



# Quantitative trait loci related to growth and wood quality traits in *Eucalyptus grandis* W. Hill identified through single- and multi-trait genome-wide association studies

Lucas Fernandes Rocha<sup>1</sup> · Thiago Romanos Benatti<sup>2</sup> · Leandro de Siqueira<sup>2</sup> · Izabel Christina Gava de Souza<sup>2</sup> · Isadora Bianchin<sup>2</sup> · Aguinaldo José de Souza<sup>2</sup> · Aline Cristina Miranda Fernandes<sup>2</sup> · Shinitiro Oda<sup>2</sup> · José Luiz Stape<sup>2</sup> · Rafael Massahiro Yassue<sup>3</sup> · Humberto Fanelli Carvalho<sup>4</sup> · Niels A. Müller<sup>5</sup> · Matthias Fladung<sup>5</sup> · Juan Jose Acosta<sup>6</sup> · Roberto Fritsche-Neto<sup>7</sup> · Evandro Vagner Tambarussi<sup>1,8,9</sup>

Received: 18 February 2022 / Revised: 26 September 2022 / Accepted: 1 October 2022 / Published online: 12 October 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

The aim of this study was to conduct single- and multi-trait genome wide association studies (GWAS) and identify quantitative trait loci (QTLs) for the expression of phenotypic traits in *Eucalyptus grandis*. We evaluated an open-pollinated breeding population with 1772 genotypes composed of 25 different families established using a randomized complete block design. We performed single-trait GWAS using the fixed and random model circulating probability unification (FarmCPU) and multi-trait GWAS for genetically correlated phenotypic traits using the multi-trait mixed model (MTMM). Then, gene annotation was identified through the Phytozome database. The FarmCPU model identified 43 and 38 QTLs that are significantly associated with growth and wood quality traits, respectively. Similarly, 40 pleiotropic QTLs were discovered using the MTMM model. Gene ontology for single-trait analysis identified loci responsible for regulating several important biological processes in different tissues and at different stages of maturation. On the other hand, the multi-trait model identified loci associated with gibberellin signaling, which regulates several aspects of plant growth and development, as well as loci related to the reinforcement of cell wall composition. Our study demonstrates the complex nature of *E. grandis* quantitative traits and provides new evidence of loci not described before which are associated with the expression of important phenotypic traits.

**Keywords** Association mapping · Genome-wide association study · *Eucalyptus* · Gene ontology

## Introduction

Genome-wide association studies (GWAS) are used to identify significant associations among quantitative traits and genetic loci in plant and animal genomes (Bush and Moore 2012). GWAS have been used extensively to understand the genetic complexity of economically important traits in tree species (Korte and Farlow 2013). Within the *Eucalyptus* genus, the species *Eucalyptus grandis* stands out because of its fast growth, high adaptability, and superior wood quality (Malan 1993). It is the most commonly planted hardwood tree globally, with a diverse range of applications in

cellulose, paper, timber, and charcoal production (Malan and Gerischer, 1987; Grattapaglia 2008; Carocha et al. 2015).

Cellulose in particular is a key wood product that meets a wide variety of primary human needs, such as paper (Hollertz et al. 2017; Jin et al. 2021), pharmaceuticals (Beyger and Nairn 1986; Giri et al. 2020), biofuels (Carere et al. 2008; Rubin 2008; Carroll and Somerville 2009), and food (Lavanya et al. 2011; Shi et al. 2014). Therefore, tree breeding strategies should focus on selecting genotypes considering not only growth characteristics, but also wood quality traits (Byram et al. 2005; Grattapaglia and Kirst, 2008; Apiolaza et al. 2013). To improve the quality of wood production, several studies have emphasized the importance of finding the genetic basis of wood quality traits such as lignin (Li et al. 2008; Hisano et al. 2009; Mizrachi et al. 2017), syringyl/guaiacyl ratio (Stackpole et al. 2011; Denis and Bouvet, 2013), wood density (Osorio et al. 2001; Stackpole et al. 2010), total extractives

Communicated by F. Isik

✉ Evandro Vagner Tambarussi  
tambarussi@gmail.com

Extended author information available on the last page of the article

(Gallo et al. 2018; Makouanzi et al. 2018), and cellulose yield (Schimleck et al. 2004; Kien et al. 2009). In this context, the application of genomics on forest improvement (Grattapaglia et al. 2009) and the development of new GWAS strategies are essential for identifying associations between genomic regions of the traits of interest and those significantly associated with the phenotype (Hirschhorn and Daly 2005). Several studies have identified loci related to the expression of growth and wood quality traits in *Eucalyptus* (Thavamanikumar et al. 2014; Lamara et al. 2016; Resende et al. 2017; Müller et al. 2017, 2019). Generally, growth traits tend to be more correlated with moderate levels of heritability, while wood quality traits are less correlated, but commonly present higher levels of heritability (Mphahlele et al. 2020). For *Eucalyptus*, Kainer et al. (2019) examined the genetic effects on oil yield, while Resende et al. (2017) conducted regional heritability mapping for growth and wood quality traits to identify quantitative trait loci (QTLs). Nevertheless, few studies have sought to understand the genetic effect of pleiotropic loci in *Eucalyptus* by comparing single-trait and multi-traits GWAS (Tan and Ingvarsson 2018; Ram-bolarimanana et al. 2018).

Pleiotropic effects occur when genetic loci have an influence on more than one trait (Solovieff et al. 2013). The application of pleiotropy in breeding means a movement away from selecting for one trait at the genetic level to selecting for multiple traits at a phenotypic level (Paaby and Rockman 2013). Although single-trait GWAS has identified the polygenic inheritance effect of markers, several efforts have been made to understand the pleiotropism between quantitative traits (Liu and Yan 2019), such as multiple trait selection assisted by genetic markers. Among single-trait GWAS algorithms, the fixed and random model circulating probability unification procedure (FarmCPU) performs a multi-locus linear mixed model (MLMM) to effectively control for spurious associations (Liu et al. 2016). On the other hand, multi-trait mixed models (MTMMs) were developed by Korte et al. (2012) to perform multi-trait GWAS and examine the common genetic effects that act in pleiotropy on two correlated phenotypic traits.

The MTMM algorithm performs three different analyses, categorized as full, common, and interaction. While the full model considers both common and interaction effects, the common and interaction models separate these effects individually. Thus, the common model performs a statistical analysis that demonstrates the coincident effects on two traits. Meanwhile, the interaction model identifies interacting genetic effects that act in the opposite direction between two traits (Korte et al. 2012). In the presence of pleiotropy, the power of the multi-trait GWAS is superior to single-trait GWAS because of the additional accuracy obtained when data for two traits

are considered together (Korte et al. 2012; Korte and Farlow 2013; Oladzad et al. 2019).

The present study focused on using GWAS to assess the genetic architecture of growth and wood quality traits of an open-pollinated *E. grandis* seed orchard. The specific objectives of the present study were to (1) develop and compare the significant loci using single- and multi-trait GWAS models in the identification of significant SNP markers related to growth and wood quality traits, (2) identify QTLs significantly associated with the expression of phenotypic traits, and (3) understand the pleiotropic effects and the genetic architecture of important traits.

## Material and methods

### Plant material and phenotypes

The study population was an open-pollinated seed orchard of *E. grandis* located in the municipality of São Miguel Arcanjo, São Paulo, Brazil (− 23.890188, − 47.937138). The population was established in September 2012 by the Suzano company's breeding team. The experiment consisted of a randomized complete block design, with four blocks, each containing 25 families (treatments) and one clonal control test (commercial clone), with four plots of 20 individuals each (five plants per plot). The spacing between plants was 3 m × 2 m, resulting in a planted area of 1.344 ha with 2240 trees. The open-pollinated seeds used to establish the experiment were collected from seven different locations across Brazil (Rio Claro — São Paulo (SP); Teixeira de Freitas — Bahia; Biritiba Mirim — SP; Salto — SP; Sarapui — SP; Mogi Guaçu — SP; and São Simão — SP) and one from Zimbabwe, Africa. The 25 families are originally from Coff's Harbour (New South Wales, NSW) and Ather-ton (Queensland, QLD), Australia.

For the analysis, we considered the genomic and phenotypic information from 1772 individuals. The control genotype was an *E. grandis* commercial clone used by the SUZANO company. The phenotypic information was subdivided into growth traits (GWTs) and wood quality traits (WQTs). Growth traits were measured at two different ages (3 and 6 years after planting) and were classified as height (HEI3 and HEI6) in meters and diameter at breast height (DBH3 and DBH6) in centimeters. The DBH (DBH3/DBH6) and height (HEI3/HEI6) were used to estimate tree volume at 3 and 6 years of age (VOL3 and VOL6, respectively) in cubic meters according to the formula described by Schumacher and Hall (1933):

$$\text{VOL} = \text{DBH}^2 \times \frac{\pi}{40000} \times \text{HEI} \times f$$

Furthermore, we analyzed six wood quality traits related to cellulose production. To do so, an increment borer was used at

breast height to collect wood cores of 12 mm at 6.5 years after planting. Then, wood material was sent to the laboratory for processing to obtain spectral information using near-infrared spectroscopy (NIRS).

Sawdust samples from 69 genotypes were used to create the curve calibration. The wood material was retained in a mesh sieve and placed in circular cells. The NIR reflectance spectra were obtained using scans of wavelength ranges. Curve calibration was based on samples from five different species (*Eucalyptus grandis*, *Eucalyptus urophylla*, *Eucalyptus brassiana*, *Eucalyptus tereticornis*, and *Eucalyptus pellita*) collected in three different regions of Brazil (Maranhão, São Paulo, and Bahia) at 6 years after planting. An internal company calibration (SUZANO S. A.) model was developed using the Bruker FT-NIR spectrophotometer MPA II. The resulting calibration database containing NIR wood spectra was obtained through following methods outlined in the reference literature. The prediction of constituent values based on existing calibration curves was used to estimate the following wood quality traits: pure cellulose yield (PCY) in percentage, basic wood density (WBD) in cubic meters, syringyl/guaiacyl ratio (SGR), soluble lignin (SOL) in percentage, total solid content (TSC), and total extractives (TEX) in percentage.

### Phenotypic data analysis

Each of the 1772 samples was evaluated using the Bonferroni outlier test to find the mean-shift outlier with studentized residuals in linear mixed models. Thus, outliers were removed by deleting observations based on standard deviation with the car package in the R software environment (Fox et al. 2012). Then, the normal distribution of phenotypic data was verified using the Shapiro–Wilk test, and data normalization was performed using the bestNormalize package in R (Peterson 2021). Finally, with the normalized dataset, the best linear unbiased predictions (BLUPs) (Rodriguez et al. 2020) were estimated for each trait with the breedR package in R (Muñoz and Sanchez, 2015) using the following mixed model:

$$Y_{ijk} = \mu + Xb_j + Zt_i + Zp_k + \varepsilon_{ij}$$

where  $\mu$  is the average mean;  $b_j$  is the fixed effect of the  $j^{\text{th}}$  block;  $t_j$  is the fixed effect of the  $j^{\text{th}}$  family effect (progeny);  $P_k$  is the random effect of the  $j^{\text{th}}$  plot with  $p \sim N(0, \sigma_p^2)$ ; and  $\varepsilon_{ij}$  is the residual error that represents the nongenetic effects. The matrices  $X$  and  $Z$  are the incidence matrices for the fixed and random effects, respectively. Deregressed best linear unbiased prediction/predictor (dBLUP) was then estimated to avoid shrinkage properties (Henderson 1975) according to the formula  $\frac{\hat{g}}{r^2}$  (Garrick et al. 2009), where  $\hat{g}$  is the genomic BLUP, and  $r^2$  is the reliability, estimated as  $1 - (\text{PEV}/\sigma_g^2)$ , where PEV is the prediction error variance, and  $\sigma_g^2$  is the genotypic variance. Pearson's genetic correlation tests were

then performed using the BLUPs to verify the correlation between the 12 growth and wood quality traits. Correlation distributions were plotted using the ggcorrplot package in R (Kassambara 2019). The significant  $p$ -values were estimated using function “p.mat”.

### DNA extraction and quality control

Cambium tissue was collected individually from 1772 trees and processed using the CTAB Lysis Buffer. DNA was extracted using the CTAB method (Doyle and Doyle 1987). DNA integrity was confirmed in 1% agarose gel electrophoresis and quantified by the Nanodrop spectrophotometer (Thermo Fisher, Waltham, MA, USA). DNA genotyping was performed using the EUChip60K high-density Illumina Infinium SNPchip for *Eucalyptus* species (Silva-Junior et al. 2015). Duplicate SNPs were eliminated from the raw dataset based on markers with the lowest call rate. Quality control was conducted using the R package snpReady (Granato et al. 2018). Markers were removed if they were monomorphic or had a call rate lower than 95%. Alleles with minor allele frequency (MAF) lower than or equal to 0.05 were also excluded. The genotypes were coded as “0” and “2” for homozygotes and “1” for heterozygotes. The remaining genotypic data was imputed using the R package snpReady considering Wright's equilibrium of the probability of occurrence considering the combination of allelic frequency and heterozygosity observed from the markers (Granato et al. 2018). Later, the filtered markers were submitted to linkage disequilibrium (LD) pruning, removing markers with a pairwise  $r^2$  higher than 0.99. This step was performed using the SNPRelate package in R (Zheng et al. 2012). After quality control, high-quality SNPs were selected for association mapping.

### SNP repositioning

We repositioned the markers using the information from the SNP probes in Illumina. Probe sequences were used to align with the second version of the *Eucalyptus grandis* reference genome (v2.0) ([https://data.jgi.doe.gov/refine-download/phytozome?genome\\_id=297](https://data.jgi.doe.gov/refine-download/phytozome?genome_id=297)) with the bowtie 2 aligner (Langmead and Salzberg 2012) and sensitive global alignment settings. The SNP position from version 2.0 was used in the GWAS analysis. We removed all scaffolds from the Brasuz v2.0 that were not in the linkage groups (from chromosome 1 to 11). The success of the repositioning was analyzed using a comparison map, and the dotplot coincidence graphs of the positioning of the two reference genomes (v1.0 and v2.0) were plotted using the R packages RIdeogram (Hao et al. 2020) and ggplot2 (Wickham 2011), respectively.

## Genetic parameters and population structure

The effective population size ( $N_e$ ) was estimated using the molecular linkage disequilibrium method (Waples and Do 2008) as implemented in NeEstimator V2.1 (Do et al. 2014). Population genetic parameters were estimated using the popgen function in the R package SNPReady (Granato et al. 2018), and include Nei's genetic diversity, as  $GD = 1 - p_j^2 - q_j^2$ ; polymorphic information content, where  $PIC = 1 - (p_j^2 + q_j^2) - (2p_j^2q_j^2)$ ; and minor allele frequency using the formula  $MAF = \min(p_j, q_j)$ . The observed heterozygosity ( $H_o$ ) was obtained with the formula:  $H_o = nH_j/N$ , where  $H_j$  is the number of heterozygous individuals, and  $N$  is the number of individuals. For each trait, we estimated the narrow-sense ( $h_a^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$ ) and broad-sense ( $h_g^2 = \sigma_a^2 + \sigma_d^2 / (\sigma_a^2 + \sigma_d^2 + \sigma_e^2)$ ) genomic heritability, where  $\sigma_a^2$  represents the additive variance,  $\sigma_e^2$  is the residual variance, and  $\sigma_d^2$  is the dominance variance. The narrow and broad sense heritabilities were estimated using the ASReml R package (Gilmour et al. 2017). Then, the degree of differentiation between the two origin populations ( $F_{ST}$ ) was estimated using the formula  $F_{ST} = 1 - H_S/H_T$ , where  $H_S$  is the average expected heterozygosity for each population (two different origins), and  $H_T$  is the expected heterozygosity in the total population.

The population structure was first analyzed by a principal component analysis (PCA) using genotypic data, where the first two principal components (PC1 and PC2) were used to determine the extent of population structuration. The two different origins were represented by different colors. We subsequently used the ADMIXTURE software to identify different genetic clusters with a fixed number of populations ( $K$ ) ranging from 1 to 40. Genetic correlation between phenotypes was estimated using the BreedR package in R (Munoz and Rodriguez 2014). Correlation was estimated in pairs considering the same model used to estimate BLUPs (Item 2.2). The genomic kinship matrix ( $G_a$ ) was obtained using the SNPReady package in R (Granato et al. 2018), following VanRaden (2008), with the following equation:

$$G_a = \frac{Z_A Z_A^T}{2 \sum_1^m p_i (1 - p_i)}$$

where  $Z_A$  is a matrix coded as 0 for homozygote  $A_1A_1$ , 1 for heterozygote  $A_1A_2$ , and 2 for homozygote  $A_2A_2$ ;  $p_i$  is the frequency of an allele from locus  $i$ ; and  $Z$  is an  $n \times m$  matrix of marker incidence ( $n$  is the number of genotypes, and  $m$  is the number of markers). In order to compare the difference between genomic and pedigree information, we estimated

the pedigree relationship matrix ( $A$ ) using the R package pedigree (Bates and Vazquez, 2013).

## LD decay

Genome-wide pairwise linkage disequilibrium (LD) was estimated for each chromosome using the function *LD.decay* from the sommer package v 2.9 in R v 4.0.2 (Covarrubias-Pazarán 2016). LD was estimated by the squared allele frequency correlation  $r^2$  between marker pairs, and the decay was plotted considering the first distance classes based on the marker matrix and a map with distances between SNPs on a loess curve. To investigate the average LD decay in the whole genome and within chromosomes, significant intra-chromosomal  $r^2$  values were plotted against the genetic distance between markers using the ggplot2 package in R (Wickham 2011).

## Genome-wide association study

We performed single-trait GWAS using the fixed and random model circulating probability unification (FarmCPU) (Liu et al. 2016) and multi-trait GWAS using the multi-trait mixed model (MTMM) (Korte et al. 2012) to identify genetic factors associated with the expression of phenotypic traits. The corrected phenotypic data (BLUP) and the genotypic information were used for single- and multi-trait GWAS. The single-trait association was performed using the genome association and prediction integrated tool (GAPIT) (Lipka et al. 2012; Tang et al. 2016). The population structure based on PCA matrix (Q) and kinship (K) were automatically generated (VanRaden 2008; Lipka et al. 2012) using genotypic data and the default GAPIT parameters. Using the GWAS results, we estimated the phenotypic variance explained by a significant marker (PVE), described as follows:

$$PVE = 2 * (\beta^2) * MAF * (1 - MAF)$$

where  $\beta$  is the effect of allele substitution, and MAF is the minor allele frequency of markers. The pleiotropic effect among phenotypic traits, which is a SNP marker having an effect on two or more traits, was estimated using the multi-trait mixed model (MTMM) (Korte et al. 2012). We performed multi-trait GWAS in pairs for the significantly associated growth and wood quality phenotypic traits. The R scripts provided by Korte et al. (2012) partition the interaction effects into three different analysis models: interaction, common, and full. Thus, considering two traits using a single marker model, the MTMM model can be written as (Korte, 2012):

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = s_1 \mu_1 + s_2 \mu_2 + x\beta + (x \times s_1)\alpha + \nu$$



where  $y_1$  and  $y_2$  are phenotypic values for genotype interactions of two traits. The  $y$  value is estimated as  $X\beta + u_G + u_{G \times E} + e$ , considering  $u_G$  and  $u_{G \times E}$  are the genotype and genotype-by-environment interaction values;  $s_1$  and  $s_2$  are vectors of 1 or 0 for all values of the trait in question;  $\mu_1$  and  $\mu_2$  are the means;  $x$  is the marker effect;  $\beta$  represents the effect size of fixed effects; and  $v$  is the prediction error. The interaction and common models identify markers that act differentially or in the same direction for two traits. On the other hand, the full model identifies SNPs with either an interaction or common effect. The significance threshold used for the  $p$ -values estimated by single- and multi-trait GWAS was calculated using the Bonferroni method ( $\alpha=0.05$ ). The  $p$ -values ( $-\log_{10} P$ ) for each evaluated SNP and model was used to generate Manhattan and QQ (quantile–quantile) plots using the R package CMPlot (Yin, 2018).

## Gene ontology

The significant SNPs for growth and wood quality traits were used to conduct a gene ontology analysis according to the physical distance within the GWAS peak regions. Since there were no strong LD blocks along the genome, which is probably related to LD-pruning, the downstream and upstream distance to search for candidate loci were estimated considering the distance of the two nearest flanking markers to the significant SNP. The genetic annotation and predicted functional effect of each gene were obtained by searching the database for version 2.0 of *E. grandis* from Phytozome v11.0 (Egrandis\_297\_v2.0.gene.gff3.gz). Venn diagrams were developed using the jvenn plot (Bardou et al. 2014).

## Results

### Phenotypic data

The number of outliers removed varied among the 12 phenotypic traits (DBH3: 63; HEI3: 111; VOL3: 2; DBH6: 0; HEI6: 6; VOL6: 1; PCY: 22; WBD: 21; SGR: 111; TSC: 23; SOL: 78; and TEX: 83). The genetic correlation among phenotypic traits ranged from  $-0.96$  (PCY/TSC) to 1 (DBH6/VOL6) (Fig. 1a and b). Similarly, the highest correlation between wood quality traits was found between TEX and SOL (0.62). The PCA biplot represents the first two components for the full set of 12 traits (six growth and six wood quality). The first two axes account for 50.7% and 17.9% of the variation in the phenotypic data (Fig. 1c).

### Population structure and genetic diversity parameters

The PCA using genotypic data revealed that the first component was mainly responsible for the genetic variation (55%) (Fig. S1). Although there was a slight grouping of genotypes according to their origin by PCA, the ADMIXTURE analysis showed an absence of population genetic structure (Fig. S2). Accordingly, the genetic differentiation ( $F_{ST}$ ) between individuals from two different origins presented a value of 0.036, indicating limited genetic divergence between them. A similar pattern was found for the kinship matrix (VanRaden 2008), where different subpopulations were identified but with no evidence of a strong population structure (Fig. 2a). Also, we notice that the genomic relationship matrix increased the prediction accuracy when compared with the pedigree information (Fig. 2b), with genotypes more and less related. Although the genotypes evaluated are originally from two native populations, the seeds which were used to establish the breeding population are from open-pollinated trials installed in eight different locations. Thus, we believe that crossings among genotypes from different origins may have generated stratification in the population.

In general, genetic diversity parameters showed moderate values. Nei's genetic diversity of the whole population ranged from 0.07 to 0.50, with an average of 0.35. The marker polymorphic information content (PIC) ranged from 0.07 to 0.38, with an average of 0.28. The minor allele frequency (MAF) showed a mean value of 0.26, ranging from 0.06 to 0.50. The observed heterozygosity ( $H_o$ ) had an average of 0.40, ranging from 0.24 to 0.47. Similarly, the inbreeding coefficient ranged from 0.04 to 0.50, with a mean value of 0.26. We found an effective population size ( $N_e$ ) of 31.5 considering linkage disequilibrium between markers ( $LDN_e$ ).

### SNP repositioning and quality control

In general, several SNPs changed their original relative position between the first (Myburg et al. 2014) and second (Bartholomé et al. 2015) versions of the *E. grandis* reference genome, and some even changed chromosomes. However, the genome-scale SNP collinearity (Fig. 3) between the two versions showed that most SNPs maintained similar positions. We did note a high collinearity pattern and more reliable linkage maps with version 2.0 (Bartholomé et al. 2015). Thus, we chose SNP positions estimated using the second version to perform GWAS analysis and identify QTLs and candidate loci related to trait expression.

For quality control, from the initial total of 64,639 markers, 3425 duplicate SNPs were removed considering the call rate, leaving 61,214 markers. After SNP repositioning, 1946

**Fig. 1 a** Genotypic correlation and **b** distribution of phenotypic traits for growth and wood quality categories across the 1772 *Eucalyptus grandis* genotypes; **c** principal component analysis for wood quality and growth traits. DBH3 diameter at breast height at 3 years; DBH6 DBH at 6 years; VOL3 volume at 3 years, VOL6 volume at 6 years; HEI3 height at 3 years; HEI6 height at 6 years; PCY pure cellulose yield; WBD basic wood density; SGR Syringyl/guaiacyl ratio; SOL = soluble lignin; TSC total solid content; and TEX total extractives

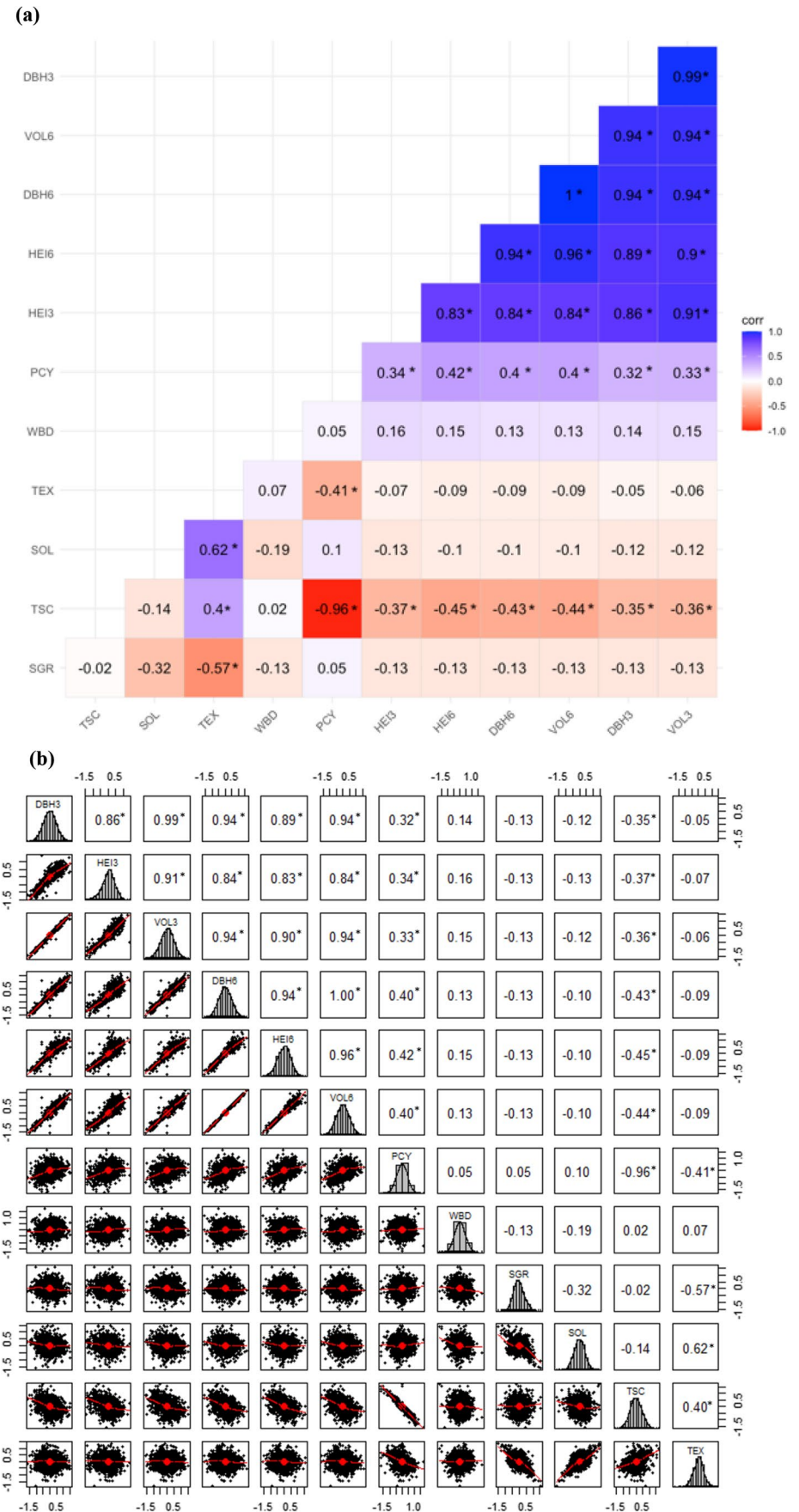
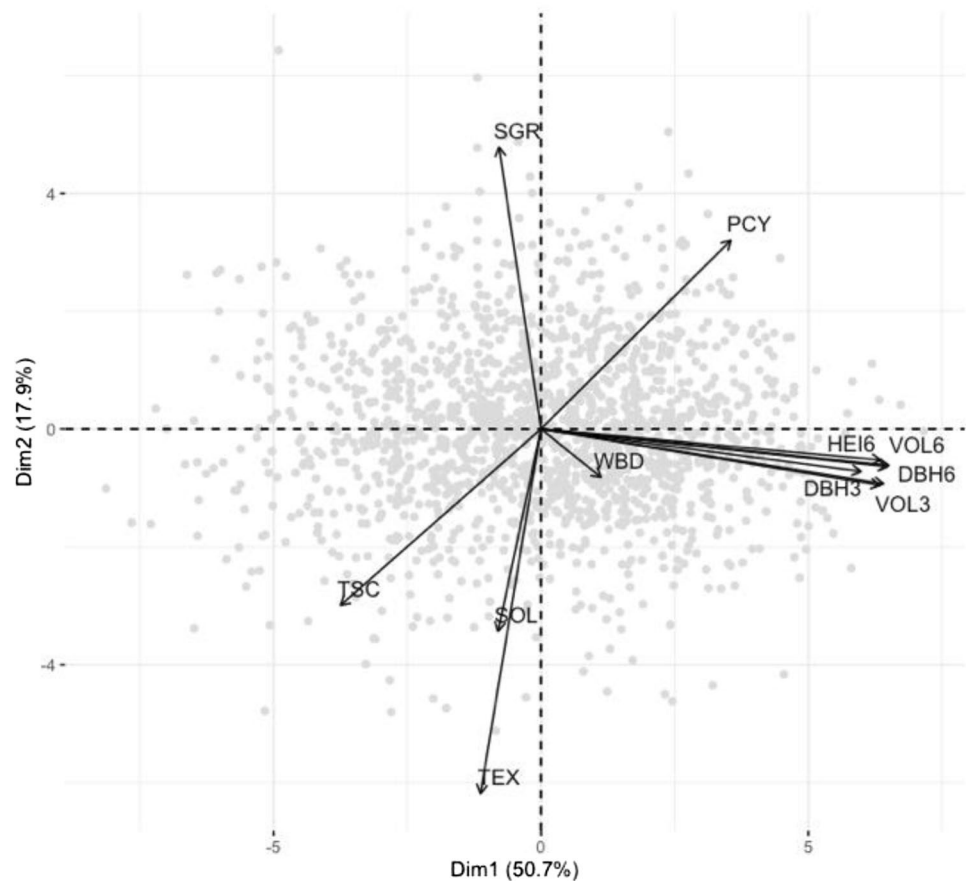


Fig. 1 (continued)

(c)



markers were removed as they were located in small scaffolds. A total of 28,957 markers were removed due to MAF (0.05), and 8,229 markers were removed due to the call rate (0.9), leaving 22,082 markers. Furthermore, 1.08% missing points were imputed. Finally, after LD pruning, 828 SNPs with high linkage disequilibrium were removed, leaving a final total of 21,254 markers for the analysis.

The informative SNPs selected were uniformly distributed across the 11 chromosomes of the *E. grandis* genome. Figure 3a shows the occurrence of SNPs along the *E. grandis* chromosomes, where the number of SNPs is summed within adjacent 1 Mb windows. LD showed a quick and similar decay pattern across the 11 *E. grandis* chromosomes (Fig. 3b). The ad hoc value of  $r^2$  (0.10) indicated an average LD across chromosomes ranging from 150 to 200 kb (Fig. S4).

### Genome-wide association studies

#### Broad- and narrow-sense heritability and single-trait genome-wide association study

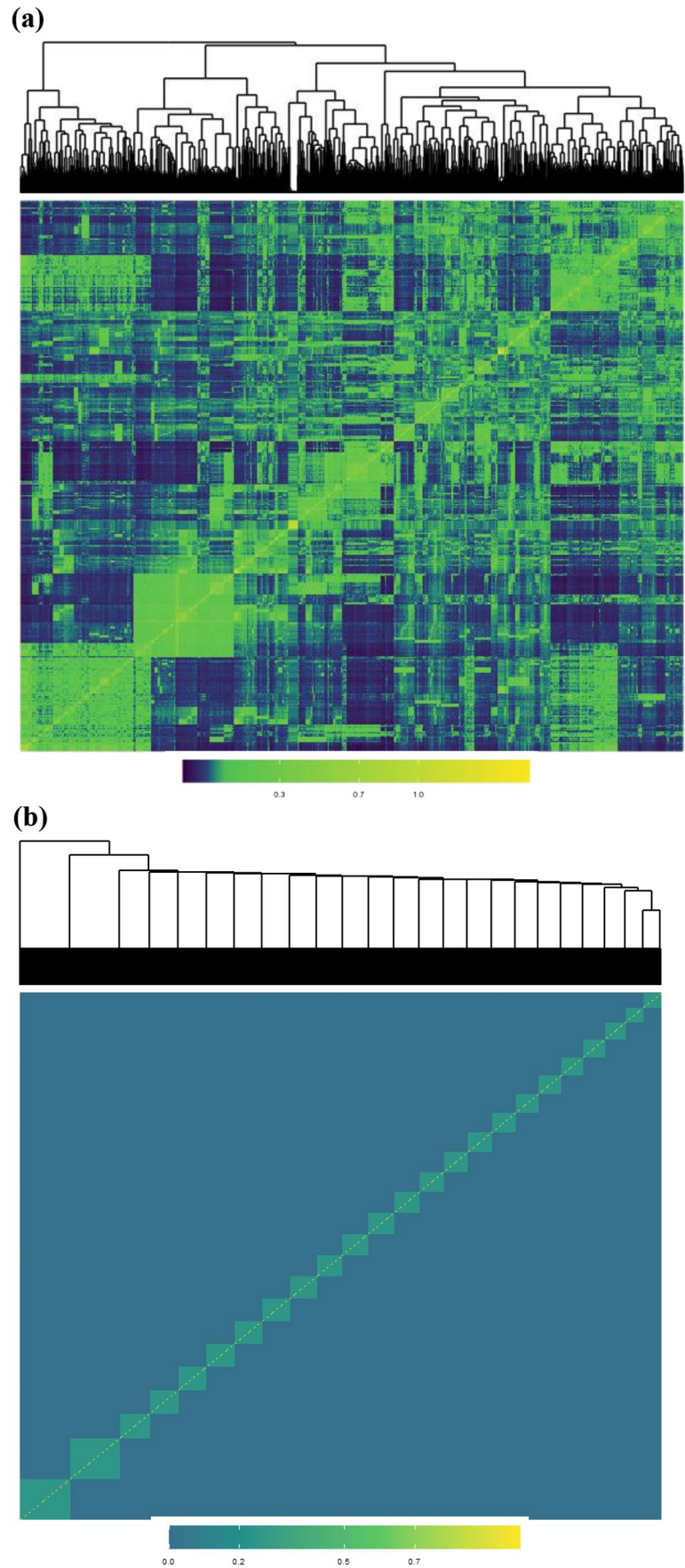
For growth traits, we found moderate values of narrow-sense heritability, ranging from 0.4299 (HEI3) to 0.5816

(DBH6) (Table 1). Three wood quality traits (SGR, SOL, and TEX) presented relatively low narrow-sense heritability (0.1599, 0.1845, and 0.1515, respectively). On the other hand, pure cellulose yield (PCY) presented the highest broad-sense heritability (0.7107) among all growth and wood quality traits.

The FarmCPU model successfully performed single-trait GWAS, indicating significant associations between growth and wood quality traits in *E. grandis*. After Bonferroni correction, a total of 81 SNPs with a significant association were identified for six growth traits (43 SNPs) and five wood quality traits (38 SNPs). Only the wood quality trait total extractives (TEX) showed no significant associations with markers (Table 1). The number of significant markers associated with phenotypic traits ranged from 2 (DBH6) to 14 (WBD) (Fig. 4a and b, respectively).

The average minor allele frequency (MAF) for the significant markers ranged from 0.0946 (DBH6) to 0.3164 (TSC). For all significant SNPs, the total phenotypic variance explained by a given SNP (PVE) was low, ranging from 0.0529 (SOL) to 0.2110 (PCY). Marker EuBR04s9558885 (PCY) showed the highest phenotypic variance (0.1014), suggesting a strong influence of this marker on phenotypic expression. Several markers were found associated with

**Fig. 2** Kinship heatmaps for the **a** marker relationships matrix estimated using the 21,254 SNPs based on the VanRaden method and **b** pedigree relationship matrix for the *Eucalyptus grandis* breeding population of 1772 individuals





multiple phenotypic traits for trait expression and candidate gene annotation (Fig. 4).

The number of annotated loci associated with the expression of phenotypic traits ranged from 0, with no gene annotation for the significant SNP (DBH6) to 46 (WBD). We found QTLs significantly related to more than one trait for both categories (Table S1; Fig. S3). In general, functional gene annotation presented several categories and descriptions associated with tissue growth on cell walls, cellulose biosynthetic process, transporter activity, DNA, ion and protein binding, oxidation–reduction process, and catalytic activity, among others. The function and description of all candidate loci for both growth and wood quality traits are shown in Table S1.

The pleiotropic effect among loci for growth traits was first seen for the SNP marker EuBR09s24960947, which presented the most significant association for traits DBH3 and HEI3, with  $p$ -values of  $9.34 \times 10^{-9}$  and  $1.91 \times 10^{-9}$ , respectively. This marker tags seven different loci (Eucgr.I01459, Eucgr.I01460, Eucgr.I01461, Eucgr.I01462, Eucgr.I01463, Eucgr.I01464, and Eucgr.I01465). In general, the single-trait GWAS revealed 13 candidate loci significantly associated with DBH3. Similarly, most loci found for HEI3 also showed comparable genome locations and molecular functions. Considering the trait HEI3, we found no annotation for candidate loci located near three significant SNPs (EuBR07s925067, EuBR06s38139098, and EuBR08s70063929) (Table S1). On the other hand, for HEI6, we found 44 annotated loci related to trait expression with different descriptions and gene ontology terms.

The SNP EuBR11s17004419 (HEI6) showed four different flanking loci (Eucgr.K01383, Eucgr.K01384, Eucgr.K01385, and Eucgr.K01386). The marker EuBR06s23565060 was identified for both ages for volume (VOL3 and VOL6) and for DBH3 ( $p$ -values  $8.29 \times 10^{-8}$ ,  $1.01 \times 10^{-8}$ ,  $8.44 \times 10^{-8}$ , respectively). We found a similar pattern of significant SNPs correlated with more than one phenotypic trait for wood quality in single-trait GWAS. For PCY and TSC, three different SNPs (EuBR07s252985, EuBR08s57640594, and EuBR10s1696823) were significantly correlated with the expression of these traits. These two phenotypic traits (PCY and TSC) presented the highest negative correlation ( $-0.96$ ), indicating that negative correlations can be effective in identifying pleiotropic loci.

Regarding pure cellulose yield, SNP marker EuBR04s9558885 had the highest significance ( $p$ -values  $1.06 \times 10^{-11}$ ), presenting five different flanking loci related to trait expression (Eucgr.D00522, Eucgr.D00523, Eucgr.D00525, Eucgr.D00526, and Eucgr.D00527). Additionally, SNP EuBR10s1696823 was related to PCY, with the presence of gene Eucgr.J00155, a wound-induced protein. Similarly, for TSC, we found 12 annotated loci significantly associated with trait expression. The marker EuBR01s39512949

presented six different loci related to its expression (Eucgr.A02909, Eucgr.A02910, Eucgr.A02911, and Eucgr.A02912).

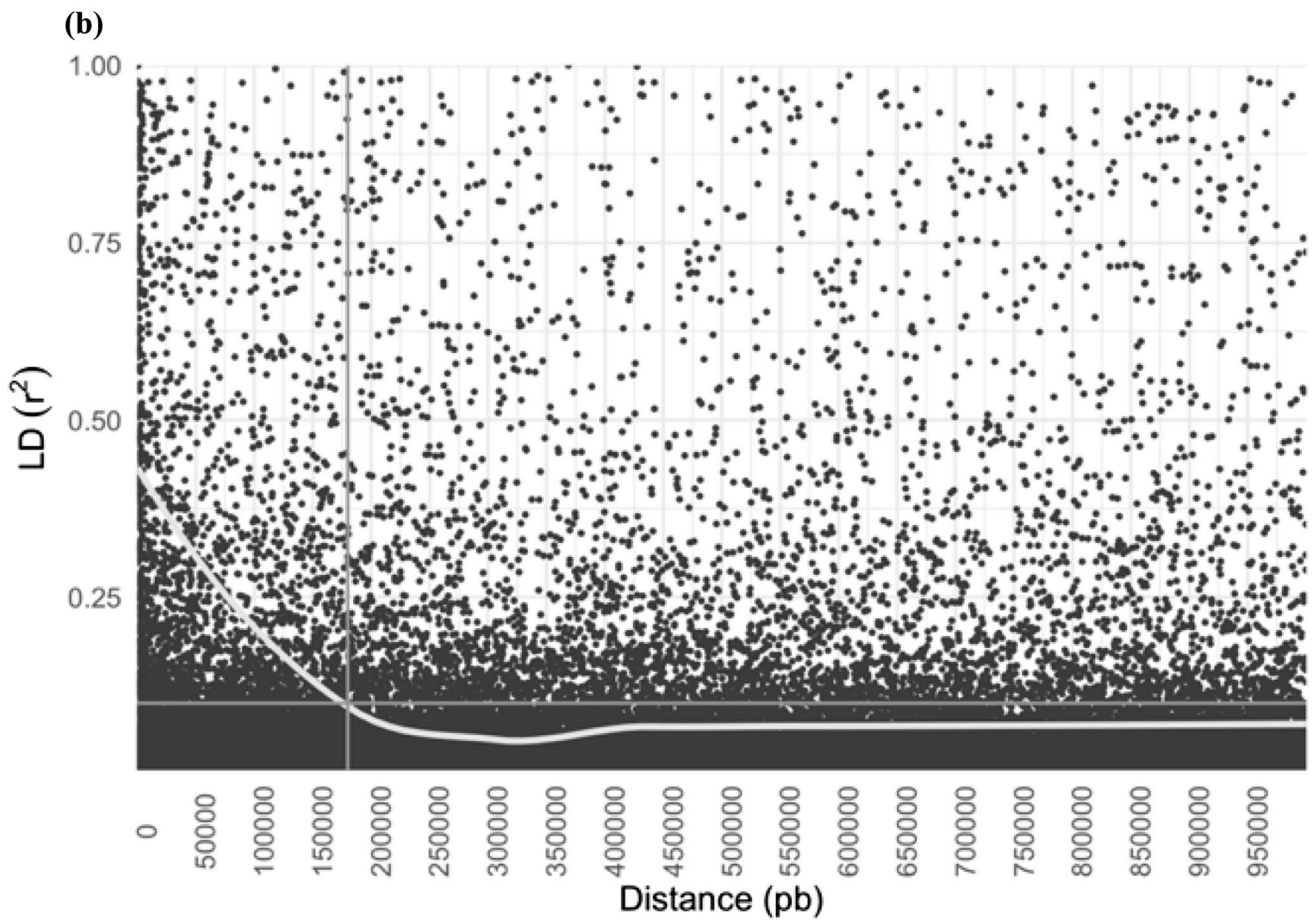
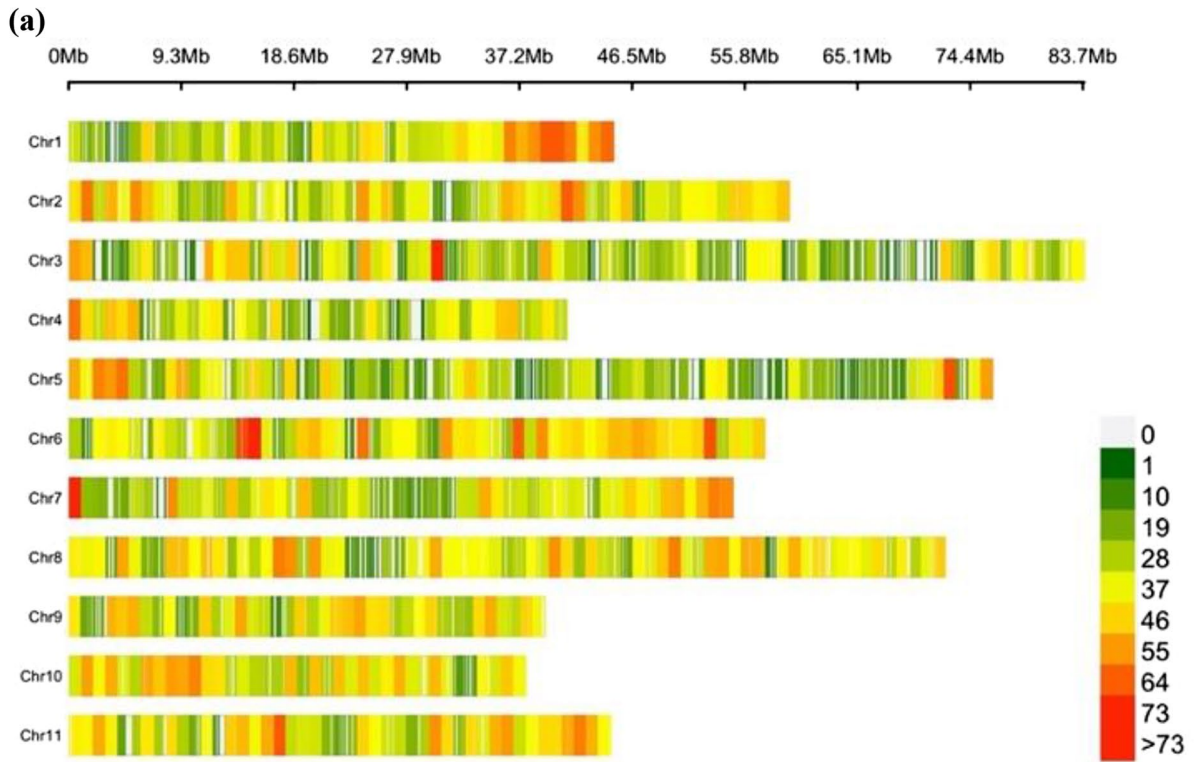
### Multi-trait genome-wide association study

The multi-trait GWAS showed good performance for all significant combinations among traits (Fig. S3). We found significant marker-phenotype associations for growth and wood quality traits that were not identified with the single-trait GWAS. Considering the 33 phenotypic correlations among the 12 traits, 22 combinations showed significant associations (Table 1; Fig. S3). The MT models (full, common, and/or interaction) for some models were unable to properly perform the GWAS since the  $p$ -values seemed to be deflated, and the QQ-plot showed more noise (e.g., Fig. S4d). These models were also unable to find significant associations considering the Bonferroni correction, which suggests no influence of possible false positives on the results.

The combinations among growth traits in multi-trait GWAS resulted in the highest number of significant SNPs with pleiotropic effects (24). Furthermore, the multi-trait GWAS analysis among wood quality (6) and between the two categories (GWT and WQT) (10) tended to express less significant markers. The multi-trait GWAS revealed 40 SNPs influencing the expression of multiple phenotypic traits (Table S6). Not surprisingly, the multi-trait methodology showed greater power to identify associations considering that most single-trait analyses using the MTMM methodology (Korte et al. 2012) could not identify associations due to the strict Bonferroni cutoff ( $\alpha=0.05$ ;  $p$ -value =  $1.63 \times 10^{-5}$ ).

Significant associations between the same trait in different years of data sampling (e.g., EuBR03s72654230 for DBH3 and DBH6; EuBR04s246324 for HEI3 and HEI6; EuBR02s2712998 for VOL3 and VOL6) indicate a strong pleiotropic effect on trait association. The SNP EuBR06s39120397 presented a strong  $p$ -value, and this marker was also statistically significant in the expression of HEI3, DBH6, and VOL6. This pattern may be related to the strong genetic correlation among these traits. Similarly, SNP EuBR01s28498846 was also found for the combination of traits TSC, DBH6, VOL6, and VOL3, indicating evidence of a pleiotropic effect of the marker on growth and wood quality traits.

We identified one SNP that was significantly associated with two trait combinations (HEI3 and DBH6; HEI3 and VOL6) (EuBR06s39120397;  $p$ -values  $7.57 \times 10^{-6}$  and  $1.14 \times 10^{-5}$ , respectively), with nine candidate loci related to its expression (Eucgr.F02939 ~ Eucgr.F02947). One SNP was found to be significant by the full model between the combinations of traits DBH3 and DBH6 and DBH3 and VOL6 (EuBR03s72654230;  $p$ -values  $1.59 \times 10^{-5}$  and  $1.02 \times 10^{-5}$ , respectively).



**Fig. 3** **a** SNP density plot across each chromosome representing the number of SNPs after quality control within a 1 Mb window size; **b** pairwise LD-decay across the 11 chromosomes of the 1772 individuals genotyped using the EUChip60K. Different colors represent different SNP density, and “Chr” represents the *E. grandis* chromosomes

The SNP EuBR03s43394028 was significant for three combinations of traits (HEI6 and VOL6, VOL3 and HEI3, and DBH3 and HEI6) ( $p$ -values  $9.33 \times 10^{-7}$ ,  $1.01 \times 10^{-5}$ , and  $1.54 \times 10^{-6}$ ). However, between traits HEI3 and HEI6, although SNP marker EuBR03s22449999 ( $p$ -value  $8.41 \times 10^{-6}$ ) was identified as significant by the common model, there were no annotations for candidate loci. On the other hand, SNP EuBR04s246324 was significant for the common and full models with a high  $p$ -value ( $9.05 \times 10^{-7}$ ). Considering that both traits HEI3 and HEI6 represent plant height, the power of multi-trait GWAS to detect significant candidate loci proved to be effective even for the same trait, considering different developmental stages.

One SNP marker was detected as significant for traits DBH3 and VOL3 (EuBR05s62102817;  $p$ -value  $2.19 \times 10^{-6}$ ) (Table S1). Similarly, between traits VOL3 and DBH6, the marker EuBR07s16969079 showed significant association ( $p$ -value  $1.38 \times 10^{-5}$ ). We found two significant markers for the first WQT combination SOL and TEX (EuBR06s19529730 and EuBR06s52964694;  $p$ -values  $2.66 \times 10^{-7}$  and  $5.06 \times 10^{-6}$ , respectively). The multi-trait GWAS combination between the traits SGR and TEX identified three significant markers (EuBR11s43922247, EuBR11s44284539, and EuBR03s16484895;  $p$ -values  $7.26 \times 10^{-6}$ ,  $1.16 \times 10^{-5}$ , and  $9.83 \times 10^{-6}$ ) through the full and interaction models. Similarly, the genomic regions for

marker EuBR11s44284539 revealed two flanking candidates loci (Eucgr.K03516 and Eucgr.K03517).

## Discussion

Single and multi-trait GWAS were effective in properly identifying QTLs as well as annotated loci related to phenotypic expression in the studied *E. grandis* breeding population. Additionally, the quality control process was able to remove uninformative markers, leaving a total of 21,254 highly informative markers that were used in the GWAS analysis. In general, most of the markers removed (28,957) during quality control were due to a low minor allele frequency (<5%), which is the frequency of the second most common allele in the population. In relation to the rearrangement of the *E. grandis* genome assembly, Bartholomé et al. (2015) identified 43 non-collinear and 13 non-synthetic regions. Thus, although there are modifications in marker collinearity found by the linear trend between the two versions of the genome, the new arrangement may be related to modifications in genome assembly.

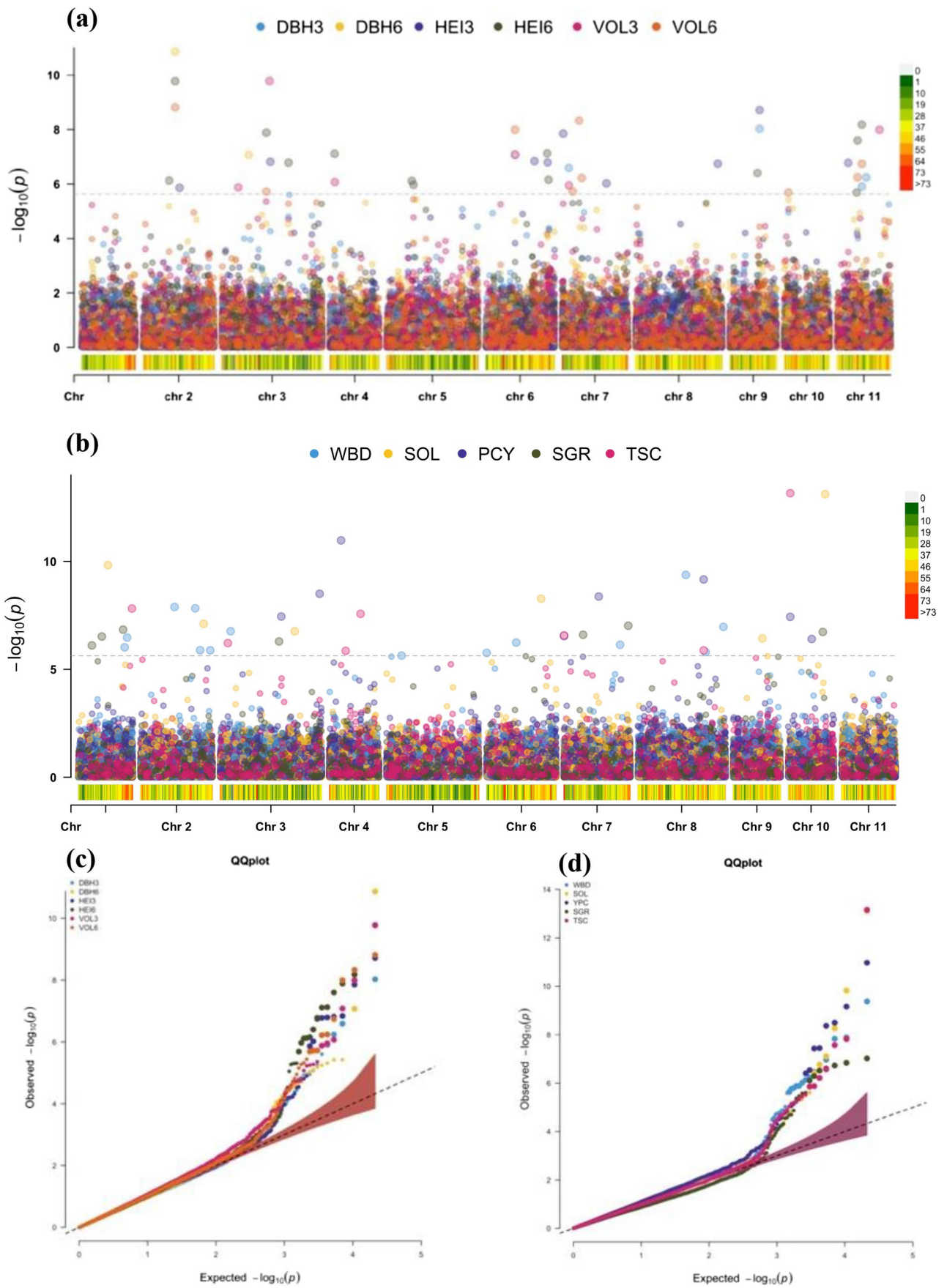
Although there were some rearrangements during SNP reposition, using new SNP positions for v2.0 of the genome was effective in finding QTLs and annotated loci. We reinforce that as far as we know, this is the first GWAS study developed using repositioned SNP probes that compares the positions of the two genome versions. Furthermore, considering that gene annotation is based on the second version of the *Eucalyptus* genome (Bartholomé et al. 2015), we believe that the possibility of errors was reduced.

**Table 1** Significant associations for growth and wood quality traits using the single-trait model (FarmCPU) for a *Eucalyptus grandis* breeding population. Traits are divided into growth (GWT) and wood quality (WQT). The number of SNPs, MAF, PVE, and number of loci are related to the significant number of associations found by the FarmCPU model

Type	Trait	$h_a^2$	s.e	$h_g^2$	s.e	SNPs	MAF	PVE	Loci
GWT	DBH3	0.57	0.08	0.68	0.06	5	0.1868	0.0976	10
	HEI3	0.43	0.07	0.55	0.06	8	0.2549	0.2108	28
	VOL3	0.56	0.08	0.65	0.06	6	0.1615	0.1039	15
	DBH6	0.58	0.08	0.68	0.06	2	0.0946	0.0551	0
	HEI6	0.56	0.08	0.68	0.06	13	0.2285	0.1502	44
	VOL6	0.55	0.08	0.63	0.06	9	0.1789	0.1910	21
WQT	PCY	0.61	0.07	0.71	0.06	8	0.3146	0.2110	22
	WBD	0.57	0.07	0.59	0.06	14	0.2404	0.1991	45
	SGR	0.16	0.05	0.18	0.05	7	0.2363	0.0867	10
	SOL	0.18	0.05	0.19	0.05	6	0.2423	0.0529	28
	TSC	0.54	0.05	0.59	0.05	3	0.3164	0.1197	14
	TEX	0.11	0.05	0.15	0.05	0	–	–	0

$h_a^2$  narrow-sense heritability,  $h_g^2$  broad-sense heritability, *SNPs* number of significant SNPs, *MAF* average minor allele frequency, *PVE* sum of phenotypic variance explained by the significant SNPs, *Loci* number of loci found, *GWT* growth traits, *WQT* wood quality traits, *DBH3* diameter at breast height at 3 years, *DBH6* DBH at 6 years, *VOL3* volume at 3 years, *VOL6* volume at 6 years, *HEI3* height at 3 years, *HEI6* height at 6 years, *PCY* pure cellulose yield, *WBD* basic wood density, *SGR* syringyl/guaiacyl ratio, *SOL* soluble lignin, *TSC* total solid content, *TEX* total extractives







**Fig. 4** Manhattan and QQ-plots of GWAS for growth traits (a, c) and wood quality traits (b, d), respectively, using the FarmCPU model for an *Eucalyptus grandis* breeding populations with 21,254 markers. Different colors represent different tested traits. Dashed line indicates the Bonferroni threshold ( $\alpha=0.05$ )

Another important point to consider is related to population structure. Herein, we found no clear structuration of the population between individuals, which may be related to the population's breeding history. Although there are two origins and it is likely that there would have been population structure, the breeding population was established from eight different provenances, which might have promoted out-crossing between individuals from different origins. According to Hayes (2013), not considering population structure in GWAS can cause false-positive associations. Thus, both models (single- and multi-trait GWAS) were tested against population structure, and we believe that this effect did not have an impact on our results as they were considered in the analysis.

Several genetic mapping through association studies have been used to assess the complexity of the genetic architecture of growth (Freeman et al. 2013; Müller et al. 2017, 2019), wood quality traits (Cappa et al. 2013; Freeman et al. 2013; Resende et al. 2017; Dasgupta et al. 2021), and non-wood traits (Resende et al. 2017; Kainer et al. 2019; Mhoswa et al. 2020) of *Eucalyptus*. Using the second version of the *Eucalyptus* genome, it was possible to more accurately identify QTLs. Many studies have also developed single-trait GWAS for growth, wood quality, and disease resistance in *Eucalyptus* spp. (Resende et al. 2017; Kainer et al. 2019; Müller et al. 2019; Ballesta et al. 2020; Mhoswa et al. 2020; Valenzuela et al. 2021). However, few studies have evaluated the multi-trait association models for growth and even fewer for wood quality in eucalypts (Rambolarimanana et al. 2018; Tan and Ingvarsson 2018). As expected, although several markers were found to be significant, the results from the single- and multi-trait GWAS indicate limited genetic variance, which can explain the relatively low number of associations. This pattern might be related to the polygenic nature of quantitative traits (Grattapaglia et al. 2018), indicating that there are many loci related to trait expression, as predicted by Fisher's infinitesimal model (Fisher 1918).

Although the complexity of multiple loci influences the expression of quantitative traits, the number of significant SNPs identified herein, and consequently the number of QTLs for both single and multi-trait GWAS, was similar to previous studies (Müller et al. 2019; Ballesta et al. 2020). Additionally, besides the reliable accuracy achieved by the single- and multi-trait GWAS, the phenotypic information used in the present study was obtained from a single environment, which may have limited the phenotypic precision of each individual. Thus, our study reinforces the importance

of using multi-trait models combined with single-trait models for highly complex quantitative traits. According to Liu et al. (2016), the FarmCPU model offers the best trade-off between predictive power and false positives. On the other hand, the power of the MTMM approach considering the correlation between two traits (multi-trait GWAS) can improve the identification of more evident pleiotropic effects than those found using a single marginal trait analysis (Korte et al. 2012).

The implementation of GWAS using phenotypic information from different traits can lead to the discovery of effects stronger than those identified by single-trait analysis (Korte et al. 2012). To increase the statistical power of GWAS, several studies have used multi-trait analysis to identify significant genetic-phenotypic associations (Jaiswal et al. 2016; Thoen et al. 2017; Yoshida and Yáñez 2021). Thus, multi-trait GWAS can increase the power of single-trait GWAS using different measures or multiple traits with a high pattern of genetic correlation (Porter and O'Reilly 2017). Regarding Pearson's genetic correlations between phenotypic traits, the strongest associations between growth variables found herein are expected because diameter, height, and volume are directly related. On the other hand, wood quality traits did not show strong patterns of association, except for PCY which presented several significant and positive associations with growth traits. This finding suggests that selection for growth traits might lead to a large increase in cellulose yield, which, for example, could have a further effect of reducing the total solid content production. Thus, pleiotropic QTLs are important when using marker-assisted selection for multiple traits.

Generally, our results show that multi-trait GWAS was able to increase the power of single-trait GWAS (FarmCPU) to identify loci that directly affect mutual traits, thus increasing the capacity to identify markers with minor effects. Furthermore, compared to the multi-trait GWAS (MTMM), the FarmCPU showed a lack of power to identify pleiotropic markers and correlated traits with low phenotypic correlation, as shown in previous studies (Korte et al. 2012). The joint association analysis, which considered the full, common, and interaction models, suggested genetic factors acting in the same direction, differentially, or with an interaction or common effect for the expression of the growth and wood quality traits.

Considering single-trait GWAS, several studies identified that FarmCPU increased the power of GWAS for complex traits (Tang et al. 2016; Kusmec and Schnable 2018; Miao et al. 2019). Our study corroborated this finding, identifying 81 significant markers for growth (43) and wood quality (38) traits. Furthermore, FarmCPU was able to control for false positives caused by population structure and kinship because of the distribution of quantile–quantile (QQ) plots. On the other hand, the MTMM model performed in the multi-trait

GWAS identified a smaller number of significant markers (31) among all significant trait combinations (Table S6). The importance of finding pleiotropic QTLs is related to marker-assisted selection, which can be used together to select multiple regions related to the expression of both growth and wood quality traits (Gupta et al. 2010). Regarding genomic heritability, low/moderate heritability levels were found for growth traits. The high/moderate heritability for the wood quality traits PCY and WBD indicates that they are less influenced by the environment. However, three wood quality traits (SGR, SOL, and TEX) showed a critically low heritability, making GWAS not appropriate for these traits.

Herein, the pleiotropic effect of loci influencing the expression of phenotypic traits was primarily found through single-trait GWAS analysis. A similar tendency was found by Ward et al. (2019) comparing yield traits in soft red winter wheat, where several markers presenting loci with pleiotropic effects were identified by the FarmCPU model. Here, pleiotropy was identified for both growth and wood quality traits in *E. grandis*. However, the markers with a pleiotropic effect identified for different traits by single-trait GWAS were not identified when using the multi-trait GWAS. The difference of significant SNPs found in these analyses might result from the different statistical methodologies that explore GWAS associations (Hayes 2013).

## Conclusion

Our study highlights the importance of examining associations between markers and phenotypes for eucalypt species. Herein, we identified markers that act individually on each trait using the single-trait GWAS and markers that have pleiotropic effects and influence several traits using multi-trait GWAS. The results corroborate previously published data for eucalypt species using moderate-size populations along with high-density SNP datasets. As far as we know, most of the markers identified herein have never been described in previous GWAS for eucalypt species. The results discussed herein provide a better understanding of gene expression and offer important information to inform marker-assisted selection.

In terms of identifying QTLs using single- and multi-trait GWAS, we were able to find clear results related to gene interaction. Gene ontology analysis of GWAS was also important in identifying the biological context of loci. The different GWAS methodologies applied involved the scanning of the whole genome from different trees and identifying genetic markers that can be used to predict phenotypic traits. As a result, GWAS effectively identified candidate loci related to the expression of phenotypic traits. We believe that the results can be used in genetic selection to increase the productivity of eucalypt plantations and improve future breeding programs. Nevertheless, further studies should be

conducted to identify significant associations with multiple environmental conditions. Thus, it is essential to continue evaluating the genetic effects and the complexity of the genetic architecture of economically important traits to continue to accumulate genetic gains in each breeding cycle.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11295-022-01570-x>.

**Acknowledgements** We acknowledge Suzano S.A. for providing the phenotypic and genotypic data.

**Author contribution** LFR conceptualization, methodology, software, formal analysis and writing — original draft. TRB conceptualization, resources, visualization, investigation and writing — review & editing. LS conceptualization, resources, visualization and funding acquisition. ICGS conceptualization, visualization, resources and methodology. AJS methodology and visualization. ACMF methodology, software and writing — original draft. SO methodology, conceptualization and visualization. JLS conceptualization and methodology. RMY methodology and visualization. HFC methodology, software, formal analysis and writing — original draft. NAM methodology, software, formal analysis and writing — original draft. MF methodology, software, formal analysis and writing — original draft. JJA methodology, software, formal analysis and writing — original draft. RFN methodology, software, formal analysis and writing — original draft. EVT writing — review and editing, visualization, supervision and project administration.

**Funding** This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) — Finance Code 001. The German Academic Exchange Service (DAAD) co-financed a short-term research grant (ref. no.: 91781916). Evandro V. Tambarussi is supported by a research productivity fellowship (grant number 304899/2019–4) and Post-Doctoral Scholarship (grant number 200727/2020–6) from “Conselho Nacional de Desenvolvimento Científico e Tecnológico” (CNPq).

**Code availability** Not applicable.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Conflict of interest** The authors declare no competing interests.

## References

- Ballesta P, Bush D, Silva FF, Mora F (2020) Genomic predictions using low-density SNP markers, pedigree and GWAS information: a case study with the non-model species *Eucalyptus cladocalyx*. *Plants* 9:99
- Bardou P, Mariette J, Escudié F et al (2014) jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* 15:1–7
- Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol* 65:415–442
- Bartholomé J, Mandrou E, Mabilia A et al (2015) High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytol* 206:1283–1296

- Bates D, Vazquez AI (2013) Pedigreemm: pedigree-based mixed-effects models. <http://pedigreemm.r-forge.r-project.org/>
- Beyger JW, Nairn JG (1986) Some factors affecting the microencapsulation of pharmaceuticals with cellulose acetate phthalate. *J Pharm Sci* 75:573–578
- Bush WS, Moore JH (2012) Genome-wide association studies. *PLoS Comput Biol* 8:e1002822
- Cappa EP, El-Kassaby YA, Garcia MN 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(11):1–16
- Carere CR, Sparling R, Cicek N, Levin DB (2008) Third generation biofuels via direct cellulose fermentation. *Int J Mol Sci* 9:1342–1360
- Carocha V, Soler M, Hefer C et al (2015) Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol* 206:1297–1313
- Carroll A, Somerville C (2009) Cellulosic biofuels. *Annu Rev Plant Biol* 60:165–182
- Chakraborty T, Akhtar N (2021) Biofertilizers: prospects and challenges for future. *Biofertilizers: Study and Impact* 575–590
- Covarrubias-Pazarán G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* 11:1–16
- Dasgupta M, Parveen ABM, Shanmugavel S et al (2021) Targeted re-sequencing and genome-wide association analysis for wood property traits in breeding population of *Eucalyptus tereticornis* × *E. grandis*. *Genomics* 113:4276–4292
- Denis M, 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
- Do C, Waples RS, Peel D et al (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size ( $N_e$ ) from genetic data. *Mol Ecol Resour* 14:209–214
- Doyle J, Doyle JL (1987) Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull* 19:11–15
- Fisher RA (1918) XV. The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ Sci Trans R Soc Edinburgh* 52:399–433
- Fox J, Weisberg S, Adler D, et al (2012) Package ‘car.’ Vienna R Found Stat Comput 16:
- Freeman JS, Potts BM, Downes GM et al (2013) Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus*. *New Phytol* 198:1121–1134
- Gallo R, Pantuza IB, dos Santos GA et al (2018) Growth and wood quality traits in the genetic selection of potential *Eucalyptus dunnii* Maiden clones for pulp production. *Ind Crops Prod* 123:434–441
- Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 41:1–8
- Gilmour A, Gogel B, Cullis B, et al. (2015) ASReml User Guide Release 4.1. Hemel Hempstead: VSN International
- Giri BR, Poudel S, Kim DW (2020) Cellulose and its derivatives for application in 3D printing of pharmaceuticals. *J Pharm Investig* 1–22
- Granato I, Galli G, de Oliveira Couto E, et al (2018) snpReady: a tool to assist breeders in genomic analysis
- Grattapaglia D (2008) Genomics of *Eucalyptus*, a global tree for energy, paper, and wood. In: *Genomics of tropical crop plants*. Springer, pp 259–298
- Grattapaglia D, Plomion C, Kirst M et al (2009) Genomics of growth traits in forest trees. *Curr Opin Plant Biol* 12(2):148–156
- Grattapaglia D, Silva-Junior OB, Resende RT et al (2018) Quantitative genetics and genomics converge to accelerate forest tree breeding. *Front Plant Sci* 22(9):1–16
- Hao Z, Lv D, Ge Y et al (2020) RIdiogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* 6(e251):1–11
- Hayes B (2013) Overview of statistical methods for genome-wide association studies (GWAS). *Genome-wide Assoc Stud genomic Predict* 149–169
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 423–447
- Hirsch S, Oldroyd GED (2009) GRAS-domain transcription factors that regulate plant development. *Plant Signal Behav* 4:698–700
- Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6:95–108
- Hisano H, Nandakumar R, Wang Z-Y (2009) Genetic modification of lignin biosynthesis for improved biofuel production. *Vitr Cell Dev Biol* 45:306–313
- Hollertz R, Durán VL, Larsson PA, Wågberg L (2017) Chemically modified cellulose micro- and nanofibrils as paper-strength additives. *Cellulose* 24:3883–3899
- Jaiswal V, Gahlaut V, Meher PK et al (2016) Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*T aestivum* L.). *PLoS One* 11:e0159343
- Jin K, Tang Y, Liu J et al (2021) Nanofibrillated cellulose as coating agent for food packaging paper. *Int J Biol Macromol* 168:331–338
- Kainer D, Padovan A, Degenhardt J et al (2019) High marker density GWAS provides novel insights into the genomic architecture of terpene oil yield in *Eucalyptus*. *New Phytol* 223:1489–1504
- Kassambara MA (2019) Package ‘ggcorrplot.’ R Package version 01
- Kien ND, Quang TH, Jansson G et al (2009) Cellulose content as a selection trait in breeding for kraft pulp yield in *Eucalyptus urophylla*. *Ann for Sci* 66:1–8
- Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29
- Korte A, Vilhjálmsson BJ, Segura V et al (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet* 44:1066–1071
- Kusmec A, Schnable PS (2018) Farm CPU pp efficient large-scale genomewide association studies. *Plant Direct* 2:e00053
- Lamara M, Raherison E, Lenz P et al (2016) Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. *New Phytol* 210(1):240–255
- Laskin JD, Heck DE, Laskin DL (2002) The ribotoxic stress response as a potential mechanism for MAP kinase activation in xenobiotic toxicity. *Toxicol Sci* 69:289–291
- Lavanya D, Kulkarni PK, Dixit M et al (2011) Sources of cellulose and their applications—a review. *Int J Drug Formul Res* 2:19–38
- Li X, Weng J, Chapple C (2008) Improvement of biomass through lignin modification. *Plant J* 54:569–581
- Li S-M, Zheng H-X, Zhang X-S, Sui N (2021) Cytokinins as central regulators during plant growth and stress response. *Plant Cell Rep* 40:271–282
- Lipka AE, Tian F, Wang Q et al (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399
- Liu H, Yan J (2019) Crop genome-wide association study: a harvest of biological relevance. *Plant J* 97:8–18
- Liu X, Huang M, Fan B et al (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* 12:e1005767
- MacMillan CP, Mansfield SD, Stachurski ZH et al (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and *Eucalyptus*. *Plant J* 62:689–703
- Makouanzi G, Chaix G, Nourissier S, Vigneron P (2018) Genetic variability of growth and wood chemical properties in a clonal population of *Eucalyptus urophylla* × *Eucalyptus grandis* in the Congo. *South for a J for Sci* 80:151–158
- Malan FS (1993) The wood properties and qualities of three South African-grown eucalypt hybrids. *South African for J* 167:35–44



- Malan FS, Gerischer GFR (1987) Wood property differences in South African grown *Eucalyptus grandis* trees of different growth stress intensity. *Holzforschung* 41:331–335
- Mhoswa L, O'Neill MM, Mphahlele MM, et al (2020) A genome-wide association study For resistance to the insect pest *Leptocye invasa* In *Eucalyptus grandis* reveals genomic regions and positional candidate defence genes. *Plant Cell Physiol*
- Miao C, Yang J, Schnable JC (2019) Optimising the identification of causal variants across varying genetic architectures in crops. *Plant Biotechnol J* 17:893–905
- Mizrachi E, Verbeke L, Christie N et al (2017) Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proc Natl Acad Sci* 114(5):1195–1200
- Mphahlele MM, Isik F, Mostert-O'Neill MM et al (2020) Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. *Tree Genet Genomes* 16:1–12
- Müller BSF, Neves LG, de Almeida Filho JE 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
- Müller BSF, de Almeida Filho JE, Lima BM et al (2019) Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *New Phytol* 221:235–254
- Munoz F, Rodriguez LS (2014) breedR: statistical methods for forest genetic resources analysis. In: *Trees for the future: plant material in a changing climate*. pp 13–p
- Myburg AA, Grattapaglia D, Tuskan GA et al (2014) The genome of *Eucalyptus grandis*. *Nature* 510:356–362
- Oladzad A, Porch T, Rosas JC et al (2019) Single and multi-trait GWAS identify genetic factors associated with production traits in common bean under abiotic stress environments. *G3 Genes, Genomes, Genet* 9:1881–1892
- Osoerio LF, White TL, Huber DA (2001) Age trends of heritabilities and genotype-by-environment interactions for growth traits and wood density from clonal trials of *Eucalyptus grandis* Hill ex Maiden. *Silvae Genet* 50:108–116
- Paaby AB, Rockman MV (2013) The many faces of pleiotropy. *Trends Genet* 29:66–73
- Peterson RA (2021) Finding optimal normalizing transformations via bestNormalize. R J
- Porter HF, O'Reilly PF (2017) Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci Rep* 7:1–12
- Rambolarimanana T, Ramamonjisoa L, Verhaegen D et al (2018) Performance of multi-trait genomic selection for *Eucalyptus robusta* breeding program. *Tree Genet Genomes* 14:1–13
- Resende RT, Resende MDV, Silva FF et al (2017) Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. *New Phytol* 213:1287–1300
- Rodriguez M, Scintu A, Posadinu CM et al (2020) GWAS based on RNA-Seq SNPs and high-throughput phenotyping combined with climatic data highlights the reservoir of valuable genetic diversity in regional tomato landraces. *Genes (basel)* 11:1387
- Rubin EM (2008) Genomics of cellulosic biofuels. *Nature* 454:841–845
- Schimleck LR, Kube PD, Raymond CA (2004) Genetic improvement of kraft pulp yield in *Eucalyptus nitens* using cellulose content determined by near infrared spectroscopy. *Can J for Res* 34:2363–2370
- Schumacher FX (1933) Logarithmic expression of timber-tree volume. *J Agric Res* 47:719–734
- Shi Z, Zhang Y, Phillips GO, Yang G (2014) Utilization of bacterial cellulose in food. *Food Hydrocoll* 35:539–545
- Silva-Junior OB, Faria DA, 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 Phytol* 206:1527–1540
- Solovieff N, Cotsapas C, Lee PH et al (2013) Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* 14:483–495
- Stackpole DJ, Vaillancourt RE, de Aguiar M, Potts BM (2010) Age trends in genetic parameters for growth and wood density in *Eucalyptus globulus*. *Tree Genet Genomes* 6:179–193
- Stackpole DJ, Vaillancourt RE, Alves A et al (2011) Genetic variation in the chemical components of *Eucalyptus globulus* wood. *G3 Genes Genomes Genet* 1:151–159
- Tan B, Ingvarsson PK (2018) Multivariate genome-wide association identify loci for complex growth traits by considering additive and over-dominance effects in hybrid *Eucalyptus*
- Tang Y, Liu X, Wang J, et al (2016) GAPIT version 2: an enhanced integrated tool for genomic association and prediction. *Plant Genome* 9:plantgenome2015–11
- Thavamanikumar S, McManus LJ, Ades PK, Bossinger G et al (2014) Association mapping for wood quality and growth traits in *Eucalyptus globulus* ssp *globulus* Labill identifies nine stable marker-trait associations for seven traits. *Tree Genet Genomes* 10(6):1661–1678
- Toenio MPM, Davila Olivias NH, Kloth KJ et al (2017) Genetic architecture of plant stress resistance: multi-trait genome-wide association mapping. *New Phytol* 213:1346–1362
- Valenzuela CE, Ballesta P, Ahmar S et al (2021) Haplotype-and SNP-based GWAS for growth and wood quality traits in *Eucalyptus cladocalyx* trees under arid conditions. *Plants* 10:148
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423
- Waples RS, Do CHI (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Mol Ecol Resour* 8:753–756
- Ward BP, Brown-Guedira G, Kolb FL et al (2019) Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. *PLoS ONE* 14:e0208217
- Wickham H (2011) ggplot2. *Wiley Interdiscip Rev Comput Stat* 3:180–185
- Yin L (2018) CMplot: circle manhattan plot. <https://cran.r-project.org/package=CMplot>
- Yoshida GM, Yáñez JM (2021) Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics* 22:1–13
- Zheng X, Levine D, Shen J et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



## Authors and Affiliations

Lucas Fernandes Rocha<sup>1</sup> · Thiago Romanos Benatti<sup>2</sup> · Leandro de Siqueira<sup>2</sup> · Izabel Christina Gava de Souza<sup>2</sup> · Isadora Bianchin<sup>2</sup> · Aguinaldo José de Souza<sup>2</sup> · Aline Cristina Miranda Fernandes<sup>2</sup> · Shinitiro Oda<sup>2</sup> · José Luiz Stape<sup>2</sup> · Rafael Massahiro Yassue<sup>3</sup> · Humberto Fanelli Carvalho<sup>4</sup> · Niels A. Müller<sup>5</sup> · Matthias Fladung<sup>5</sup> · Juan Jose Acosta<sup>6</sup> · Roberto Fritsche-Neto<sup>7</sup> · Evandro Vagner Tambarussi<sup>1,8,9</sup> 

Lucas Fernandes Rocha  
lucasrochamg@gmail.com

Thiago Romanos Benatti  
tbenatti@suzano.com.br

Leandro de Siqueira  
lsiqueira@suzano.com.br

Izabel Christina Gava de Souza  
izabelsouza@suzano.com.br

Isadora Bianchin  
ibianchin@suzano.com.br

Aguinaldo José de Souza  
ajsouza@suzano.com.br

Aline Cristina Miranda Fernandes  
alinef@suzano.com.br

Shinitiro Oda  
shinitiro.oda@gmail.com

José Luiz Stape  
jlstape@gmail.com

Rafael Massahiro Yassue  
rafael.yassue@usp.br

Humberto Fanelli Carvalho  
humberto.fanelli@gmail.com

Niels A. Müller  
niels.mueller@thuenen.de

Matthias Fladung  
matthias.fladung@thuenen.de

Juan Jose Acosta  
jjacosta@ncsu.edu

Roberto Fritsche-Neto  
rfneto@agcenter.lsu.edu

<sup>1</sup> Department of Forest Science, FCA/UNESP, São Paulo State University, Botucatu, Sao Paulo, Brazil

<sup>2</sup> SUZANO S.A, Salvador, Bahia, Brazil

<sup>3</sup> “Luiz de Queiroz” College of Agriculture, Department of Genetics, (ESALQ/USP), University of São Paulo, Piracicaba, Brazil

<sup>4</sup> Universidad Politécnica de Madrid, 28040 Madrid, Spain

<sup>5</sup> Thünen Institute of Forest Genetics, 22927 Grosshansdorf, Germany

<sup>6</sup> North Carolina State University, NCSU, Raleigh, USA

<sup>7</sup> H. Rouse Caffey Rice Research Station - LSUAgCenter, Baton Rouge, LA, USA

<sup>8</sup> Department of Forest Engineering, State University of the Central-West, Irati, PR, Brazil

<sup>9</sup> “Luiz de Queiroz” College of Agriculture, Department of Forest Sciences, University of Sao Paulo, Sao Paulo, Piracicaba, Brazil