Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*

Makobatjatji M. Mphahlele[1,2], Fikret Isik[3], Marja M. Mostert-O'Neill[2], S. Melissa Reynolds[2], Gary R. Hodge[3,4¶] and Alexander A. Myburg[2*]

[1]Mondi Forests, Research and Development Department, Trahar Technology Centre - TTC, P.O. Box 12, Hilton, 3245, South Africa

[2]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa

[3]Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA

[4]Camcore, North Carolina State University, PO Box 7626, Raleigh, NC 27695, USA

***Corresponding author**

Alexander A. Myburg

Email: zander.myburg@up.ac.za

Telephone: +27 12 420 4945

Abstract

Genomic selection (GS) can substantially reduce breeding cycle times in forest trees compared to traditional breeding cycles. Practical implementation of GS in tree breeding requires an assessment of significant drivers of genetic gains over time, which may differ among species and breeding objectives. We present results of a GS study of growth and wood quality traits in an operational *Eucalyptus grandis* breeding program in South Africa. The training population consisted of 1,575 full and half-sib individuals, genotyped with the *Eucalyptus* (EUChip60K) SNP chip resulting in 15,040 informative SNP markers. The accuracy of the GS models ranged from 0.47 (diameter) to 0.67 (fibre width). We compared a four-year GS breeding cycle equivalent to half of a traditional eight-year *E. grandis* breeding cycle and obtained GS efficiencies ranging from 1.20 (wood density) to 1.62 (fibre length). Simulated over 17 years, the ratio of the accumulated genetic gains between three GS cycles and two traditional breeding cycles ranged from 1.53 (diameter) to 3.35 (wood density). To realise these genetic gains per unit time in *E. grandis* breeding, we show that significant adjustments have to be made to integrate GS into operational breeding steps.

Keywords: *Eucalyptus grandis, molecular breeding, genomic selection, selection efficiency, genetic gains*

Introduction

Eucalyptus species constitute about 14 million of 261 million hectares of plantation forestry (Carle and Holmgren 2008). The genus *Eucalyptus* has more than 800 different species of which "the Big Nine" species *E. urophylla, E. tereticornis, E. camaldulensis, E. saligna, E. dunnii, E. grandis, E. pellita, E. nitens, E. globulus,* and their hybrids support the bioenergy, biochemical and biomaterials industries (Shepherd et al. 2011; Stanturf et al. 2013). The sustainability of these industries relies on the high adaptability, fast growth and superior wood quality of *Eucalyptus* species in forestry plantations. Climate change is affecting traditionally productive forestry plantation areas (Booth 2013; Irland et al. 2001), as are pests and pathogens, aggravated by global trade (Wingfield et al. 2008; Wingfield et al. 2015). Faster and more agile breeding approaches are therefore needed to ensure the future sustainability of eucalypt forestry plantations.

Traditional forestry tree breeding systems still face challenges, mainly due to the long rotation cycles associated with reproductive maturity and the time-to-maturity of commercially important traits such as wood quality. Selection of superior families and individuals in these tree breeding system primarily relies on the empirical breeding values of valuable traits such as tree growth, wood quality and tolerance to biotic and abiotic stresses (Namkoong et al. 1980; White et al. 1988). Genetic relatedness is a critical consideration in determining genetic merit. Pedigree information in BLUP analysis uses the additive relationships among individuals to derive the variance-covariance relationships among all observations when making genetic value predictions (Piepho et al. 2008). However, pedigree relationships represent the average proportion of shared alleles (at infinite loci) that are identical by descent (IBD), often ignoring Mendelian sampling effects among segregating individuals in families. This average

proportion leads to an overestimation of genetic parameters, thereby affecting the correlated response of the variables (Veerkamp et al. 2011).

The inclusion of single nucleotide polymorphism (SNP) markers, the most abundant form of DNA polymorphism in plant genomes (Agarwal et al. 2008; Mammadov et al. 2012), makes tracking of the Mendelian sampling effects of individuals in families possible (Hill and Weir 2010). Genomic relationship matrices derived from SNP markers accurately estimates the genomic proportions that are IBD to capture Mendelian segregation (within families) and allows detection of cryptic relationships (between families) and the correction of erroneous pedigree records (Hayes et al. 2009). Moreover, the genomic relationship matrix can be blended with the pedigree matrices of a much broader set of non-genotyped individuals to adjust the pedigree relationship coefficients of the non-genotyped individuals to predict genetic merit better, an approach called single-step genomic BLUP (Christensen and Lund 2010; Isik et al. 2017; Legarra et al. 2009; Misztal et al. 2013), with applications in tree breeding (Cappa et al. 2019; Klapste et al. 2018; Ratcliffe et al. 2017). The blending of pedigree and genome markers-derived matrices is a cost-effective approach to maximise the accuracy of breeding value predictions in the normally largely un-genotyped tree breeding populations with shallow open-pollinated pedigree structures.

Genomic selection (GS), as a breeding tool, is the prediction of the genetic merit of genotyped individuals without phenotypes (defined as *genomic estimated breeding values*, GEBVs). The prediction is based on the aggregate modeling of the genomic and phenotypic information of the training population (Meuwissen et al. 2001). In practice, GS approaches maximise genetic gain per unit time and cost by predicting breeding values early in the breeding cycle, eliminating field testing. GS studies have demonstrated encouraging results in its application in many genetic improvement systems, such as livestock (Bouquet and Juag 2013; Garcia-Ruiz et al. 2016; Hayes et al. 2008; Luan et al. 2009; Schaeffer et al. 2006; Wiggans et al. 2017;

Wolc et al. 2016) and crops (Bassi et al. 2016; Crain et al. 2018; Cros et al. 2019; Haile et al. 2018; Voss-Fels et al. 2019). GS continues to revolutionise breeding approaches, by not just enabling accurate prediction of related individuals, but by allowing complex interrogation of genetic and environmental interactions and identifying genomic regions that are stable or responsive to specific environments (Crossa et al. 2017).

In forest trees, GS is of particular benefit due to the extended breeding cycles because of delayed reproductive maturity and the need for early selection of late expressing (mature) growth and wood quality traits. Grattapaglia et al. (2011) performed the first simulation study to demonstrate the potential of GS in tree breeding and highlighted crucial factors to consider. Since then, numerous studies have demonstrated GS as a tool for accelerated tree improvement (Grattapaglia et al. 2018). GS studies in forest species have shown acceptable prediction accuracies for adoption in conifers such as *Pinus taeda* (Resende Jr et al. 2012b; Zapata-Valenzuela et al. 2013), *Picea glauca* (Beaulieu et al. 2014) and *Pinus pinaster* (Batholome et al. 2016; Isik et al. 2015). The reference genome sequence of *E. grandis* (Myburg et al. 2014) and the development of a robust genome-wide SNP genotyping (EUChip60K) chip platform (Silva-Junior et al. 2015) have created opportunities to study and apply GS in *Eucalyptus* species and their hybrids. Prediction of genetic merit using GS has performed well in interspecific *Eucalyptus* hybrids (Resende et al. 2012; Tan et al. 2017) where linkage disequilibrium (LD) is high, as well as in open-pollinated pure-species breeding populations of *E. pellita* and *E. benthamii* (Müller et al. 2017). Suontama et al. (2019) was able to demonstrate improved breeding value accuracy as well as increased genetic gains in *E. nitens* solid wood breeding population. Whereas, in an elite clonal population of *E. globulus* Duran et al. (2017), was able to demonstrate encouraging predictive ability estimates for wood density and stem volume. Predicting genetic merit of individuals earlier in the breeding cycle has the potential to increased gains per unit time in trees (Batholome et al. 2016; Li and Dungey 2018; Resende

et al. 2017). Adopting GS as an alternative or a complementary tree breeding strategy for growth and wood quality traits requires a practical demonstration of its benefits concerning traditional tree breeding (TB) approaches.

In this study, we interrogate the benefits of implementing GS for growth and wood quality traits in the context of an established *E. grandis* breeding programme. First, we evaluated the predictive ability of GS based on genome-wide SNP markers. We then analysed the efficiency of an accelerated GS breeding strategy compared to the conventional approach. Finally, we investigated the genetic gains that can be achieved by GS for growth and wood quality traits in *E. grandis*.

Materials and Methods

Training population and phenotype assessment

The training population was derived from a series of *E. grandis* (Hill ex Maiden) breeding trials provided by the Mondi South Africa tree breeding programme. The training population is composed of 1,548 trees, genotyped and phenotyped from four trials. The genotyped and phenotyped individuals per family were selected from Tygerskloof (eight per family), Montigny (seven per family) and Port Durnfort (nine per family) trials to span the phenotypic range relative to the diameter plot average, whereas, for the Ncalu trial, all available standing trees were sampled. The Ncalu trials is a full-sib progeny trial, whereas the Montigny, Tygerskloof and Port Durnfort trails are half-sib progeny age-age correlation trials. An additional 27 parental selections were also genotyped. The number of related parents between the trials ranged from 5 to 40, all progeny were from the same breeding generation. **Table S1** presents the information of the trial designs, environmental conditions, the family relatedness as well as summary statistics of the measured growth and wood quality traits. Growth and wood quality traits, such as diameter at breast height and tree height were measured at age seven using a diameter tape and a Vertex Hypsometer (Haglof, Sweden), respectively. Non-destructive wood sampling was performed at breast height from bark-to-bark by extracting 22 mm increment cores and wood shavings. Extractives were removed from the increment cores by overnight soaking in acetone after which basic wood density was determined using the water displacement method (Tappi methods 258 om-02). Weighted fibre length and fibre width were measured using the MorFi Compact Fibre and Shive Analyzer (TachPap, France) from wood shavings macerated with a 1:1 solution of acetic acid and 50% hydrogen peroxide for four hours at 90°C. Percentage α-cellulose, as well as the syringyl to guaiacyl (S/G) lignin monomer ratio, were estimated using near-infrared (NIR) spectroscopy models with the OPUS/QUANT

Spectroscopy Software Version 6 (Bruker, Massachusetts) calibrated from *E. grandis* breeding material partially related to the training population **(Table S2)**. The raw phenotypic data was used directly in downstream analyses. Pedigree records showed that nine of the seed parents were shared between the full-sib and half-sib families as both seed and pollen contributors **(Table S3)**. The full-sib progeny in the training population were confirmed with microsatellite DNA fingerprinting (Brondani et al. 1998) as presented in **Table S3**.

Genotyping

DNA was extracted from immature xylem and cambium tissue scrapings of mature trees using the NucleoSpin DNA extraction kit (Machery-Nagel, Germany). The *Eucalyptus* (EUChip60K) SNP chip as described by (Silva-Junior et al. 2015) available from GeneSeek (Neogen, Lansing, MI, USA) was used for genotyping the 1,575 *E. grandis* trees. Informative SNP markers were retained with a call rate over 90% and a minor allele frequency above 0.05 using the SVS software v8.4.3 (Golden Helix, Inc. Bozeman, MT). The genotype information was coded based on the additive gene content model to zero, one, and two representing major homozygous (0.48), heterozygous (0.33) and homozygous minor (0.19) alleles, respectively (frequencies given in bracket). Missing allele data were imputed based on allelic distribution, assuming Hardy–Weinberg equilibrium using the *synbreed* R package (Wimmer et al. 2012).

Statistical analysis

Linear mixed models analysis

Due to the differences in progeny types and the trials design of single tree plots in Ncalu with 15 replication versus multiple plots at single replication in Montigny, Tygerskloof and Port Durnfort, we could only consider the site of the trials as our fixed model term in the analysis. The G×E random model term was not included because of the low connectedness of the trials. Linear mixed models were fitted to estimate variance components and solve for fixed and random effects. The matrix notation for the linear mixed models used was as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{e} \qquad \textbf{(1)}$$

where $\boldsymbol{y}$ is a vector of measured phenotypes, **X** and **Z** are the incidence matrix for the fixed and random effects, respectively. The $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are the vectors of fixed and random effect coefficients, respectively; $\boldsymbol{e}$ is the vector of residual effects. The site effect was considered as a fixed factor, while tree effect was treated as random. The expectations of $\boldsymbol{y}$, $\mathbf{u}$ and $\boldsymbol{e}$ are $\boldsymbol{E}(\boldsymbol{y}) = \mathbf{X\beta}$, $\boldsymbol{E}(\mathbf{u}) = \mathbf{0}$ and $\boldsymbol{E}(\mathbf{e}) = \mathbf{0}$ and the variances are $\boldsymbol{Var}(\boldsymbol{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$, $\mathbf{Var}(\mathbf{e}) = \mathbf{R} = \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$, and $\mathbf{Var}(\mathbf{u}) = \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}$, respectively, where **I** is the identity matrix, $\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$ is the variance associated with the residuals, and $\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}$ is the variance associated with the random effect. The $\boldsymbol{Var}(\mathbf{u})$ was scaled by the numerator relationship matrix **A** derived from the pedigree or by the matrix **G** derived from the SNP markers. Restricted maximum likelihood approach was used to estimated variance components with the ASReml-R 3.0 R package (Butler et al. 2009) in the R environment (R_Core 2016).

The Henderson (1975) mixed model equations were solved based on pedigree (ABLUP) to predict the empirical breeding values (EBV) of individuals:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\boldsymbol{\lambda} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\boldsymbol{y} \\ \mathbf{Z}'\boldsymbol{y} \end{bmatrix} \quad \textbf{(2)}$$

where $\mathbf{A}^{-1}$ is the inverted additive genetic relationship matrix derived from the pedigree, $\boldsymbol{\lambda} = \boldsymbol{\sigma}_e^2/\boldsymbol{\sigma}_u^2$ is the shrinkage factor. The direct genetic value (DGV) of individuals were predicted by solving the mixed model equations by substituting the $\mathbf{A}^{-1}$ matrix with the inverted realised genomic relationship matrix $\mathbf{G}^{-1}$. The genomic relationship was computed as described in VanRaden et al. (2008):

$$\boldsymbol{G} = \frac{(\mathbf{Z}-\mathbf{P})(\mathbf{Z}-\mathbf{P})'}{2\sum \boldsymbol{p}_i(1-\boldsymbol{p}_i)} \quad \textbf{(3)}$$

where $\boldsymbol{Z}$ and $\boldsymbol{P}$ are two matrices of dimension $\boldsymbol{n}$ (individuals) × $\boldsymbol{p}$ (markers). The gene content values in matrix $\boldsymbol{Z}$ are −1 (homozygote major allele), 0 (heterozygote), and 1 (homozygote minor allele). The allele frequencies in matrix $\boldsymbol{P}$ are presented as $\mathbf{2}(\boldsymbol{p}_i - \mathbf{0.5})$, where $\boldsymbol{p}_i$ is the allele frequency observed at the marker $\boldsymbol{i}$ for all individuals. The variance of alleles summed across all loci is $\mathbf{2}\sum \boldsymbol{p}_i\,(\mathbf{1} - \boldsymbol{p}_i)$.

The prediction accuracy of empirical breeding values (EBV) from the ABLUP models and direct genetic values (DGV) from the GBLUP models were estimated as

$$\boldsymbol{r} = \sqrt{\mathbf{1} - (\boldsymbol{SE}^2/(\mathbf{1}+\boldsymbol{F})\boldsymbol{\sigma}_u^2)} \quad \textbf{(4)}$$

where $\boldsymbol{SE}^2$ is squared standard error of the predictions, $\boldsymbol{\sigma}_u^2$ is the genetic variance component, and $\boldsymbol{F}$ is the inbreeding coefficient of the individuals (Gilmour et al. 2015), which is assumed to be zero for non-inbred individuals.

Genomic selection validation

A condition-free cross-validation approach was implemented to investigate the predictive ability of GBLUP models based on the above linear mixed model approaches. The cross-

validation sampling strategy involved 50 replications of random re-sampling of the whole population split into 90% training set ($\boldsymbol{TS}, n = 1425$) and 10% validation set ($\boldsymbol{VS}, n = 150$). This sampling mimics the situation in which 90% of the population is phenotyped and genotyped, and 10% is genotyped but without phenotypes. Genomic estimated breeding values (GEBV) of individuals in the validation set (genotyped individuals without phenotypes) were predicted by solving Henderson's linear mixed model equations in GBLUP. The variance components were fixed in the cross-validation analysis. The accuracy of the genomic selection models was determined as the correlation ($\boldsymbol{r_{GEBV:EBV}}$) between EBV and the GEBV. The prediction accuracy of the training set was determined as the correlation ($\boldsymbol{r_{GEBV:DGV}}$) of the GBLUP model predictions (DGV) with the GEBV.

Genomic selection efficiency

Genomic selection efficiency determines the effectiveness of the selection response achieved by the adoption of a GS breeding strategy versus the TB strategy. The selection response for GS was presented as the ratio between the selection accuracy ($\boldsymbol{r_{GEBV:EBV}}$) of the GS strategy and its breeding cycle time ($\boldsymbol{t_{GS}}$), whereas the selection response for TB is obtained as the ratio between the breeding value prediction accuracy and its breeding cycle time ($\boldsymbol{t_{TB}}$). The relative GS efficiency, in turn, was calculated as the ratio between GS and TB selection responses (Grattapaglia et al. 2011).

$$\boldsymbol{GSE} = \left(\frac{\boldsymbol{r_{GEBV:EBV}/t_{GS}}}{\sqrt{\boldsymbol{1-\left(SE^2/(1+F)\sigma_u^2\right)}}\,\boldsymbol{/t_{TB}}} \right) \qquad \textbf{(5)}$$

A sliding scale of GS breeding cycle times from three to nine years was used to simulate and compare the relative GS efficiency to that of TB cycle, which takes eight years to complete for *E. grandis*.

Genetic gains

Variance components were fixed when solving the linear mixed models to estimate EBV, DGV and GEBV for genetic gain calculations on the same scale. The expected genetic gain (%) for each of the traits was estimated by selecting the top 10% of individuals from the population expected to contribute to the next generation.

$$\%\boldsymbol{G} = \left(\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{BV}_{\boldsymbol{ij}} / \boldsymbol{\mu_i}\right) * \mathbf{100} \qquad \textbf{(6)}$$

where $\boldsymbol{\mu_i}$ is the population mean for trait $\boldsymbol{i}$ and $\boldsymbol{BV_{ij}}$ is the breeding value for individual $\boldsymbol{j}$ for trait $\boldsymbol{i}$. The genetic gains for GS and TB were simulated over 17 years, accounting for all breeding operations, including nursery activities **(Fig. 3)**. We assumed that the genetic gain is fixed at 10%, and the performance of GS is the same in every breeding cycle. The proposed 10% increase in genetic gains every breeding cycle corresponded with reported generational gains of *E. pellita* (Leksono et al. 2008) and *E. grandis* (Verryn et al. 2009). The ratio of the expected genetic gains accumulated over the 17 years for GS versus for TB should indicate the genetic gains benefit per unit time of GS over TB.

Results

Genetic parameters and relationship

A total of about 15,040 informative SNP markers were retained with 0.7% of the SNP marker alleles missing and imputed. The genomic relationship matrix was able to identify 63 selfed individuals from half-sib families with a coefficient of relationship greater than 0.75. The correction of these individuals in the pedigree resulted in a mean inbreeding coefficient increase from 1.0000 to 1.0041 **(Table 1)**. We observed lower than expected heritability estimates for growth and wood quality traits such as diameter (0.06), height (0.05) and cellulose (0.05) with the ABLUP models, with higher heritability estimates for wood quality traits such as S/G ratio (0.44) and fibre width (0.67) (**Table 2)**. The GBLUP models resulted in increased heritability estimates for both growth and wood quality traits with the highest increase for wood density (0.18 to 0.33) and the lowest for S/G ratio (0.44 to 0.45) (**Table 2)**. However, the heritability for fibre width decreased from 0.67 to 0.58. Pairwise Pearson correlations of the EBV of the growth and wood quality, as well as their distributions, are presented in **Fig S1.**

Table 1 Summary statistics of the pedigree and realised genetic relationship matrices. The corrected pedigree includes a revision of 63 (7%) individuals identified as selfs in the half-sib families (coefficient of relationship > 0.75) based on the genomic relationship matrix.

	Minimum	Maximum	Mean
Uncorrected Pedigree			
Coefficient of relationship	0.0000	0.5000	0.0202
Inbreeding coefficient	1.0000	1.0000	1.0000
Corrected Pedigree			
Coefficient of relationship	0.000	1.0000	0.0199
Inbreeding coefficient	1.000	1.5000	1.0041
Genomic Relationship			
Coefficient of relationship	-0.2631	1.2085	-0.0008
Inbreeding coefficient	1.0884	1.6258	1.2324

Table 2. Genetic variance components. The additive genetic (σ_u^2), and residual (σ_e^2) variance components and narrow-sense heritability (h^2) estimates and their standard errors (se) for the growth and wood quality traits from the ABLUP and GBLUP models.

	ABLUP			GBLUP		
	$\boldsymbol{\sigma_u^2}$	$\boldsymbol{\sigma_e^2}$	$\boldsymbol{h^2(se)}$	$\boldsymbol{\sigma_u^2}$	$\boldsymbol{\sigma_e^2}$	$\boldsymbol{h^2(se)}$
Fibre length	0.001	0.003	0.17 (0.055)	0.001	0.003	0.22 (0.035)
Fibre width	0.774	0.382	0.67 (0.093)	0.652	0.465	0.58 (0.036)
Cellulose	0.190	3.362	0.05 (0.026)	0.481	3.084	0.13 (0.031)
S/G ratio	0.020	0.020	0.44 (0.085)	0.020	0.024	0.45 (0.039)
Density	249.283	1101.067	0.18 (0.075)	440.855	903.703	0.33 (0.040)
Diameter	1.048	16.520	0.06 (0.032)	1.597	16.037	0.09 (0.030)
Height	0.690	12.049	0.05 (0.026)	1.099	11.646	0.09 (0.027)

Genomic selection accuracy and expected genetic gains

The use of the genomic relationship matrix resulted in an average increase of 7% in the breeding value prediction accuracy of DGV (ranging from 0.81 to 0.94) compared to the EBV (ranging from 0.76 to 0.87) across all traits **(Table 3)**. This resulted in an average increase of 24% in the expected genetic gain of the growth and wood quality traits for DGV compared to EBV **(Fig. 1)**. This suggested that using genomic realised relationship is efficient to capture true genetic relationships. The genetic gains of GEBV are on average 7% lower compared to DGV; however, still higher than the EBV across all traits except for cellulose and diameter, which were marginally lower **(Fig. 1)**. The accuracy of the GS model ($r_{GEBV:EBV}$) was calculated as an average of the 50 cross-validation folds ranging from 0.54 (density) to 0.67 (fibre width), whereas the prediction accuracy of the training set of the GS model ($r_{GEBV:DGV}$) ranged from 0.86 (diameter) to 0.98 (S/G ratio) **(Fig. 2)**.

Table 3 The average prediction accuracy (*r*) of the estimated breeding value (EBV) and direct genetic value (DGV) for the growth and wood quality traits. The range of the prediction accuracy and the standard deviations are presented.

	r_{EBV}			r_{DGV}		
	Accuracy	**Min-Max**	**Std.dev**	**Accuracy**	**Min-Max**	**Std.dev**
Fibre length	0.81	0.71-0.96	0.039	0.87	0.77-0.97	0.030
Fibre width	0.92	0.71-0.99	0.015	0.94	0.84-0.99	0.016
Cellulose	0.76	0.71-0.92	0.034	0.84	0.71-0.96	0.034
S/G ratio	0.87	0.71-0.99	0.024	0.92	0.81-0.98	0.021
Density	0.81	0.71-0.97	0.038	0.89	0.79-0.98	0.026
Diameter	0.77	0.71-0.92	0.035	0.82	0.66-0.94	0.037
Height	0.76	0.71-0.92	0.035	0.81	0.65-0.94	0.037

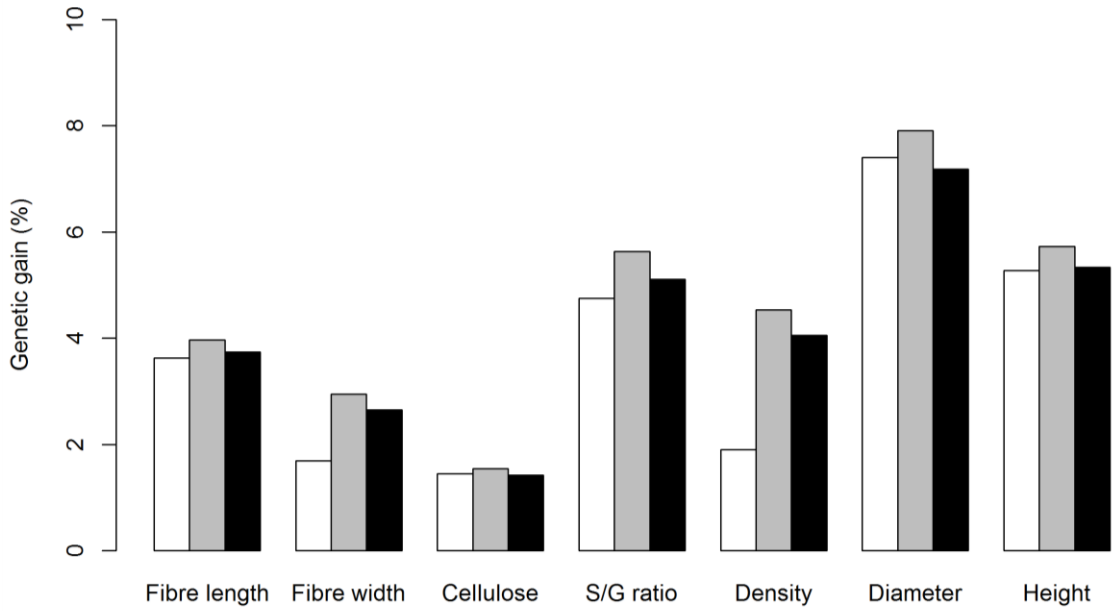

Fig. 1 The genetic gains for growth and wood quality traits. The gains were estimated from selecting the top 10% individuals from the EBV (*white*), DGV (*grey*) and GEBV (*black*) predictions for the individual traits. The variance components from the models were fixed in solving the mixed models

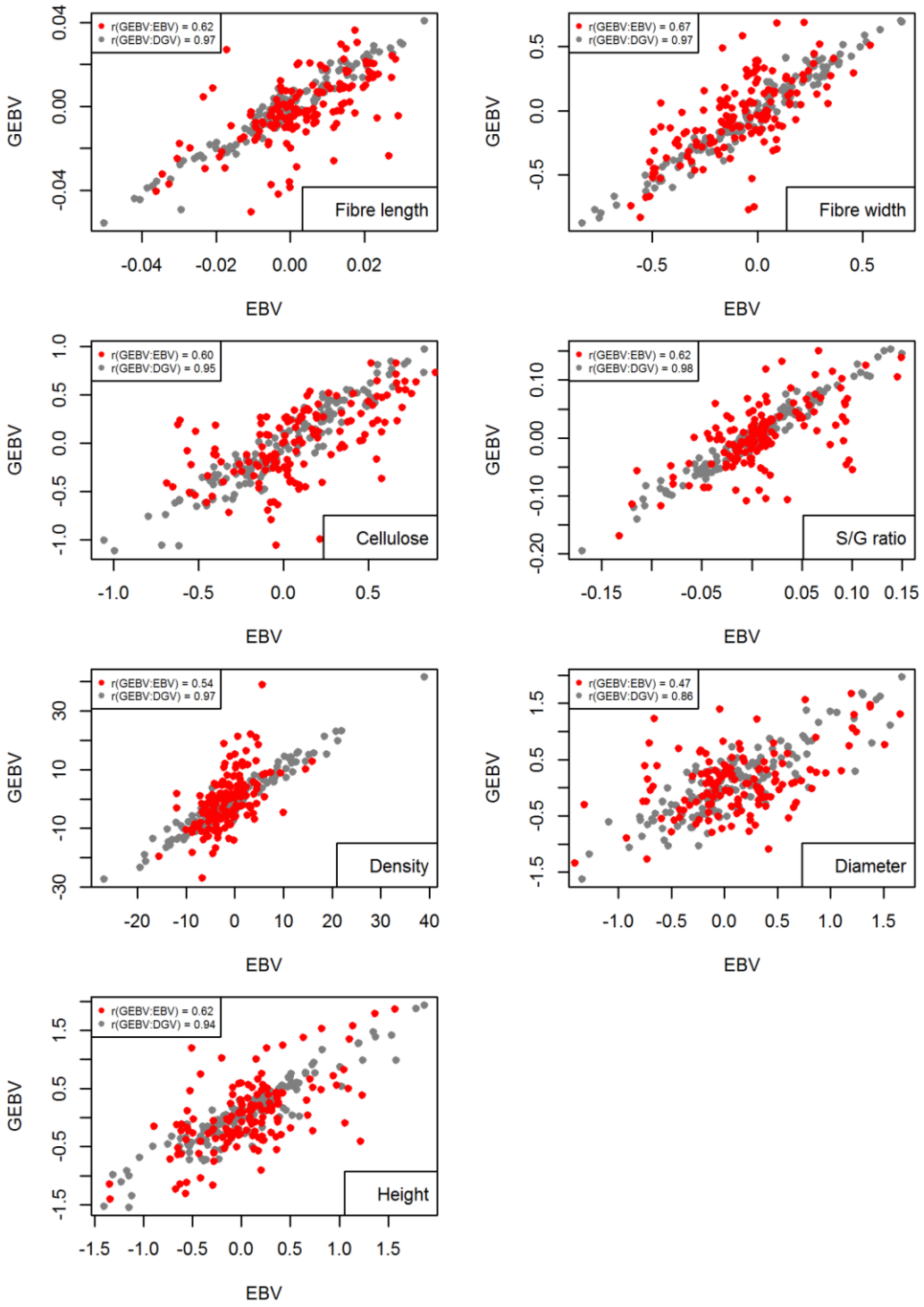

Fig. 2 Genomic selection accuracy (*r*) scatter plots for growth and wood quality traits for *E. grandis*. Depicted by the *grey dots* is the relationship between DGV (*x-axis*) and GEBV (*y-axis*) demonstrating the accuracy of the training set, while *red dots* show the relationship between EBV (*x-axis*) and the GEBV (*y-axis*) demonstrating the accuracy of the validation set

Genomic selection efficiency and genetic gains

The relative efficiency of GS compared to TB cycles, considered the selection responses of both breeding systems in a determined period (17 years), taking into account all operational breeding steps in the breeding cycles. We showed that for *E. grandis,* we could complete two 8-year conventional breeding cycles over 17 years, and three 4-year GS breeding cycles in the same period including a full-sib clonal trial to validate and update the GS model **(Fig. 3)**. For *E. grandis,* a 4-year GS breeding cycle necessitates flower induction treatments in a non-trials environment to enable controlled pollination between GEBV selected individuals. With a 4-year GS breeding cycle (seed-to-seed), the relative efficiency of the GS strategy was higher than an 8-year TB breeding cycle, ranging from 1.20 (wood density) to 1.62 (fibre length) **(Fig. 4)**. Note that GS predictions did not consider the actual genetic and phenotypic correlations of the traits at half-rotation (4 years) versus full-rotation (8 years), because the training and updating of the GS model would use full-rotation phenotype data for implementation **(Fig. 3)**. However, should the GS strategy take 5 years to complete the breeding cycle, then the relative efficiency for traits was lower, ranging from 1.17 (fibre width) to 1.29 (fibre length), with wood density not efficient at 0.96 **(Fig. 4)**. The relative efficiency of GS compared to TB diminished with increasing GS breeding cycle time, underscoring the significant effect of reproductive biology (seed-to-seed) on the feasibility of implementing GS in *E. grandis*.

The accumulated genetic gains as a benefit of the TB strategy over the 17 years ranged from 3.0% (cellulose) to 15.5% (diameter), whereas for the GS strategy it ranged from 4.7% (cellulose) to 23.8% (diameter) **(Table 4)**. The genetic gains ratio, which is the benefit of GS compared to the TB over the 17 years ranged from 1.53 (diameter) to 3.35 (wood density) **(Table 4)**, suggesting that there was an improved benefit in genetic gains over the 17 years with GS compared to a TB approach in *E. grandis.*

Table 4 The ratio of the genetic gains of genomic selection (GS) compared to traditional breeding (TB) accumulated over 17 years. The conditions of the simulation over the 17 years are a constant 10% increase in the percentage of genetic gain every breeding cycle and similar performance of GS Model A every breeding cycle.

Traits	Traditional breeding (TB)			Genomic selection (GS)				Genetic gain ratio GS/TB
	Cycle1	Cycle2	Total	Cycle1	Cycle2	Cycle3	Total	
Fibre length	3.6	4.0	**7.6**	3.7	4.1	4.5	**12.4**	1.62
Fibre width	1.7	1.9	**3.6**	2.7	2.9	3.2	**8.8**	2.47
Cellulose	1.5	1.6	**3.0**	1.4	1.6	1.7	**4.7**	1.55
S/G ratio	4.8	5.2	**10.0**	5.1	5.6	6.2	**16.9**	1.69
Density	1.9	2.1	**4.0**	4.1	4.5	4.9	**13.4**	3.35
Diameter	7.4	8.1	**15.5**	7.2	7.9	8.7	**23.8**	1.53
Height	5.3	5.8	**11.1**	5.3	5.9	6.5	**17.7**	1.60

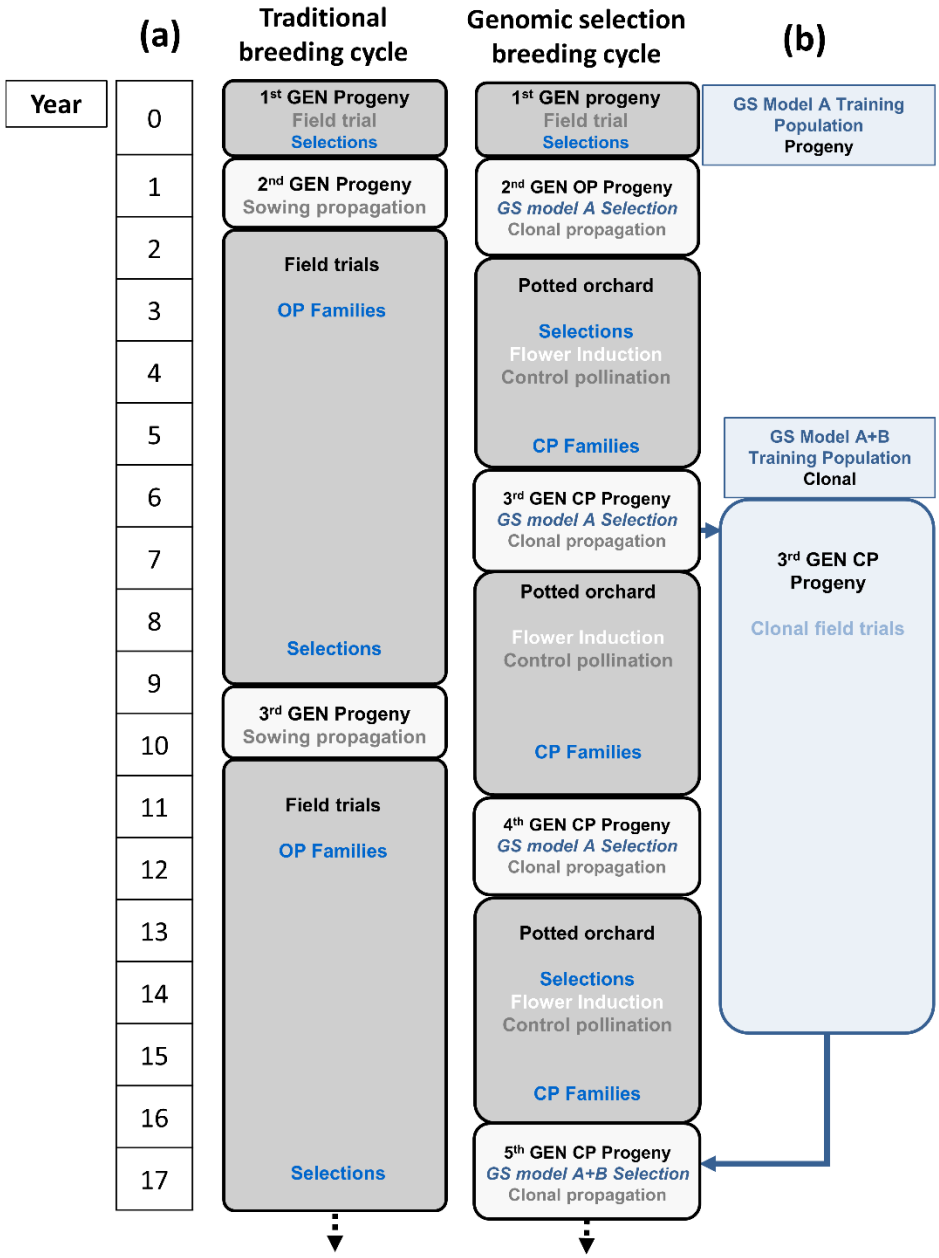

Fig. 3 Comparison of traditional and genomic selection breeding cycles over a 17 year period for *E. grandis*. (a) Two complete traditional eight-year TB cycles and (b) three complete four-year GS breeding cycles with a full rotation full-sib clonal trial to validate and update the current GS model. GS model A predictions are performed in year 1, 6, and 12, with the updated model (GS model A+B) from the control pollinated clonal progeny field trials used for prediction in year 17

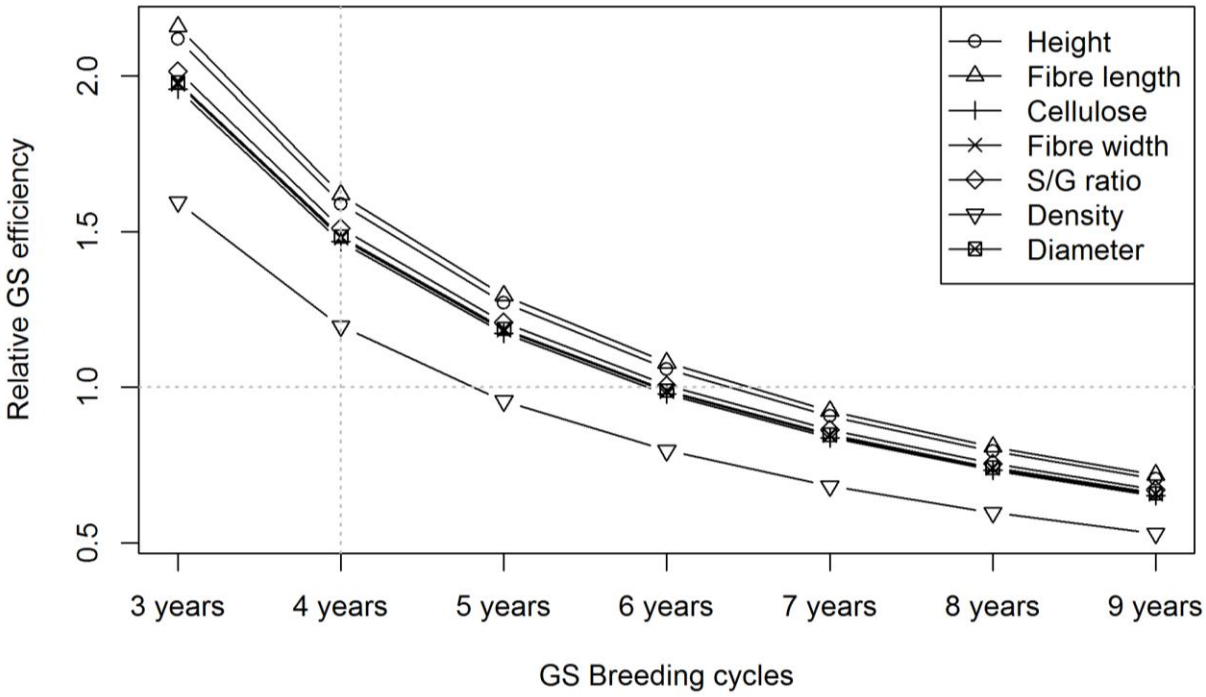

Fig. 4 The relative efficiency of GS overtime for the growth and wood quality traits compared to the eight-year TB cycle. GS efficiency decreases with the increase of its breeding cycle times. The *grey dotted line* indicates the intercept between the earliest time point (*x-axis*) to achieve reproductive maturity with flowering induction treatments under non-trial conditions and when the 8-year TB cycle is as efficient as the GS breeding (*y-axis*). Where the relative efficiency of GS is better than the TB, then the *y-axis* is >1.0, and when GS is less efficient than the TB, then the *y-axis* is <1.0

Discussion

The benefits of genomic selection are well documented foremost the acceleration of breeding cycles, and improvement of the accuracy of breeding values. Genome-wide DNA markers have also enabled better estimation of genetic parameters compared to pedigree records (Klapste et al. 2014). The sustained decrease in genotyping cost means that more innovative and complex strategies can be explored with GS approaches pursuing different breeding objectives (Grattapaglia et al. 2018). Such strategies may include regular updating of GS models to ensure their continued efficacy across generations, improving the accuracy of marker-trait associations and developing GS models that perform across multiple and changing environments. In this study, we investigated the use and practical implications of implementing GS strategies in an operational *E. grandis* breeding programme under realistic conditions including the challenge of using individual trees from unbalanced trials as opposed to a clonally replicated training population. The study is unique in that it demonstrates in practice the actual operational breeding cycle times required for the implementation of GS strategies compared to traditional breeding strategies. For this to happen, vital adjustments to operational breeding steps are required to accommodate and realise the benefits of adopting a GS breeding strategy for *E. grandis*.

In our training population, the heritability estimates of diameter (0.05) and height (0.06) were lower than expected for *E. grandis.* Harrand et al. (2009) previously reported heritability estimates of 0.16 and 0.14 for diameter and height, respectively, in *E. grandis*. The low heritability estimates observed in our study could be attributed to experimental design inefficiency due to the unbalanced trial design, resulting in the increases of the residual errors. Nevertheless, the use of the genomic relationship matrix resulted in higher and more precise heritability estimates **(Table 2)**, higher breeding value prediction accuracies **(Table 3)** and

higher genetic gains **(Fig. 1)** for growth and wood quality traits compared to the pedigree relationship matrix.

The prediction accuracy of the GS models for growth and wood quality traits ranged from 0.47 to 0.67 **(Fig. 2)**. The accuracies shown in this study are similar to those obtained in GS studies of other forest trees, for example, growth and stem-form in maritime pine *Pinus pinaster* ranging from 0.54 to 0.65 (Batholome et al. 2016) and from 0.39 to 0.49 (Isik et al. 2015). Similar accuracies were observed in the selection of *E. globulus* clones for density (0.60) and volume (0.73) (Duran et al. 2017). The prediction accuracy of disease, growth and wood quality traits of *Pinus taeda* breeding ranged from 0.20 to 0.46 (Resende Jr et al. 2012a). The genetic gains from the GEBV were lower than the DGV but higher than that of the EBV **(Fig. 1)**. The genetic gains from GEBV are lower than that from the DGV, because of the lower GS model prediction accuracy, but higher than that of the EBV because of the improved average breeding value prediction accuracy (ranging 0.81 to 0.87) and heritability estimates associated with the genomic relationship matrix. Together, these results suggest that there is sufficient GS prediction accuracy as well as improved genetic gains over pedigree to justify the implementation of GS strategies in *E. grandis* breeding.

Next, we investigated the relative efficiency of adopting the GS breeding strategy in *E. grandis* using simulation. We took into account the operational breeding steps and the respective breeding cycle times required for a current TB strategy versus the proposed GS breeding strategy **(Fig 4)**. One TB cycle takes eight years (full-rotation) for *E. grandis.* There is a strong correlation between growth and wood quality trait measurements at half-rotation vs full-rotation (Luo et al. 2010; Osorio et al. 2003; Rweyongeza 2016; Wu et al. 2007). Age-age genetic correlations higher than 0.90 have been reported in *E. nitens* for growth traits such as heights, diameter and volume, as well as basic density (Greaves et al. 1997). Although *E. grandis* can reach reproductive maturity at four years (half-rotation) under experimental (field

trials) conditions, the proportion of these individuals is small. Therefore, there is a need to wait until year eight to increase the percentage of flowering individuals so that most of the chosen selections have seed to turn over the generation. The GS breeding cycle should take four years incorporating all the necessary adjustments to advance GEBV selected seedlings into the next generation **(Fig. 3)** because, under non-trial conditions, flowering precocity in *Eucalyptus* species can be enhanced with growth regulators such as paclobutrazol (Griffin et al. 1993; Hasan and Reid 1995; Williams et al. 2003). In the case of *E. grandis*, flowering can be induced well within four years. This means that a GS strategy can be implemented with a plausible four-year seed-to-seed breeding cycle. Therefore, the GS strategy, together with accelerated flowering, can produce seedlings with predicted GEBVs, unlike TB, where seedlings from early flowering may not have EBVs. A four year GS breeding cycle would represent a 50% reduction compared to the traditional breeding cycle as simulated by (Grattapaglia et al. 2011). The operational adjustments required for GS approaches will differ when considering different species, breeding objectives, reproductive biology, propagation requirements and performances of seedlings vs clones to realise the expected accelerated gains per unit time (Isik 2014; Li and Dungey 2018; Resende et al. 2017). In our projection, the four-year GS breeding cycle will result in a relative efficiency ranging from 1.20 to 1.62 for growth and wood quality traits compared to the eight-year TB breeding cycle of *E. grandis* **(Fig. 3)**. Thus far, the GS breeding study has demonstrated encouraging GS prediction accuracy for growth and wood quality traits, higher genetic gains, and higher efficiency compared to a TB strategy.

Over the full 17-year period, the predicted ratio of the accumulated genetic gains for the GS and TB cycles range from 1.53 to 3.35, demonstrating the benefit of GS in terms of increased gains per unit time (**Table 4**). There is a realistic expectation that the cost of DNA isolation and operational expenses, including infrastructure maintenance, will increase with inflation over time. However, despite the need for repeated genotyping over the generations to update

the GS model, the cost of genotyping will reduce as high-throughput technology evolve over the 17 years, approaching zero compared to phenotyping cost. Our results suggest that a GS breeding strategy can be implemented for *E. grandis* and is likely to realise the expected genetic gains per unit time. However, it will be important to manage the significant adjustments and additional cost of operational breeding steps required to accommodate the GS strategy, as these may increase the cost per unit genetic gain in the short term, but will be offset in the longer term by increased profits at the mill.

The increased cost per unit genetic gain does come with additional benefits in that the TB and GS breeding strategies can advance two separate, but complementary crucial breeding objectives, which are to maintain genetic diversity and accelerate genetic gains, respectively. Current breeding strategies combine these two breeding objectives, mainly to maximise resource allocation. However, when the two objectives are separated, proper attention can be given to maximize their respective outputs. For example, the proposed three control-pollinated GS breeding cycles compared to the two open-pollinated TB cycles will more rapidly accumulate favourable alleles. The control pollination strategy embedded in the GS breeding approach should allow mating designs that maximise gains while limiting inbreeding, increase population LD, and allow controlled infusion of genotypes from the TB breeding cycle into the GS breeding cycle. Controlled pollination is essential for developing desirable training populations for GS, in that it confines allelic diversity within the training population (Habier et al. 2007; Rutkoski et al. 2015). The rapid development of a deep full-sib pedigree within a GS strategy has the potential to surpass the benefit provided by genome-wide markers in predicting genetic merit (Batholome et al. 2016). The replicated clonal full-sib progeny field trials running concurrently with the GS approach simulated in this study will serve to update and validate the GS model for genotype-by-environmental interactions. We have only highlighted the breeding implications of GS in *E. grandis* in this study. However, it is important

to note that the GS strategy will realize commercial products at least four years earlier than the TB approach offering an opportunity to recoup some of the investments required for the practical implementation of GS.

Conclusions and future prospects

The use of genome-wide DNA markers presents multiple advantages over shallow pedigree. These benefits include more accurate estimates of genetic relationships among individuals, higher and more precise heritability estimates, improved breeding value prediction accuracy and increased genetic gains. Significant practical adjustments to the TB cycle are required to realise the efficiency of GS in *E. grandis*. Enhanced benefits in the form of gains per unit time are achieved through the shortened operational breeding cycle of GS, mainly by overcoming the reproductive limits such as the time from unimproved seed to improved seed.

Looking forward, GS will become a common practice that will provide breeders with much more critical information to achieve breeding goals and produce elite clones for deployment. More sophisticated analysis models will be needed to help breeders accurately compute all of the information gained from interactions between the genome, phenome and the environment. It is important that practical studies like this assess tree breeding programs not only to inform the research objectives of tree breeders but also strategic decisions for managers.

Acknowledgements

We thank Mondi South Africa for the use of their *E. grandis* breeding population as well as the use of facilities and technical support to generate phenotypic data. SNP genotyping, wet chemistry data generation used to develop the NIR models and pedigree reconstruction was supported through funding from Mondi South Africa for the Forest Molecular Genetics (FMG) Programme at the University of Pretoria. The authors acknowledge North Carolina State University for training on R and ASReml analysis. We further thank the financial support from the funding organization Mondi South Africa (Pty) Ltd: Forests: Research and Development, National Research Foundation (NRF), Department of Science and Technology: Bioinformatics

and Functional Genomics Programme (BFG) Grant UID 86936 and 97911 and Department Trade and Industry South Africa: Technology and Human Resource Industry Programme (THRIP) Grant 80118 and 964134.

Compliance with ethical standards

The authors declare that they have no conflict of interest. The author M.M. is an employee of Mondi South Africa (Pty) Ltd.

Author's contributions

M.M. carried out the experimental design, data collection, data analysis, and drafting of this paper as part of his PhD thesis. F.I. assisted with the modelling and scripts of the data analysis tools. F.I., G.H., and A.A.M. supervised the study and helped with data interpretation and the drafting of the manuscript. M.O. assisted with SNP data analysis and M.R. assisted with DNA extraction and pedigree reconstruction. All authors have read and approved the final manuscript.

Data archiving statement

The full SNPs, pedigree, and phenotypic data are provided in the supplementary materials as supporting data.

References

Agarwal, M., Shrivastava, N., and Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports **27**(4): 617-631.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). Plant Science **242**: 23-36.

Batholome, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., and Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. BCM Genomics **17**(604).

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

Booth, T. H. (2013). Eucalypt plantations and climate change. Forest Ecology and Management **301**: 28-34.

Bouquet, A., and Juga, J. (2013). Integrating genomic selection into dairy cattle breeding programmes: a review. Animal **7**(5): 705-713.

Brondani, R. P. V., Brondani, C., Tarchini, R., and Grattapaglia, D. (1998). Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theoretical and Applied Genetics **97**: 816-827.

Butler, D. G., Cullis, B. R., Gilmour, A. R., and Goqel, B. J. (2009). ASReml-R reference manual.

Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Carcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. Plant Science **284**: 9-15.

Carle, J., and Holmgren, P. (2008). Wood from planted forests a global outlook 2005 - 2030. Forest Product Journal **58**(12): 6-18.

Christensen, O. F., and Lund, M.S. (2010). Genomic prediction when some animals are not genotyped. Genetics Selection Evolution **42:** 2.

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., and Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. The Plant Genome **11**(1): 1-14.

Cros, D., Mbo-Nkoulou, L., Bell, J. M., Oum, J., Masson, A., Soumahoro, M., Tran, D. M., Achour, Z., Guen, V. L., and Clement-Demanage, A. (2019). Within-family genomic selection in rubber tree (*Hevea brasiliensis*) increases genetic gain for rubber production. Industrial Crops and Products **138**: 111464.

Crossa, J., de los Campos, G., Maccaferri, M., Tuberosa, R., Burgueno, J., and Perez-Todriguez, P. (2017). Extending the marker x environment interaction model for genomic-enabled prediction and genome-wide association analysis in Durum Wheat. Crop Science **56**: 2193-2209.

Duran, R., Isik, F., Zapara-Valenzuela, J., Balocchi, C., and Valanzuela, A. (2017). Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. Tree Genetics & Genome **13**: 74.

Garcia-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-Lopez, F. J., and Van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. PNAS **113**(33): E4928.

Gilmour, A. R., Gogel, B.J., Cullis, B.R., Welham, S.J., and Thompson, R. (2015). ASReml guide release 4.1 functional specification. Hemel Hempstead, HP1 1ES, UK, VSN International Ltd (www.vsni.co.uk).

Greaves, B. L., Borralhoo, n. M. G., Raymond, C. A., Evans, R., and Whitehead, Ph. (1997). Age-age correlation in, and relationships between basic density desity and growth in *Eucalyptus nitens.* Silvae Genetica **45**(5): 264-270.

Grattapaglia, D., Macos, D., and Resende, D.V. (2011). Genomic selection in forest tree breeding. Tree Genetics and Genomes **7**(2): 241-255.

Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Muller, B. S. F., Tan, B., Isik, F., Ratcliffe, D., and El-Kassaby, Y. A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. Frontiers in Plant Science **9**: 1693.

Griffin, A. R., Whiteman, P., Rudge, T., Burgess, I. P., and Moncur, M. (1993). Effect of paclobutrazol on flower-bud production and vegetative growth in two species of *Eucalyptus*. Canadian Journal of Forest Research **23**(4): 640-647.

Habier, D., Fernando, R. L., and Dekker, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. Genetics **177**: 2389-2397.

Haile, J. K., N'Diaye, A., Clarke, F., Clarke, J., Knox, R., Rutkoski, J., Bassi, F. M., and Pozniak, C. J. (2018). Genomic selection for grain yield and quality traits in durum wheat. Molecular Breeding **38**: 78.

Harrand, L., Hernandez, J. J. V., Upton, J. L., and Valverde, G. R. (2009). Genetic parameters of growth traits and wood density in *Eucalyptus grandis* progenies planted in Argentina. Silvae Genetica **58**: 1-2.

Hasan, O., and Reid, J .B. (1995). Reduction of generation time in *Eucalyptus globulus*. Plant Growth Regulation **17**(1): 53-60.

Hayes, B. J., Bowman, P. J., Camberlain, A. J., and Goddard, M. E. (2008). Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science **92**: 433-443.

Hayes, B. J., Visscher, P.M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research **91**: 47-60.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. Biometrics **31**: 423-447.

Hill, W. G., and Weir, B. S. (2010). Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genetic Resources **93**: 47-64.

Irland, L. C., Adams, D., Alig, R., Betz, C.J., Chen, C-C., Hutchins, M., McCarl, B.A., Skog, K., and Sohngen, B.L. (2001). Assessing Socioeconomic Impacts of Climate Change on US Forests, Wood-Product Markets, and Forest Recreation: The effects of climate change on forests will trigger market adaptations in forest management and in wood-products industries and may well have significant effects on forest-based outdoor recreation. BioScience **51**(9): 753-764.

Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. New Forests **45**: 379-401.

Isik, F., Bartholome, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C., and Bouffier, L. (2015). Genomic selection in maratime pine. Plant Science **242**: 108-119.

Isik, F., Holland, J., and Maltecca, C. (2017). Genetic data analysis for plant and animal breeding. USA, Springer International Publishing.

Klapste, J., Lstiburek, M., and El-Kassaby, Y. A. (2014). Estimates of genetic parameters and breeding values from western larch open-pollinated families using marker-based relationship. Tree Genetics & Genomes **10**: 241-249.

Klapste, J., Suontanna, M., Dungey, H. S., Telfer, E. J., Graham, N. J., Low, C. B., and Stovold, G. T. (2018). Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. Journal of Heredity **109**(7): 802-810.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. Journal of Dairy Science **92**: 4656-4663.

Leksono, B., Kurinobu, S., and Ide, Y. (2008). Realized genetic gains observed in second generation seedling seed orchards of *Eucalyptus pellita* in Indonesia. Journal of Forest Research **13**(2): 110-116.

Li, Y., and Dungey, H. S. (2018). Expected benefit of genomic selection over forward selection in conifer breeding and deployment. PloS One **13**(12): e0208232.

Luan, T., Woolliams, J.A., Lien, S., Kent, M., Svendsen, M., and Meuwissen, T.H.E. (2009). The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics **183**: 1119-1126.

Luo, J., Zhou, G., Wu, B., Chen, D., Cao, J., Lu, W., Pegg, R. E., and Arnold, R. J. (2010). Genetic variation and age-age correlations of *Eucalyptus grandis* at Dongmen Forest Farm in southern China. Australian Forestry **73**(2): 67-80.

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. International Journal of Plant Genomics **2012**: 11.

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

Misztal, I., Aggrey, S. E., and Muir, W. M. (2013). Experiences with a single-step genome evaluation. Poultry Science **92**: 2530-2534.

Müller, B. S. F., Neves, L. G., de Almeida Filho, J. E., Resende, M. F. R., Muñoz, P. R., dos Santos, P. E. T., Filho, E. P., Kirst, M., and Grattapaglia, D. (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**(1): 524.

Myburg, A. A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D.,Goodstein, D.M., Dubchak, I., Poliakov, A., Mizrachi, E., Kullan, A.R.K., Hussey, S.G., Pinard, D., van der Merwe, K., Signh, P., van Jaarsveld, I., Silva-Junior, O.B., Togawa, R.C., Pappas, M.R., Faria, D.A., Sansaloni, c.P., Petroli, C.D., Yang, X., Ranjan, P., Tschaplinski, T.J., Ye, C-Y., Li, T., Sterck, L., Vanneste, K., Murat, F., Soler, M., San Clemente, H., Saidi, N., Cassan-Wanh, H., Dunand, C., Hefer, C.A., Bornberg-Bauer, E., Kersting, A.R., Vining, K., Amarasinghe, V., Ranik, M., Naithani, S., Elser, J., Boyd, A.E., Liston, A., Spatafora, J.W., Dharmwardhana, P., Rija, R., Sullivan, C., Romanel, E., Alves-Ferreira, M., Kulheim, C., Foley, W., Carocha, V., Paiva, J., Kudrna, D., Brommonschenkel, S.H., Pasquali, G., Byrne, M., Rigault, P., Tibbits, J., Spokeviciuos, A., Jones, R.C., steane, D.A., Vaillancourt, R.E., Potts, B.M., Joubert, F., Barry, K., Pappas Jr, G.J., Strauss, S.H., Jaiswal, P., Grima-Pettenati, J., Salse, J., Van de Peer, Y., Rokhsar, D.S., and Schmutz, J. (2014). The genome of *Eucalyptus grandis*. Nature **510**: 356-375.

Namkoong, G., Barnes, R.D., and Burley, J. (1980). Screening for yield in forest tree breeding. Commonwealth Forestry Review **59**(1): 61-68.

Osorio, L. F., White, T. L., and Huber, D. A. (2003). Age-age and trait-trait correlations for *Eucalyptus grandis* Hill ex Maiden and their implications for optimal selection age and design of clonal trials (Article). Theoretical and Applied Genetics **106**(4): 735-743.

Piepho, H. P., Mohring, J., Melchinger, and Buchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. Euphytica **161**: 209-228.

R_Core (2016). R: A language and environment for statistical computing. Vienna, Austria, https://www.R-project.org/, R Foundation for Statistical Computing.

Ratcliffe, B., El-Dien, O. G., Cappa, E . P., Porth, I., Klápste, J., Chen, C., and El-Kassaby, Y. A. (2017). Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. Genes Genomics Genetics **7**: 935-942.

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

Resende Jr, M. F. R., Munoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F., and Kirst, M. (2012b). Accuracy of genomic selection methods in standard data set of loblolly pine (*Pinus taeda* L.). Genetics **190**: 1503-1510.

Resende, M. D. V., Resende Jr, M.F.R., Sansaloni, C.P., Petroli, C.D., Missiaggia, A.A., Aguiar, A.M., Abad, J.M., Takahashi, E.K., Rosado, A.M., Faria, D.A., Pappas Jr, G.J., Kilian, A., and Grattapaglia, D. (2012). Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytologist **193**: 617-624.

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevero, C. F., Takahashi, E. K., Silva-Junior, O. B., and Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity **119**(4): 245-255.

Rutkoski, J., Singh, R. P., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J. L., and Sorrells, M. E. (2015). Efficient use of historical data for genomic selection: A case study of stem rust resistance in wheat. The Plant Genome **8**(1).

Rweyongeza, D. W. (2016). A new approach to prediction of the age-age correlation for use in tree breeding. Annals of Forestry Science **73**: 1099-1111.

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. Journal Animal Breeding and Genetics **123**(4): 218-223.

Shepherd, M., Bartle, J., Lee, D, J., Brawner, J., Bush, D., Turnbull, P., Macdonel, P., Simmons, B., and Henry, R. (2011). Eucalypts as a biofuel feedstock. Biofuels **2**(6): 639-657.

Silva-Junior, O. B., Faria, D.A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. New Phytologist **206**(4): 1527-1540.

Stanturf, J. A., Vance, E. D., Fox, T. R., and Kirst, M. (2013). *Eucalyptus* beyond its native range: Environmental issues in exotic bioenergy plantations. International Journal of Forestry Research **2013**(Article ID 463030).

Suontama, M., Klápště, J., Telfer, E., Graham, N., Stovold, T., Low, C., McKinley, R., and Dungey, H. (2019). Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. Heredity **122**: 370-379.

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. BCM Plant Biology **17**: 110.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science **91**: 4414-4423.

Veerkamp, R. F., Mulder, H. A., Thompson, R., and M. P. L. Calus. (2011). Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. Journal of Dairy Science **94**: 4189-4197.

Verryn, S. D., Snedden, C. L., and Eatwell, K. A. (2009). A comparison of deterministically predicted genetic gains with those realised in a South African *Eucalyptus grandis* breeding program. Southern Forests: a Journal of Forest Science **71**(2): 141-146.

Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. Theoretical and applied genetics **132**(3): 669-686.

White, T. L., and Hodge, G.R. (1988). Best linear prediction of breeding values in forest tree improvement programs. Theoretical and applied genetics **76**(5): 719-727.

Wiggans, G. R., Cole, J. B., Hubbard, S. M., and Sontegard, T. S. (2017). Genomic selection in dairy cattle: THe USDA experience. Annual Review of Animal Bioscieces **5**: 309-327.

Williams, D. R., Potts, B. M., and Smethurst, P. J. (2003). Promotion of flowering in *Eucalyptus nitens* by paclobutrazol was enhanced by nitrogen fertilizer. Canadian Journal of Forest Research **33**: 74-81.

Wimmer, V., Albrecht, T., Auinger, H-J., and Schon, C-C. (2012). synbreed: a framework for the analysis of genomic predictions data using R. Bioinformatics **28**: 2086-2087.

Wingfield, M. J., Brocherhoff, E.G., Wingfield, B.D., and Slippers, B. (2015). Planted forest health: The need for a global strategy. Science **349**(6250): 832-836.

Wingfield, M. J., Slippers, B., Hurley, B. P., Coutinho, T. A., Wingfield, B. D., and Roux, J. (2008). Eucalypt pest and diseases: Growing threats to plantation productivity. Southern Forests **70**(2): 139-144.

Wolc, A., Kranis, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Avendano, A., Watson, K. A., de los Campos, G., Fernando, R. L., Garrick, D. J., and Dekkers, J. C. M. (2016). Implementation of genomic selection in the poultry industry. Animal Frontiers **6**(1): 23-31.

Wu, H. X., Powell, M. b., Yang, J. L., Ivkovic, M., and McRae, T. A. (2007). Efficiency of early selection for rotation-aged wood quality traits in radiata pine. Annual Forestry Science **64**: 1-9.

Zapata-Valenzuela, J., Whetten, R. W., Naele, D., McKeand, S., and Isik, F. (2013). Genomic estimated breeding values using genomic relationship matrices in a cloned population of Loblolly Pine. Genes Genomics Genetics **3**: 909-916.

Supplementary Figures and Tables

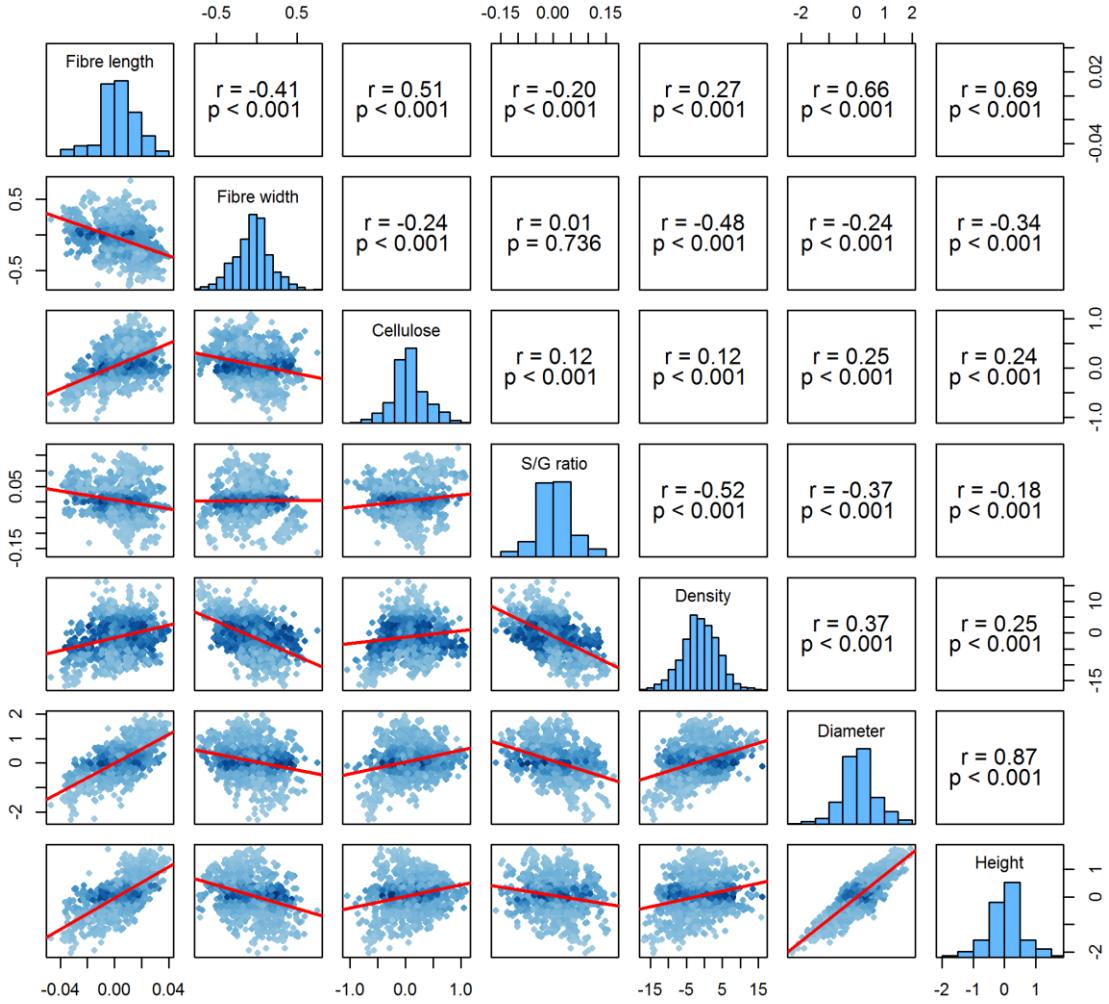

Fig. S1 Relationship of estimated breeding values (EBV) of the growth and wood quality traits. The lower triangle shows the scatter plots between the traits with a fitted linear regression (*red*), the distribution of the traits are shown on the diagonal, and the upper triangle provides the Pearson correlations (*r)* between the traits (H_0: $r = 0$).

Table S1 Tabular view of the plant materials and sites for the *E. grandis* study population.

Trial design, environmental conditions, overlap in families and summary statistics of measured phenotypes are presented.

	Mistley (CSO)	**Ncalu (NCA)**	**Montigny (MNT)**	**Tygerskloof (TYG)**	**Port Durnfort (PDF)**
Environment					
Latitude (South)	29^0 36'	30^0 20'	32^0 10'	31^0 16'	27^0 47'
Longitude (East)	30^0 22'	30^0 08'	28^0 35'	27^0 47'	27^0 47'
Altitude (m)	1023	1134	47	1118	32
MAP (ml)	832	809	1086	785	1454
MAT (^{0}C)	17	16	21	17	21
MAT min. (^{0}C)	5	4	11	4	11
MAT max. (^{0}C)	25	25	29	26	28
Trial design					
Plot design	Single	Single	Block	Block	Block
Progeny type	Parents	Full-sib	Half-sib	Half-sib	Half-sib
Tree per plot design	1	1x1	10x10	10x10	10x10
Families (Controls)	30	88 (5)	49	40	20
Replications	10	15	1	1	1
Number of plots	300	1395	49	40	20
Genotyped trees	27	709	343	320	176
Phenotyped trees		709	343	320	176
Family overlap					
Mistley		27	9	8	6
Ncalu			8	7	5
Montigny				40	20
Tygerskloof					18
Traits	**n**	**Min.**	**Max.**	**Mean**	**Std.dev**
Fibre length (mm)	1528	0.46	0.93	0.73	0.06
Fibre width (µm)	1528	15.86	23.52	19.70	1.07
α-Cellulose (%)	1538	36.00	54.10	46.06	2.08
S/G ratio	1538	1.26	2.77	2.14	0.22
Basic wood density ($kg.m^{-3}$)	1537	278.41	574.68	391.03	40.07
Diameter (cm)	1539	4.80	28.30	16.49	4.60
Height (m)	1539	7.90	30.90	20.74	4.76

MAP – mean annual precipitation. **MAT** – mean annual temperature. **Footnote:** The effective population size is 47.5, ranging from 21.9 for the full-sib trial to 46.3 for the half-sib trials. We used the formula described by Kimura, M., and Crow, J. F. (1963) for monoecious diploids**:** $\boldsymbol{N_e = \frac{N\underline{k}-2}{\underline{k}-1+\frac{V}{\underline{k}}}}$ where $\boldsymbol{\underline{k}}$ is the mean number of progeny per parent, $\boldsymbol{V}$ is the variance of the number of progeny per parent and $\boldsymbol{N}$ is the number of parents. **Reference:** Kimura, M., and Crow, J. F. (1963). The measurement of effective population number. Evolution **17**(3): 279-288.

Table S2 Near-infrared (NIR) spectroscopy models used to predict cellulose content and S/G lignin monomer ratio in the training population. The NIR model was calibration using 200 samples obtained from two *E. grandis* breeding trials (128 samples) and a subset from the Tygerskloof (TYG) trial (72 samples). NIR calibration set (fitted vs true) and the validation set (predicted vs true) parameters are presented. Data processing of the NIR scans for model development involved a combination of the first (1^{st}) derivative and vector normalization (SNV) computation using OPUS/QUANT Spectroscopy Software Version 6 (Bruker, Massachusetts).

Calibration	**Method**	**Samples**	**Rank**	**Slope[a]**	**R^2 [b]**	**RPD[c]**	**RMSEE[d]**	
α-Cellulose	SNV	124	9	0.919	92	3.52	0.661	
S/G ratio	1^{st} derivative + SNV	96	10	0.918	92	3.5	0.0591	
Validation	**Method**	**Samples**	**Rank**	**Slope[a]**	**R^2 [b]**	**RPD[c]**	**RMSECV[e]**	**Bias[f]**
α-Cellulose	SNV	67	9	0.907	90	3.08	0.745	0.00353
S/G ratio	1^{st} derivative + SNV	67	10	0.885	87	2.75	0.0733	-0.00168

[a]Linear regression slope of the calibration and validation data

[b]Coefficient of determination

[c]Residual prediction deviation

[d]Root mean square error of estimation

[e]Root mean square error of cross-validation

[f]Deviation from the linear regression slope

Table S3 Full and half-sib pedigree information of the training population. The full-sib progeny (706) in the Ncalu trial were confirmed with microsatellite DNA fingerprinting. The half-sib progeny (203) belong to nine of the seed parents using breeders' pedigree records with unknown pollen parents. A large number of trees (639) from half-sib families are not shown in the table because their seed and pollen parents are unknown. The total SNP genotyped population is 1,575 individuals, including the full-sib and half-sib progeny and the 27 parental selections.

	Pollen parents							
Seed parents	Unknown	Parent 03	Parent 07	Parent 08	Parent 15	Parent 16	Parent 18	Total
Parent 01			9					9
Parent 02		10	7	7		5	2	31
Parent 03	56		7		5	11	4	83
Parent 04		8	8	11				27
Parent 05	24	8	9					41
Parent 06		10	8		9	5	7	39
Parent 07	15	12		8		8		43
Parent 08		10			9	11	9	39
Parent 09		10			11			21
Parent 10	1	9	10		7	7	7	41
Parent 11		6		6				12
Parent 12			7		10	9	6	32
Parent 13		8	9		8		8	33
Parent 14		8	9	13	8			38
Parent 15	28	7	5	8		4	4	56
Parent 16		9	9		8		9	35
Parent 17		11	6	11	3			31
Parent 19		8		10	11	7	9	45
Parent 20		2		5	7	4	4	22
Parent 21		9			13	10		32
Parent 22	23	8	5	9	10	8		63
Parent 23	15	6	10	8	12	11	10	72
Parent 24			9		6	7		22
Parent 25		1						1
Parent 26	23							23
Parent 27	18							18
Total	203	160	127	96	137	107	79	**909**