accurate predictions than pedigrees alone (Munoz et al., 2014; El-Dien et al., 2015, 2018; Bouvet et al., 2016; Cappa et al., 2017, 2018; Tan et al., 2018). Additionally, the realized genomic relationships of a small subset of the progeny testing population have been effectively combined with a substantially large proportion of un-genotyped individuals in a single-step analysis (Legarra et al., 2009). This method was dubbed "HBLUP," since the best linear unbiased predictors (BLUPs) of breeding values are derived using a single (**H**) genetic covariance matrix that combines the pedigree-based average numerator relationship matrix (**A**) with the marker-based relationship matrix (**G**). HBLUP increases the precision of the genetic parameters generated from traditional pedigrees as shown in recent studies with forest trees (Cappa et al., 2017, 2018; Ratcliffe et al., 2017).
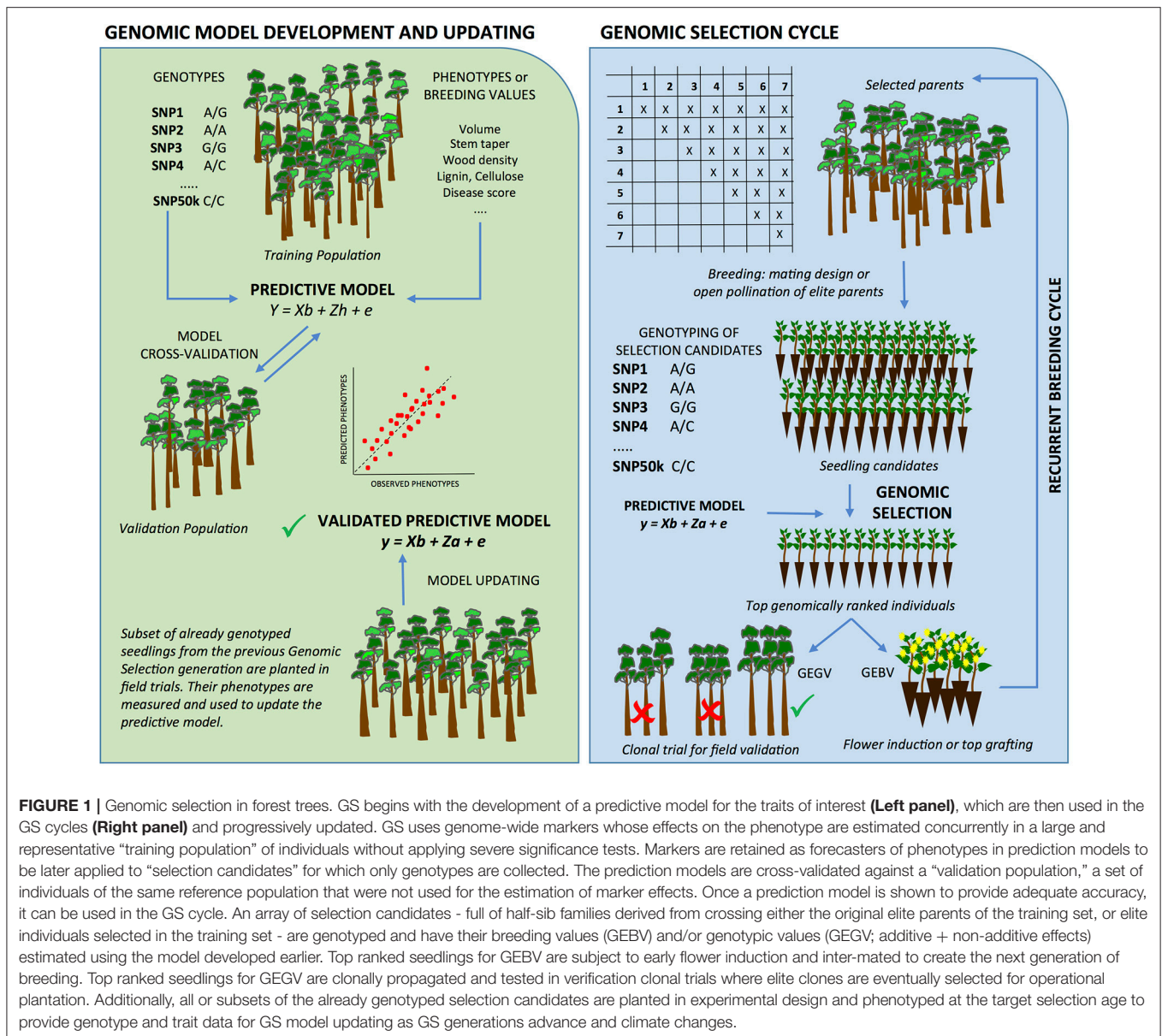
FIGURE 1 | Genomic selection in forest trees. GS begins with the development of a predictive model for the traits of interest **(Left panel)**, which are then used in the GS cycles **(Right panel)** and progressively updated. GS uses genome-wide markers whose effects on the phenotype are estimated concurrently in a large and representative "training population" of individuals without applying severe significance tests. Markers are retained as forecasters of phenotypes in prediction models to be later applied to "selection candidates" for which only genotypes are collected. The prediction models are cross-validated against a "validation population," a set of individuals of the same reference population that were not used for the estimation of marker effects. Once a prediction model is shown to provide adequate accuracy, it can be used in the GS cycle. An array of selection candidates - full of half-sib families derived from crossing either the original elite parents of the training set, or elite individuals selected in the training set - are genotyped and have their breeding values (GEBV) and/or genotypic values (GEGV; additive + non-additive effects) estimated using the model developed earlier. Top ranked seedlings for GEBV are subject to early flower induction and inter-mated to create the next generation of breeding. Top ranked seedlings for GEGV are clonally propagated and tested in verification clonal trials where elite clones are eventually selected for operational plantation. Additionally, all or subsets of the already genotyped selection candidates are planted in experimental design and phenotyped at the target selection age to provide genotype and trait data for GS model updating as GS generations advance and climate changes.

GS: ADVANCES AND CHALLENGES IN FOREST TREES

A comprehensive and time-lined list of empirical GS reports in forest tree species was recently published (Grattapaglia, 2017) and it is now updated in **Table 1**. Prediction accuracies have been largely very good, matching or surpassing those obtainable by pedigree-based phenotypic selection, in line with former simulations (Grattapaglia and Resende, 2011; Iwata et al., 2011; Denis and Bouvet, 2013). When considering the practicalities of tree breeding, however, a number of factors that affect the prospects of GS have to be considered, including the composition of training populations, analytical methods, genotype x environment interaction (G*E), age-age correlations, the long term models performance and cost and quality of DNA

marker data. All these have been the subject of research and reviewed in detail in the context of tree breeding by Grattapaglia (2017) and are briefly discussed below in light of the experimental results reported to date in forest trees.

GS experiments in forest trees have capitalized on the existing structure and diversity of breeding populations and their designs that account for the expected relationship between training and prospective selection candidates. Training populations of several hundred to a few thousand individuals sampled from existing progeny trials with effective population sizes consistent with those used in operational breeding have provided good predictions in essentially all studies and for all traits. Analytical methods differing with respect to the presumed trait architecture have been used and compared. In all studies, the ridge regression best linear unbiased prediction (RR-BLUP), with

TABLE 1 | Timeline summary of experimental genomic selection studies in forest tree species published to date.

Forest tree	DNA marker data	Traits analyzed	References
Eucalypts (*Eucalyptus grandis x E. urophylla, E. camaldulensis* - hybrids)	3,129 DArT array	Growth and wood quality	Grattapaglia et al., 2011 Resende et al., 2012a
Loblolly pine (*Pinus taeda*)	4,852 SNP chip	Growth, wood and disease	Resende et al., 2012b Resende et al., 2012c Munoz et al., 2014 De Almeida et al., 2016
Loblolly pine (*Pinus taeda*)	3,461 SNP chip	Growth and wood quality	Zapata-Valenzuela et al., 2012 Zapata-Valenzuela et al., 2013
Eucalypts (*Eucalyptus grandis x E. urophylla* - hybrids)	29,090 SNP EuCHIP60K	Growth and wood quality	Lima, 2014
White spruce (*Picea glauca*)	6,932 SNP chip	Growth and wood quality	Beaulieu et al., 2014a Beaulieu et al., 2014b Ratcliffe et al., 2017
Interior spruce (*Picea glauca x P. engelmannii*)	8,868 to 62,198 GbS SNPs	Growth and wood quality	El-Dien et al., 2015 Ratcliffe et al., 2015 El-Dien et al., 2018
Maritime pine (*Pinus pinaster*)	4,332 SNP chip	Growth and form	Isik et al., 2016 Bartholome et al., 2016
Eucalypts (*Eucalyptus grandis x E. urophylla* - hybrids)	3,303 DArTseq SNPs	Growth	Bouvet et al., 2016
Eucalypt (*Eucalyptus pellita* and *E. benthamii*)	19,506 SNP EuCHIP60K	Growth	Müller et al., 2017
Eucalypt (*Eucalyptus grandis x E. urophylla* - hybrids)	24,806 SNP EuCHIP60K	Growth and wood quality	Resende et al., 2017b
Eucalypt (*Eucalyptus globulus*)	12,000 SNP EuCHIP60K	Growth and wood quality	Duran et al., 2017
Black spruce (*Picea mariana*)	4,993 SNP chip	Growth and wood quality	Lenz et al., 2017
Eucalypt (*Eucalyptus grandis*)	2,816 DArT array	Growth	Cappa et al., 2017 Cappa et al., 2018
Eucalypt (*Eucalyptus grandis x E. urophylla* - hybrids)	41,304 SNP EuCHIP60K	Growth and wood quality	Tan et al., 2017 Tan et al., 2018
Douglas fir (*Pseudotsuga menziesii*)	69,551 exome capture SNPs	Growth and wood quality	Thistlethwaite et al., 2017
Norway spruce (*Picea abies*)	116,765 exome capture SNPs	Growth and wood quality	Chen et al., 2018
Eucalypt (*Eucalyptus nitens*)	12,236 SNP EuCHIP60K	Growth and wood quality	Suontama et al., 2018
Eucalypt (*Eucalyptus polybractea*)	Shallow whole genome sequencing; up to 500,000 SNPs	Foliar terpene yield traits	Kainer et al., 2018
Eucalypt (*Eucalyptus grandis x E. urophylla* - hybrids)	40,932 SNP EuCHIP60K and 55,772 capture probe SNP	Growth and wood quality	De Moraes et al., 2018

Studies that investigated different aspects of GS but used partially or totally the same breeding population data (genotypes and/or phenotypes) are listed in the same entry.

marker effects treated as random, normally distributed with common variance, has been very efficient. RR-BLUP has been equivalent to Genomic BLUP (GBLUP), providing the best conciliation between prediction efficiency and fast computation, while also revealing that essentially all major traits in forest trees fit the infinitesimal model (Resende et al., 2012c; Beaulieu et al., 2014b; Ratcliffe et al., 2015; Isik et al., 2016; Müller et al., 2017; Tan et al., 2017; Chen et al., 2018). Still, additional research in this area is warranted especially as prior functional information on genomic regions of slightly larger effect might emerge, for example, for disease resistance traits as shown for prediction of fusiform rust resistance in loblolly pine (Resende et al., 2012c).

Ever since the first experimental GS studies in forest trees (Grattapaglia et al., 2011; Resende et al., 2012a,b), it became clear that prediction accuracies are mainly driven by genetic

relationship between training and validation sets and are dependent on G*E and age-age correlations. Predictions will be most effective at the same age and in the same environment where the prediction model was trained. Further studies in conifers (Zapata-Valenzuela et al., 2013; Beaulieu et al., 2014a,b; El-Dien et al., 2015; Ratcliffe et al., 2015; Thistlethwaite et al., 2017; Chen et al., 2018), and eucalypts (Müller et al., 2017; Tan et al., 2017), corroborated the key significance of genetic relationships and the impact of G*E and age-age correlations, consistent with findings in domestic animals and crop plants (Lin et al., 2014; Van Eenennaam et al., 2014). While data from G*E or age-age correlation studies will shed light on what to expect from genomic prediction, assuring that the target environment of future selection candidates will be equivalent to the one where models were originally trained is a challenging

issue for GS (Heslot et al., 2015). Regular retraining of GS models by incorporating phenotypes collected in breeding generations closer to the current (Iwata et al., 2011) are expected to mitigate this problem, and will be especially essential in light of climate fluctuations. Research efforts in this area are highly needed and will come as GS programs advance, coupled to innovations in phenotyping platforms that integrate remote sensing, spatial and geographic information systems (Dungey et al., 2018).

Notwithstanding the encouraging estimates of predictive ability, most studies in forest trees used contemporary training and validation sets and thus have not yet been able to adequately assess the realized performance of GS across generations at a larger scale, but results on this topic are imminent. However, given that the relationship between parents and progeny are accurately captured by DNA marker data, and environments should be relatively stable across close generations, it is expected that the performance will be equivalent to current estimates in contemporary sets. In *Pinus pinaster*, preliminary promising results of inter-generation prediction were reported by training models with parents and progeny in the same set (Isik et al., 2016), and later using parents and grandparents to predict in the subsequent generation, albeit with limited effective population sizes (Bartholome et al., 2016). However, this outcome was not observed in a three-generation study of *Pseudotsuga menziesii* (El-Kassaby, personal communication). Model updating strategies will therefore be crucial to counteract the decay of relatedness and LD between the original training set and selection candidates as generations of breeding advance, as shown by simulations for eucalypt breeding (Denis and Bouvet, 2013).

In the past 2 years, a number of additional experimental GS studies have been reported (Cappa et al., 2017, 2018; Duran et al., 2017; Lenz et al., 2017; Müller et al., 2017; Ratcliffe et al., 2017; Tan et al., 2017, 2018; Thistlethwaite et al., 2017) (**Table 1**; Resende et al., 2017b; Chen et al., 2018; De Moraes et al., 2018; El-Dien et al., 2018; Kainer et al., 2018; Suontama et al., 2018). Many of them in species of *Eucalyptus* for which public high-throughput genotyping platforms of DArT (Sansaloni et al., 2010) and SNPs (Silva-Junior et al., 2015) have been available. Access to such resources for eucalypts also allowed improved precision of genetic parameter estimates, pedigree reconstruction and inbreeding studies (Telfer et al., 2015; Klápště et al., 2017; Müller et al., 2017). This clearly points to the fact that the advancement of research and operational adoption of genomics into breeding is strongly dependent on the availability of public, robust, cost-accessible and portable SNP genotyping platforms. The success of GS or any other genomic-based breeding approach will rely on high data quality, as one has to be able to genotype SNPs across generations with high reproducibility and negligible missing data. Although shallow whole genome sequencing (Kainer et al., 2018), genotyping-by-sequencing (GbS) (El-Dien et al., 2015) and sequence capture (Thistlethwaite et al., 2017; Chen et al., 2018; De Moraes et al., 2018) have also been used for GS in trees, currently fixed SNP arrays provide the gold standard of data reproducibility across samples batches and laboratories. Additionally, SNP array data are breeder friendly, available from multiple service providers, easily manageable and

stored without the cost and logistics of sequence data transfer, storage and analysis. This and a significant recent drop in array costs, making them as cost-effective as sequence-based methods, has motivated a large international effort to develop SNP arrays for all main planted conifers (F. Isik pers. comm.), and a second generation, higher density optimized SNP array for species of *Eucalyptus* and *Corymbia* (O.B. Silva-Junior and D. Grattapaglia pers. comm.). The use of a common SNP genotyping array across breeding programs of different organizations will be a key issue to provide the necessary economy of scale to integrate genomics into breeding.

A LOOK TO THE NEAR FUTURE

With easy access to SNP genotyping and positive results in essentially all major forest trees, we are now at the brink of widespread adoption of genomic prediction data, thus realizing the early promises of MAS in forest tree breeding. In addition to the outstanding research challenges discussed above, a promising area to enhance the value of genomic data will involve the inclusion of environmental co-variables in GS models as already shown in crops (Jarquin et al., 2014; Saint Pierre et al., 2016). The integration of multi-environment trials data will be strategic for predicting performance in unobserved environments, identifying suitable sites for evaluating or deploying genetic material and predicting climate change scenarios. While predicting the performance of untested clones or families can be accurate when there is knowledge of genomic relatedness, correspondingly, the performance in yet unobserved or future environments could be forecasted if there is data about those environments as shown for recommendation of *Eucalyptus* clones (Marcatti et al., 2017). Resources such as ClimateNA (Wang et al., 2016) and the NASA POWER project (Stackhouse, 2014) offer multitudes of historical and predicted future environmental data. Because environmental variables that define the correlation between growing conditions are trait specific, research on those most appropriate for inclusion in genomic prediction models will be essential.

Another area that will demand research comes from the evolution of sequencing technologies in moving from sparse SNP data to sequence data for GS. Apart from the challenge and cost of managing massive next generation sequencing data sets for large numbers of individuals in a breeding program framework, in theory, if sequence data were used instead of dense SNPs, accuracy should increase because rare causal alleles would be better captured in predictive models. Until now, however, simulation and experimental studies in domestic animals have shown that whole-genome sequence data does not increase accuracy when LD has a slow decay pattern (Macleod et al., 2014; Forneris et al., 2017; Vanraden et al., 2017), unless very precise prior estimates on the functionality of particular SNPs exist (Perez-Enciso et al., 2015). Increasing the availability and quality of functional data on specific genomic regions might therefore, be warranted.

The success of whole-genome prediction and the poor outcome of dissection approaches in identifying functional

quantitative trait nucleotides, have contributed nevertheless to an exciting new perspective for the study of complex trait variation. A clear pattern has emerged in annual plants indicating that the association signal of common variants in large sample sizes, although spread across the entire genome, is heavily concentrated in regulatory DNA in open chromatin marked by deoxyribonuclease hypersensitive sites (Sullivan et al., 2014; Rodgers-Melnick et al., 2016; Swinnen et al., 2016). In these plants, cis-regulatory elements (CREs) associated with open chromatin such as promoters and enhancers regulating gene expression may contain close to half of all variants influencing traits. As GS implementation advances and large datasets of several thousand trees across unrelated populations are collected, opportunities will emerge for joint and meta-GWAS, as recently described in *Eucalyptus* (Müller et al., 2018). At the same time, chromatin accessibility and gene network data as reported for *Eucalyptus* (Hussey et al., 2017) and *Populus* (Zinkgraf et al., 2017) will become increasingly available which, combined with data from highly powered SNP-trait association studies, should provide new avenues for computational predictive discovery of key regulatory elements in the genome. The progress of

such integrative approaches based on large genotype and phenotype datasets might, thus, result in additional clues toward understanding the complex connections and interactions between discrete genomic elements and continuous phenotypic trait variation, ultimately enhancing tree breeding practice.

AUTHOR CONTRIBUTIONS

DG drafted the first version of the manuscript and all co-authors subsequently contributed to it by editing and formatting the final version.

ACKNOWLEDGMENTS

REFERENCES

Bartholome, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., et al. (2016). Performance of genomic prediction within and across generations in maritime pine. *BMC Genomics* 17:604. doi: 10.1186/s12864-016-2879-8

Beaulieu, J., Doerksen, T., Boyle, B., Clement, S., Deslauriers, M., Beauseigle, S., et al. (2011). Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. *Genetics* 188, 1197–1214. doi: 10.1534/genetics.110.125781

Beaulieu, J., Doerksen, T., Clement, S., Mackay, J., and Bousquet, J. (2014a). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* 113, 343–352. doi: 10.1038/hdy.2014.36

Beaulieu, J., Doerksen, T. K., Mackay, J., Rainville, A., and Bousquet, J. (2014b). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics* 15:1048. doi: 10.1186/1471-2164-15-1048

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131

Bouvet, J. M., Makouanzi, G., Cros, D., and Vigneron, P. (2016). Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. *Heredity* 116, 146–157. doi: 10.1038/hdy.2015.78

Cappa, E. P., El-Kassaby, Y. A., Garcia, M. N., Acuna, C., Borralho, N. M. G., Grattapaglia, D., et al. (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:e81267. doi: 10.1371/journal.pone.0081267

Cappa, E. P., El-Kassaby, Y. A., Muñoz, F., Garcia, M. N., Villalba, P. V., Klápšt,ě, J., et al. (2017). Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models. *Mol. Breed.* 37:125. doi: 10.1007/s11032-017-0725-6

Cappa, E. P., El-Kassaby, Y. A., Muñoz, F., Garcia, M. N., Villalba, P. V., Klápšt,ě, J., et al. (2018). Genomic-based multiple-trait evaluation in *Eucalyptus grandis* using dominant DArT markers. *Plant Sci.* 271, 27–33. doi: 10.1016/j.plantsci.2018.03.014

Chen, Z. Q., Baison, J., Pan, J., Karlsson, B., Gull, B. A., Westin, J., et al. (2018). Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as

genotyping platform in Norway spruce. *bioRxiv [Preprint]*. doi: 10.1101/293696

Cumbie, W. P., Eckert, A., Wegrzyn, J., Whetten, R., Neale, D., and Goldfarb, B. (2011). Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of Pinus taeda L. *Heredity* 107, 105–114. doi: 10.1038/hdy.2010.168

De Almeida, J. E. D., Guimaraes, J. F. R., Silva, F. F. E., De Resende, M. D. V., Munoz, P., Kirst, M., et al. (2016). The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity* 117, 33–41. doi: 10.1038/hdy.2016.23

De Moraes, B. F. X., Dos Santos, R. F., De Lima, B. M., Aguiar, A. M., Missiaggia, A. A., Da Costa Dias, D., et al. (2018). Genomic selection prediction models comparing sequence capture and SNP array genotyping methods. *Mol. Breed.* 38:115. doi: 10.1007/s11032-018-0865-3

Dekkers, J. C. M. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* 82, E313–E328. doi: 10.2527/2004.8213_supplE313x

Denis, M., and Bouvet, J. M. (2013). Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genet. Genomes* 9, 37–51. doi: 10.1007/s11295-012-0528-1

Dungey, H. S., Dash, J. P., Pont, D., Clinton, P. W., Watt, M. S., and Telfer, E. J. (2018). Phenotyping whole forests will help to track genetic performance. *Trends Plant Sci.* 23, 854–864. doi: 10.1016/j.tplants.2018.08.005

Duran, R., Isik, F., Zapata-Valenzuela, J., Balocchi, C., and Valenzuela, S. (2017). Genomic predictions of breeding values in a cloned Eucalyptus globulus population in Chile. *Tree Genet. Genomes* 13:74. doi: 10.1007/s11295-017-1158-4

El-Dien, O. G., Ratcliffe, B., Klapste, J., Chen, C., Porth, I., and El-Kassaby, Y. A. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genomics* 16:370. doi: 10.1186/s12864-015-1597-y

El-Dien, O. G., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., and El-Kassaby, Y. A. (2018). Multienvironment genomic variance decomposition analysis of open-pollinated Interior spruce (*Picea glauca* x engelmannii). *Mol. Breed.* 38:26. doi: 10.1007/s11032-018-0784-3

El-Kassaby, Y. A. (1982). Associations between Allozyme genotypes and quantitative traits in Douglas-Fir [*Pseudotsuga Menziesii* (Mirb.) Franco]. *Genetics* 101, 103–115.

Fahrenkrog, A. M., G., N.L., R., R.M.F., I., V. A., Gustavo, C., Christopher, D., Robert, S., Mark, D., et al. and Matias, K. (2017). Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *N. Phytol.* 213, 799–811. doi: 10.1111/nph.14154

Falconer, D. S. (1989). *Introduction to Quantitative Genetics*. Essex: Longman Scientific and Technical.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433. doi: 10.1017/S0080456800012163

Forneris, N. S., Vitezica, Z. G., Legarra, A., and Pérez-Enciso, M. (2017). Influence of epistasis on response to genomic selection using complete sequence data. *Genet. Select. Evol.* 49:66. doi: 10.1186/s12711-017-0340-3

Gion, J. M., Carouche, A., Deweer, S., Bedon, F., Pichavant, F., Charpentier, J. P., et al. (2011). Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: eucalyptus. *Bmc Genomics* 12:301. doi: 10.1186/1471-2164-12-301

Gonzalez-Martinez, S. C., Wheeler, N. C., Ersoz, E., Nelson, C. D., and Neale, D. B. (2007). Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* 175, 399–409. doi: 10.1534/genetics.106.061127

Grattapaglia, D. (2014). "Breeding forest trees by genomic selection: current progress and the way forward," in *Advances in Genomics of Plant Genetic Resources,* eds R. Tuberosa, A. Graner, and E. Frison (New York, NY: Springer), 652–682. doi: 10.1007/978-94-007-7572-5_26

Grattapaglia, D. (2017). "Status and perspectives of genomic selection in forest tree breeding," in *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*, eds R. K. Varshney, M. Roorkiwal, and M. E. Sorrells (Cham: Springer International Publishing), 199–249. doi: 10.1007/978-3-319-63170-7_9

Grattapaglia, D., Chaparro, J., Wilcox, P., Mccord, S., Werner, D., Amerson, H., et al. (1992). "Mapping in woody plants with RAPD markers: applications to breeding in forestry and horticulture," in *Proceedings of the Symposium Applications of RAPD Technology to Plant Breeding* (Minneapolis, MN: Crop Science Society of America, American Society of Horticultural Science, American Genetic Association), 37–40.

Grattapaglia, D., Plomion, C., Kirst, M., and Sederoff, R. R. (2009). Genomics of growth traits in forest trees. *Curr. Opin. Plant Biol.* 12, 148–156. doi: 10.1016/j.pbi.2008.12.008

Grattapaglia, D., and Resende, M. D. V. (2011). Genomic selection in forest tree breeding. *Tree Genet. Genomes* 7, 241–255. doi: 10.1007/s11295-010-0328-4

Grattapaglia, D., Resende, M. D. V., Resende, M., Sansaloni, C., Petroli, C., Missiaggia, A., et al. (2011). Genomic selection for growth traits in Eucalyptus: accuracy within and across breeding populations. *BMC Proc.* 5:O16. doi: 10.1186/1753-6561-5-S7-O16

Greenwood, M. S., Adams, G. W., and Gillespie, M. (1991). Stimulation of flowering by grafted black spruce and white spruce - a comparative-study of the effects of Gibberellin A4/7, cultural treatments, and environment. *Can. J. Forest Res. Revue Can. Rech. Forest.* 21, 395–400. doi: 10.1139/x91-049

Haley, C. S., and Visscher, P. M. (1998). Strategies to utilize marker-quantitative trait loci associations. *J. Dairy Sci.* 81, 85–97. doi: 10.3168/jds.S0022-0302(98)70157-2

Hasan, O., and Reid, J. B. (1995). Reduction of generation time in *Eucalyptus globulus. Plant Growth Regul.* 17, 53–60.

Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res (Camb).* 91, 47–60. doi: 10.1017/S0016672308009981

Heslot, N., Jannink, J. L., and Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Sci.* 55, 1–12. doi: 10.2135/cropsci2014.03.0249

Hussey, S. G., Loots, M. T., Van Der Merwe, K., Mizrachi, E., and Myburg, A. A. (2017). Integrated analysis and transcript abundance modelling of H3K4me3 and H3K27me3 in developing secondary xylem. *Sci. Rep.* 7:3370. doi: 10.1038/s41598-017-03665-1

Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *N. Forests* 45, 379–401. doi: 10.1007/s11056-014-9422-z

Isik, F., Bartholome, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., et al. (2016). Genomic selection in maritime pine. *Plant Sci.* 242, 108–119. doi: 10.1016/j.plantsci.2015.08.006

Iwata, H., Hayashi, T., and Tsumura, Y. (2011). Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica. Tree Genet. Genomes* 7, 747–758. doi: 10.1007/s11295-011-0371-9

Jarquin, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127, 595–607. doi: 10.1007/s00122-013-2243-1

Kainer, D., Stone, E. A., Padovan, A., Foley, W. J., and Külheim, C. (2018). accuracy of genomic prediction for foliar terpene traits in *Eucalyptus polybractea* G3 8, 2573–2583. doi: 10.1534/g3.118.200443

Kirst, M., Myburg, A., and Sederoff, R. (2004). Genetic mapping in forest trees: markers, linkage analysis and genomics. *Genet. Eng.* 26, 105–141. doi: 10.1007/978-0-306-48573-2_7

Klápště, J., Suontama, M., Telfer, E., Graham, N., Low, C., Stovold, T., et al. (2017). Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of Eucalyptus nitens. *PLoS ONE* 12:e0185137. doi: 10.1371/journal.pone.0185137

Lande, R., and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756.

Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi: 10.3168/jds.2009-2061

Lenz, P. R. N., Beaulieu, J., Mansfield, S. D., Clement, S., Desponts, M., and Bousquet, J. (2017). Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genomics* 18:335. doi: 10.1186/s12864-017-3715-5

Lima, B. M. (2014). *Bridging Genomics and Quantitative Genetics of Eucalyptus: Genome-Wide Prediction and Genetic Parameter Estimation for Growth and Wood Properties Using High-Density SNP Data*. Dissertation, Piracicaba. University of São Paulo. Available online at http://www.teses.usp.br/teses/disponiveis/11/11137/tde-25062014-085814/pt-br.php

Lin, Z., Hayes, B. J., and Daetwyler, H. D. (2014). Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* 65, 1177–1191. doi: 10.1071/CP13363

Macleod, I. M., Hayes, B. J., and Goddard, M. E. (2014). The effects of demography and long-term selection on the accuracy of genomic prediction with sequence data. *Genetics* 198, 1671–1684. doi: 10.1534/genetics.114.168344

Marcatti, G. E., Resende, R. T., Resende, M. D. V., Ribeiro, C. A. A. S., Dos Santos, A. R., Cruz, J. P., et al. (2017). GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. *For. Ecol. Manage.* 392, 144–153. doi: 10.1016/j.foreco.2017.03.006

Mckown, A. D., Klapste, J., Guy, R. D., Geraldes, A., Porth, I., Hannemann, J., et al. (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa. N. Phytol.* 203, 535–553. doi: 10.1111/nph.12815

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Müller, B. S. F., De Almeida Filho, J. E., Lima, B. M., Garcia, C. C., Missiaggia, A., Aguiar, A. M., et al. (2018). Independent and Joint-GWAS for growth traits in *Eucalyptus* by assembling genome-wide data for 3373 individuals across four breeding populations. *N. Phytol.* doi: 10.1111/nph.15449. [Epub ahead of print].

Müller, B. S. F., Neves, L. G., Almeida-Filho, J. E., Resende, M. F. R. J., Munoz Del Valle, P., Santos, P. E. T., et al. (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of Eucalyptus. *BMC Genomics* 18:524. doi: 10.1186/s12864-017-3920-2

Munoz, P. R., Resende, M. F. R., Gezan, S. A., Resende, M. D. V., De Los Campos, G., Kirst, M., et al. (2014). Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics* 198, 1759–1768. doi: 10.1534/genetics.114.171322

Namkoong, G., Kang, H. C., and Brouard, J. S. (1988). *Tree Breeding: Principles and Strategies*. New York, NY: Springer Verlag. doi: 10.1007/978-1-4612-3892-8

Neale, D. B., and Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nat. Revi. Genet.* 12, 111–122. doi: 10.1038/nrg2931

Neale, D. B., and Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends Plant Sci.* 9, 325–330. doi: 10.1016/j.tplants.2004.05.006

Neale, D. B., and Williams, C. G. (1991). Restriction-Fragment-Length-Polymorphism mapping in conifers and applications to forest genetics and tree improvement. *Can. J. Forest Res. Revue Can. Rech. Forest.* 21, 545–554. doi: 10.1139/x91-076

Nejati-Javaremi, A., Smith, C., and Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.* 75, 1738–1745. doi: 10.2527/1997.7571738x

Novaes, E., Osorio, L., Drost, D. R., Miles, B. L., Boaventura-Novaes, C. R. D., Benedict, C., et al. (2009). Quantitative genetic analysis of biomass and wood chemistry of Populus under different nitrogen levels. *N. Phytol.* 182, 878–890. doi: 10.1111/j.1469-8137.2009.02785.x

Perez-Enciso, M., Rincon, J. C., and Legarra, A. (2015). Sequence- vs. chip-assisted genomic selection: accurate biological information is advised. *Genet. Select. Evol.* 47:43. doi: 10.1186/s12711-015-0117-5

Porth, I., Klapste, J., Skyba, O., Hannemann, J., Mckown, A. D., Guy, R. D., et al. (2013). Genome-wide association mapping for wood characteristics in Populus identifies an array of candidate single nucleotide polymorphisms. *N. Phytol.* 200, 710–726. doi: 10.1111/nph.12422

Ratcliffe, B., El-Dien, O. G., Cappa, E. P., Porth, I., Klápště, J., Chen, C., et al. (2017). Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca. G3* 7, 935–942. doi: 10.1534/g3.116.037895

Ratcliffe, B., El-Dien, O. G., Klapste, J., Porth, I., Chen, C., Jaquish, B., et al. (2015). A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* x glauca) using unordered SNP imputation methods. *Heredity* 115, 547–555. doi: 10.1038/hdy.2015.57

Resende, M. D. V., Resende, M. F. R., Sansaloni, C. P., Petroli, C. D., Missiaggia, A. A., Aguiar, A. M., et al. (2012a). Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *N. Phytol.* 194, 116–128. doi: 10.1111/j.1469-8137.2011.04038.x

Resende, M. F. R., Munoz, P., Acosta, J. J., Peter, G. F., Davis, J. M., Grattapaglia, D., et al. (2012b). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *N. Phytol.* 193, 617–624. doi: 10.1111/j.1469-8137.2011.03895.x

Resende, M. F. R., Munoz, P., Resende, M. D. V., Garrick, D. J., Fernando, R. L., Davis, J. M., et al. (2012c). Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190, 1503–1510. doi: 10.1534/genetics.111.137026

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva, O. B., et al. (2017b). Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. *Heredity* 119, 245–255. doi: 10.1038/hdy.2017.37

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, O. B., et al. (2017a). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. *N. Phytol.* 213, 1287–1300. doi: 10.1111/nph.14266

Rodgers-Melnick, E., Vera, D. L., Bass, H. W., and Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3177–E3184. doi: 10.1073/pnas.1525244113

Saint Pierre, C., Burgueño, J., Crossa, J., Fuentes Dávila, G., Figueroa López, P., Solís Moya, E., et al. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Sci. Rep.* 6:27312. doi: 10.1038/srep27312

Sansaloni, C. P., Petroli, C. D., Carling, J., Hudson, C. J., Steane, D. A., Myburg, A. A., et al. (2010). A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus. Plant Methods* 6:16. doi: 10.1186/1746-4811-6-16

Silva-Junior, O. B., Faria, D. A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing 240 *Eucalyptus* tree genomes across 12 species. *N. Phytol.* 206, 1527–1540. doi: 10.1111/nph.13322

Stackhouse, P. W. (2014). "Prediction of worldwide energy resource," ed N. L. R. Ctr. (Hampton, VA). Available online at: http://power.larc.nasa.gov.

Strauss, S. H., Lande, R., and Namkoong, G. (1992). Limitations of molecular-marker-aided selection in forest tree breeding. *Can. J. Forest Res. Revue Can. Rech. Forest.* 22, 1050–1061. doi: 10.1139/x92-140

Sullivan, A. M., Arsovski, A. A., Lempe, J., Bubb, K. L., Weirauch, M. T., Sabo, P. J., et al. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana. Cell Rep.* 8, 2015–2030. doi: 10.1016/j.celrep.2014.08.019

Suontama, M., Klápště, J., Telfer, E. J., Graham, N., Stovold, T., Low, C., et al. (2018). Efficiency of genomic prediction across two Eucalyptus nitens seed orchards with different selection histories. *Heredity*. doi: 10.1038/s41437-018-0119-5. [Epub ahead of print].

Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from domestication: targeting *Cis*-regulatory elements for crop improvement. *Trends Plant Sci.* 21, 506–515. doi: 10.1016/j.tplants.2016.01.014

Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., and Ingvarsson, P. K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Genomics* 17:110. doi: 10.1186/s12870-017-1059-6

Tan, B., Grattapaglia, D., Wu, H. X., and Ingvarsson, P. K. (2018). Genomic relationships reveal significant dominance effects for growth in hybrid Eucalyptus. *Plant Sci.* 267, 84–93. doi: 10.1016/j.plantsci.2017.11.011

Telfer, E. J., Stovold, G. T., Li, Y., Silva-Junior, O. B., Grattapaglia, D. G., and Dungey, H. S. (2015). Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: a comparative analysis of marker data power and robustness. *PLoS ONE* 10:e0130601. doi: 10.1371/journal.pone.0130601

Thistlethwaite, F. R., Ratcliffe, B., Klapste, J., Porth, I., Chen, C., Stoehr, M. U., et al. (2017). Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genomics* 18:930. doi: 10.1186/s12864-017-4258-5

Thumma, B. R., Nolan, M. R., Evans, R., and Moran, G. F. (2005). Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* 171, 1257–1265. doi: 10.1534/genetics.105.042028

Thumma, B. R., Southerton, S. G., Bell, J. C., Owen, J. V., Henery, M. L., and Moran, G. F. (2010). Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens. Tree Genet. Genomes* 6, 305–317. doi: 10.1007/s11295-009-0250-9

Ukrainetz, N., Ritland, K., and Mansfield, S. (2008). Identification of quantitative trait loci for wood quality and growth across eight full-sib coastal Douglas-fir families. *Tree Genet. Genomes* 4, 159–170. doi: 10.1007/s11295-007-0097-x

Van Eenennaam, A. L., Weigel, K. A., Young, A. E., Cleveland, M. A., and Dekkers, J. C. M. (2014). Applied animal genomics: results from the field. *Ann. Rev. Anim. Biosci.* 2, 105–139. doi: 10.1146/annurev-animal-022513-114119

Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Vanraden, P. M., Tooker, M. E., O'connell, J. R., Cole, J. B., and Bickhart, D. M. (2017). Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet. Select. Evol.* 49, 32. doi: 10.1186/s12711-017-0307-4

Wang, T., Hamann, A., Spittlehouse, D., and Carroll, C. (2016). Locally downscaled and spatially customizable climate data for historical and future periods for North America. *PLoS ONE* 11:e0156720. doi: 10.1371/journal.pone.0156720

White, T. L., Adams, W. T., and Neale, D. B. (2007). *Forest Genetics*. Cambridge, MA: CABI Publishing. p. 682 doi: 10.1079/9781845932855.0000

Williams, C. G. (1988). Accelerated short-term genetic testing for loblolly-pine families. *Can. J. Forest Res. Revue Can. Rech. Forest.* 18, 1085–1089. doi: 10.1139/x88-165

Williams, C. G., and Neale, D. B. (1992). Conifer wood quality and marker-aided selection - a case-study. *Can. J. Forest Res. Revue Can. Rech. Forest.* 22, 1009–1017. doi: 10.1139/x92-135

Zapata-Valenzuela, J., Isik, F., Maltecca, C., Wegrzyn, J., Neale, D., Mckeand, S., et al. (2012). SNP markers trace familial linkages in a cloned population of *Pinus*

taeda - prospects for genomic selection. *Tree Genet. Genomes* 6, 1307–1318. doi: 10.1007/s11295-012-0516-5

Zapata-Valenzuela, J., Whetten, R. W., Neale, D., Mckeand, S., and Isik, F. (2013). Genomic estimated breeding values using genomic relationship matrices in a cloned population of loblolly pine. *G3* 3, 909–916. doi: 10.1534/g3.113.005975

Zinkgraf, M., Liu, L., Groover, A., and Filkov, V. (2017). Identifying gene coexpression networks underlying the dynamic regulation of wood-forming tissues in Populus under diverse environmental conditions. *N. Phytol.* 214, 1464–1478. doi: 10.1111/nph.14492