



## OPEN ACCESS

## EDITED BY

Shixiao Yu,  
Sun Yat-sen University, China

## REVIEWED BY

Paulino Pérez-Rodríguez,  
Colegio de Postgraduados (COLPOS),  
Mexico  
Sanushka Naidoo,  
University of Pretoria, South Africa

## \*CORRESPONDENCE

Gustavo Balmelli

✉ [gbalmelli@inia.org.uy](mailto:gbalmelli@inia.org.uy)

RECEIVED 15 December 2022

ACCEPTED 24 May 2023

PUBLISHED 03 July 2023

## CITATION

Quezada M, Giorello FM, Da Silva CC,  
Aguilar I and Balmelli G (2023) Single-step  
genome-wide association study for  
susceptibility to *Teratosphaeria nubilosa*  
and precocity of vegetative phase  
change in *Eucalyptus globulus*.  
*Front. Plant Sci.* 14:1124768.  
doi: 10.3389/fpls.2023.1124768

## COPYRIGHT

© 2023 Quezada, Giorello, Da Silva, Aguilar  
and Balmelli. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Single-step genome-wide association study for susceptibility to *Teratosphaeria nubilosa* and precocity of vegetative phase change in *Eucalyptus globulus*

Marianella Quezada<sup>1,2</sup>, Facundo Matias Giorello<sup>3</sup>,  
Cecilia Corina Da Silva<sup>3</sup>, Ignacio Aguilar<sup>1</sup> and Gustavo Balmelli<sup>4\*</sup>

<sup>1</sup>Programa Nacional de Investigación en Producción de Leche, Estación Experimental "Wilson Ferreira Adulnate", Instituto Nacional de Investigación Agropecuaria, Canelones, Uruguay, <sup>2</sup>Laboratorio de Biotecnología, Departamento de Biología Vegetal, Facultad de Agronomía, Universidad de la República, Montevideo, Uruguay, <sup>3</sup>PDU Espacio de Biología Vegetal del Noreste, sede Tacuarembó, CENUR Noreste, Universidad de la República, Tacuarembó, Uruguay, <sup>4</sup>Programa Nacional de Investigación en Producción Forestal, Estación Experimental del Norte, Instituto Nacional de Investigación Agropecuaria, Tacuarembó, Uruguay

**Introduction:** *Mycosphaerella* leaf disease (MLD) is one of the most prevalent foliar diseases of *Eucalyptus globulus* plantations around the world. Since resistance management strategies have not been effective in commercial plantations, breeding to develop more resistant genotypes is the most promising strategy. Available genomic information can be used to detect genomic regions associated with resistance to MLD, which could significantly speed up the process of genetic improvement.

**Methods:** We investigated the genetic basis of MLD resistance in a breeding population of *E. globulus* which was genotyped with the EUChip60K SNP array. Resistance to MLD was evaluated through resistance of the juvenile foliage, as defoliation and leaf spot severity, and through precocity of change to resistant adult foliage. Genome-wide association studies (GWAS) were carried out applying four Single-SNP models, a Genomic Best Linear Unbiased Prediction (GBLUP-GWAS) approach, and a Single-step genome-wide association study (ssGWAS).

**Results:** The Single-SNP (model K) and GBLUP-GWAS models detected 13 and 16 SNP-trait associations in chromosomes 2, 3 y 11; whereas the ssGWAS detected 66 SNP-trait associations in the same chromosomes, and additional significant SNP-trait associations in chromosomes 5 to 9 for the precocity of phase change (proportion of adult foliage). For this trait, the two main regions in chromosomes 3 and 11 were identified for the three approaches. The SNPs identified in these

regions were positioned near the key *miRNA* genes, *miR156.5* and *miR157.4*, which have a main role in the regulation of the timing of vegetative change, and also in the response to environmental stresses in plants.

**Discussion:** Our results demonstrated that ssGWAS was more powerful in detecting regions that affect resistance than conventional GWAS approaches. Additionally, the results suggest a polygenic genetic architecture for the heteroblastic transition in *E. globulus* and identified useful SNP markers for the development of marker-assisted selection strategies for resistance to MLD.

#### KEYWORDS

**eucalyptus, genome-wide association study (GWAS), single-step genome-wide association study (ssGWAS), leaf disease, heteroblastic transition**

## 1 Introduction

*Eucalyptus globulus* Labill. is widely planted in temperate countries mainly for the excellent quality of its wood for pulp production, which has a high market value in the paper industry (Tibbits et al., 1997). However, the susceptibility of this species to several pests and diseases limits its use in commercial plantations (Simeto et al., 2020). *Mycosphaerella* leaf disease (MLD), caused by a complex of fungal species of the genus *Mycosphaerella* and *Teratosphaeria*, is one of the most prevalent foliar diseases of *E. globulus* in natural forests and plantations worldwide (Tibbits et al., 1997). *Teratosphaeria nubilosa* is considered one of the most virulent MLD species (Mohammed et al., 2003; Hunter et al., 2009). This pathogen infects predominantly juveniles and intermediate foliage and causes severe leaf spotting, premature defoliation and shoot blight (Carnegie and Ades, 2003). Crown damage from MLD in young plantations can range from <10% to 80% which reduces the tree growth and survival, leading to concomitant loss in productivity of affected plantations (Balmelli et al., 2013; Milgate et al., 2005).

A number of silvicultural practices and management strategies have been proposed to minimize the effects of foliar diseases in *Eucalyptus* plantations. These include avoiding planting on high-risk endemic areas, the application of protectants and fungicides, as well using remedial fertilizer applications. However, these strategies have economic, environmental or operational constraints (Simeto et al., 2020). Thus, breeding for resistance to develop more resistant genotypes is the most promising strategy to effectively control MLD in commercial plantations (Milgate et al., 2005; Hunter et al., 2009; Balmelli et al., 2013). Genetic variation on *E. globulus* for susceptibility to MLD has been identified between and within families, provenances and genetics groups (Costa e Silva et al., 2013; Balmelli et al., 2014). Considering that this species is markedly heteroblastic, significant genetic variation in the timing of vegetative phase change has also been reported (Balmelli et al., 2013). As young foliage is particularly susceptible to *T. nubilosa* infection, at least two mechanisms has been proposed to manage MLD disease: increase the resistance of juvenile foliage and increase the precocity

change to the resistant adult foliage (Milgate et al., 2005). Several studies suggest that both mechanisms involved in foliar resistance are under moderate to strong genetic control in *E. globulus* (Smith et al., 2017). But the underlying molecular mechanism and genetic architecture of MLD response has been relatively little investigated (Freeman et al., 2008; Hudson et al., 2014).

Genome-wide association studies (GWAS) are a leading approach for complex trait dissection and identification of genomic segments or alleles that underlie phenotypic variation and thus can be used for breeding (Tibbs Cortes et al., 2021). A common GWAS method in plants is the unified mixed linear model (MLM) proposed by Yu et al. (2006), that accounts for relatedness at two levels: population structure and kinship. Since this method was computationally very intensive, a reparameterization of the MLM likelihood function was proposed (Kang et al., 2010). Despite the statistical improvements, a typical feature of these GWAS methods is sequentially fitting each marker one at a time as fixed effects, resulting in a lack of power to map loci for quantitative traits (Wang et al., 2012). This limitation has been overcome by methods that simultaneously fit high density genome-wide molecular markers by mixed linear models as GBLUP (Genomic Best Linear Unbiased Prediction). Following this strategy, marker's effects are estimated by a linear transformation of genomic breeding values obtained from GBLUP (Gualdrón Duarte et al., 2014). One limitation of this method is that it only includes phenotypes of genotyped individuals. Generally in forest breeding populations, as in crops and livestock populations, only a fraction of individuals in a population are genotyped. Thus, the GBLUP methods were extended to include the information of non-genotyped individuals in prediction analysis, which is the so-called Single-step genomic best linear unbiased prediction (ssGBLUP) (Aguilar et al., 2010). The ssGBLUP approach is widely adopted for genomic prediction in livestock (Misztal et al., 2021) and recently applied in forest species (Cappa et al., 2019; Quezada et al., 2022). The strength of this approach is the ability to jointly incorporate the information of all genotypes, observed phenotypes and pedigree information in one simple and single step model. This procedure was extended to estimate the marker's effects in an approach called Single-step

GWAS (ssGWAS) (Wang et al., 2012). Applying this method, marker effects and the p-values associated are estimated simultaneously for all markers while accounting for population structure using the pedigree and genomic relationship for all individuals in the population (Aguilar et al., 2019).

The advances in next generation sequencing technologies allows the development of accessible high throughput genotyping systems. Particularly for *Eucalyptus* species, in which marker number and density have been the limiting factors for a long time, dense sets of Single Nucleotide Polymorphism (SNP) genotyping platforms are now available (Silva-Junior et al., 2015). The commercial *Eucalyptus* chip has been useful to understand the genetic basis of complex quantitative traits of interest, such as growth or wood properties in *Eucalyptus* species (Du et al., 2018). For example, in *E. grandis* × *E. urophylla* breeding populations, eight SNPs were described as main associations for growth traits, which were related to genes involved in cell wall biosynthesis (Müller et al., 2019). Also, it has become possible to detect significant marker-trait associations for growth related traits in species poorly represented in the chip. Thus, a total of 87 SNPs were associated with growth and wood quality traits in *E. cladocalyx*, revealing associations with genes related to primary metabolism and biosynthesis of cell wall components (Valenzuela et al., 2021). Similarly, a study of *E. grandis* × *E. urophylla* hybrids detected 22 quantitative trait loci for growth and wood traits and four for *Puccinia psidii* rust disease resistance (Resende et al., 2017). This resistance QTLs were positioned near the major QTLs detected in *E. grandis* for Myrtle rust (*Ppr1* locus). Recently, for the same fungal disease, 33 highly significant SNPs were detected in *E. obliqua*, identifying candidate defense response genes, one of them located within the *Ppr1* locus (Yong et al., 2021). With the exception of Myrtle rust, GWAS analysis to investigate the genetic control and identify candidate defense genes for fungal diseases in *Eucalyptus* have not been implemented. For MLD resistance, a classical QTLs mapping using biparental crosses and 165 molecular markers detected two major QTLs that explained a large proportion of the phenotypic variance for *Mycosphaerella cryptica* severity (Freeman et al., 2008).

This work aimed to examine the genetic architecture of resistance to MLD caused by *T. nubilosa* in a breeding population of *E. globulus*. Although the Single-step GWAS strategy has received increasing attention in livestock species (Misztal et al., 2021), to the best of our knowledge, it has not yet been applied in forest tree species. Therefore, this study has two specific objectives: i) to compare the Single-SNP associations GWAS, the GBLUP-GWAS and the Single-step GWAS strategies in a forest tree species, ii) to elucidate the genetic base of resistance and escape to *T. nubilosa* in *E. globulus*.

## 2 Material and methods

### 2.1 Population and phenotypic data

The population and phenotypic data used in this study have been described in detail in Balmelli et al., (2013). Briefly, a progeny test of *E. globulus* was established by Instituto Nacional de

Investigación Agropecuaria (INIA) of Uruguay, in Lavalleja (Lat. 34° 110 S; Long. 54° 540 W; Alt. 206 m). The trial contain 194 open-pollinated families with a total of 4601 individuals. The experimental design was a randomized complete block design, with 3 replicates and eight-tree row plots. When the trees were ~ 1 year old, there were several consecutive days of rain and high humidity, causing a severe infection of MLD (Balmelli et al., 2016). Phenotypic responses to the natural infection were assessed at age 14 (coinciding with outbreak of MLD), 21 and 26 months. The response to disease was assessed using two parameters, the severity of leaf spots (SEV) and defoliation (DEF). Both were evaluated on the whole crown (juvenile and adult foliage) using a visual scale, recording the percentage of leaf area affected by spots (SEV) and the percentage of leaves prematurely abscised (DEF). The escape to disease was assessed as the precocity of vegetative phase change (ADFO), quantified as the proportion of the crown with adult foliage (petiolate, alternate, shiny green and pendulous leaves). The SEV was classified in percentages classes of 5%, whereas DEF and ADFO in intervals of 10%.

### 2.2 Genotyping

A total of 1008 trees were sampled for genotyping. DNA was extracted from leaf tissue of each tree using a standard CTAB protocol. Genotyping was performed using the *Eucalyptus* Illumina EUChip60K (Silva-Junior et al., 2015) by GeneSeek (Lincoln, NE, USA). Of the 194 families in the trial, 179 have genotyped trees ranging from 4 to 8 trees per family. Individual genotypes were filtered by excluding samples with more than 10% of missing data across SNPs. The SNP data were then filtered by call rate > 90% and a minor allele frequency (MAF) > 0.01. The quality control of genotyping data was performed using QCF90 software (Masuda et al., 2019).

### 2.3 Population structure and linkage disequilibrium

The genetic structure of the breeding population was assessed by the Principal Components Analysis (PCA) using PREGSF90 of BLUPF90 family of programs (Aguilar et al., 2018). Pairwise-estimates of linkage disequilibrium (LD) were calculated as the squared correlation of allele counts for each pair of two SNPs ( $r^2$ ) within each of the 11 chromosome. The LD decay was estimated using the R script by Marroni et al. (2011). The pattern of LD decay was visualized plotting the  $r^2$  against the physical distance with ggplot (Wickham, 2016) R package (R Core Team, 2022). The threshold of LD decay was identified at  $r^2 = 0.02$ .

### 2.4 Statistical analyses

#### 2.4.1 Mixed model analysis

Linear mixed models were used to estimate the variance components of proportion of adult foliage, leaf spot severity and

defoliation, considering the observed phenotypes and transformed data into normal score within each block (Cappa and Varona, 2013). The linear mixed model was:

$$y = X\beta + Wp + Za + e \tag{1}$$

where  $y$  is the vector of phenotypes;  $\beta$  is the vector of fixed effects including the overall mean and block, with incidence matrix  $X$ ;  $p \sim N(0, I\sigma_p^2)$  is the vector of random plot effects, with design matrix  $W$ ;  $a \sim N(0, A\sigma_a^2)$  is the vector of random effects of individual trees (i.e., breeding values), with incidence matrix  $Z$ ; and  $e \sim N(0, I\sigma_e^2)$  is the vector of residual errors. The matrix  $A$  is the pedigree relationship matrix and  $\sigma_a^2$  is the additive genetic variance. Genomic data were not included for variance components estimation.

Heritability of each trait was estimated as:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_p^2 + \sigma_e^2} \tag{2}$$

where  $\sigma_a^2$  is the additive genetic variance,  $\sigma_p^2$  is the plot variance, and  $\sigma_e^2$  is the residual variance.

### 2.4.2 Single-SNP models

The most common single-marker models were compared, with different combinations of population stratification and genetic relatedness correction. First, a naive model, without any correction for population structure or relatedness was fitted independently for each SNP following the linear mixed model:

$$y = X\beta + x_i b_i + Wp + e \tag{3}$$

where  $y$  is the vector of phenotypes;  $\beta$  is the vector of fixed effects including the overall mean and block, with incidence matrix  $X$ ;  $p \sim N(0, I\sigma_p^2)$  is the vector of random plot effects, with design matrix  $W$ ; and  $e \sim N(0, I\sigma_e^2)$  is the vector of residual errors. The  $x_i$  is a vector that contains the genotype for the  $i^{th}$  SNP for each individual and  $b_i$  is the  $i^{th}$  SNP effect considered as fixed. A second model was fitted to account for population structure (model P), where the vector of fixed effects ( $\beta$ ) of equation 3 include the first three principal components identified.

Alternatively, to account for family structure, a linear mixed model including a polygenic effect was tested (model K). This association model was fitted using the same naive model by adding a random effect utilizing a genomic relationship matrix as follow:

$$y = X\beta + x_i b_i + Wp + Zu + e \tag{4}$$

where  $Z$  is a design matrix and  $u$  is the vector of polygene background effects (random),  $u \sim N(0, G\sigma_u^2)$ . The rest of the components were previously defined (equation 3). The  $G$  matrix was estimated using the first method of VanRaden (2008):

$$G = \frac{ZZ'}{2\sum p_i(1 - p_i)} \tag{5}$$

where  $Z$  is the matrix of gene content adjusted for observed allele frequencies and  $p_i$  is the allele frequency of the  $i^{th}$  SNP. To account for family and population structure simultaneously, a model K + P was fitted, where the vector of fixed effects ( $\beta$ ) of

equation 4 include the first three principal components. All Single-SNP models were fitted using the Sommer R package (Covarrubias-Pazarán, 2016).

### 2.4.3 GBLUP-GWAS model

For the GBLUP-GWAS model, the follow mixed model was used:

$$y = X\beta + Wp + Za + e \tag{6}$$

where every element is defined as before (equation 1), except for the vector of random effects of the individual trees ( $a$ ) that follow a  $var(a) \sim N(0, G\sigma_a^2)$ . The matrix  $G$  is the genomic relationship matrix calculated using all the SNPs (equation 5).

Under this GBLUP-GWAS model, the vector of SNP effects was obtained from a linear transformation of the vector of breeding values in  $a$ . The vector of allele marker effects ( $\hat{u}$ ) was calculated for all SNPs simultaneously following Gualdrón Duarte et al. (2014).

$$\hat{u} = Z' \frac{1}{2\sum p_j(1 - p_j)} G^{-1} \hat{a} \tag{7}$$

The variance of SNP effects and the p-values of each SNP effect were estimated following the formulas of Gualdrón Duarte et al. (2014).

### 2.4.4 Single-step GBLUP association model

The ssGWAS model is similar to the GBLUP-GWAS previously presented, but in this model all the phenotype and pedigree information of all assessed trees (genotyped and non-genotyped) was considered. Thus, the  $G$  matrix is replaced with the  $H$  matrix that combines both pedigree and genomic information.

The inverse of the relationship matrix that combines pedigree and genomic information ( $H^{-1}$ ) was derived by Aguilar et al. (2010) as:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \tag{8}$$

where  $A^{-1}$  and  $A_{22}^{-1}$  are the inverse of the pedigree relationship matrices for all individuals and only for the genotyped individuals, respectively, and  $G^{-1}$  is the inverse of the genomic relationship matrix.

The SNP effects were estimated as

$$\hat{u} = Z' \frac{1}{2\sum p_j(1 - p_j)} G^{-1} \hat{a}_{22} \tag{9}$$

where  $\hat{a}_{22}$  is a vector of genomic breeding values only for genotyped individuals. The variance explained for each marker and the p-values were obtained as suggested by Aguilar et al. (2019).

The GLUP-GWAS and ssGWAS models were fitted with POSTGF90 of BLUPF90 family software (Aguilar et al., 2014). The GWAS results were plotted with CMPLLOT R package (Yin et al., 2021).

### 2.4.5 Model comparison

Correction for multiple testing was applied to determine a threshold to indicate significant associations for all the six models

implemented. At a genome-wide level, the Bonferroni correction at an alpha level of 0.05 was applied as a conservative criterion. In addition, a less stringent threshold for declaring significance was applied using the false discovery rate (FDR), with an alpha level of 0.05. Quantile-quantile (Q-Q) plots of observed and expected p-values were used to evaluate the control of population structure and genetic relationships in the different GWAS models. These plots were generated with QQMAN R package (Turner, 2018).

## 2.5 Candidate genes and functional annotation

The physical position of the SNPs was defined according to the reference genome of *E. grandis* (Phytozome Version 1.1 available on <https://phytozome-next.jgi.doe.gov/>) since the genome of this species is thoroughly annotated. Candidate genes were searched within 30 kb upstream and downstream of the significant SNPs. The size of this interval was chosen based on the typical LD present in the *Eucalyptus* populations and the SNPs density in our study (1 SNP per 28kb). The sequence of the candidate genes were extracted from the annotation file of *E. grandis* v1.1. Particular miRNA related to heteroblasty, and their targets, were searched based on the annotation of Hudson et al. (2014) using blastn-short (BLASTN program optimized for sequences shorter than 50 bases). Functional annotation of genes associated with significant SNPs was done by protein homology search against the NCBI non-redundant protein sequences (NR) database (Sayers et al., 2022) using BLAST (Altschul et al., 1990). When possible, the best hit with functional annotation was used.

## 3 Results

### 3.1 Phenotypic variation

In this trial, the epidemic of *T. nubilosa* affected all the trees, which showed high variability for all the diseased-related traits. The mean and standard deviation for all the traits increased along with the increasing age of measurement, indicating a high prevalence of infection (Table 1). Though a wide range of variability was observed for the proportion of adult foliage, a high proportion of trees had only juvenile foliage, resulting in a distribution with an inflated

proportion of zero counts in the three measurement ages. The disease damage traits (SEV and DEF) presented low but significant variability, following a normal distribution (Figure 1). Phenotypic correlations between measurement ages were positive and high for the proportion of adult foliage, and positive and moderate for defoliation. The disease damage traits, presented moderately positive correlation. The vegetative change to adult foliage (ADFO) was negatively correlated with response to diseased traits, confirming the mechanism of disease escape through early change to resistant adult foliage (Figure 1). The estimated pedigree-based heritabilities were moderate to high, varying from 0.33 for DEF\_21 to 0.77 for ADFO\_26 (Table 1).

### 3.2 Population structure and linkage disequilibrium

A total of 19,992 markers and 961 trees were retained after quality filtering. The SNP markers were distributed throughout the 11 chromosomes according to the *E. grandis* reference genome, with a range from 1,268 in chromosome 4 to 2,516 in chromosome 8 (Figure 2). The trend of linkage disequilibrium (LD) in the *E. globulus* population was analyzed across each chromosome. The average genome-wide  $r^2$  in this population was 0.032, with a LD decay at 19.1 kb using the significant threshold ( $r^2 = 0.2$ ) (Figure 3A). The LD decay varied across different chromosomes, with the most rapid LD decay observed for chromosome 5 (12.4 kb) compared with the slower rate on chromosome 9 (50.1 kb)(data not shown). The principal component analysis (PCA) showed the diversity and population structure present in the breeding population. The first two principal components explained 6.22% and 1.53% of the total variation, respectively. Based on these components, the majority of trees cluster into a major group, and different sub clusters can be identified (Figure 3B). Only the three first components that accounted for genetic variances greater than 1% were included in the association analysis to correct for population stratification.

### 3.3 Association analysis

#### 3.3.1 Single-SNP models

The Single-SNP model without taking into account the population structure (naive model) resulted in the detection of a

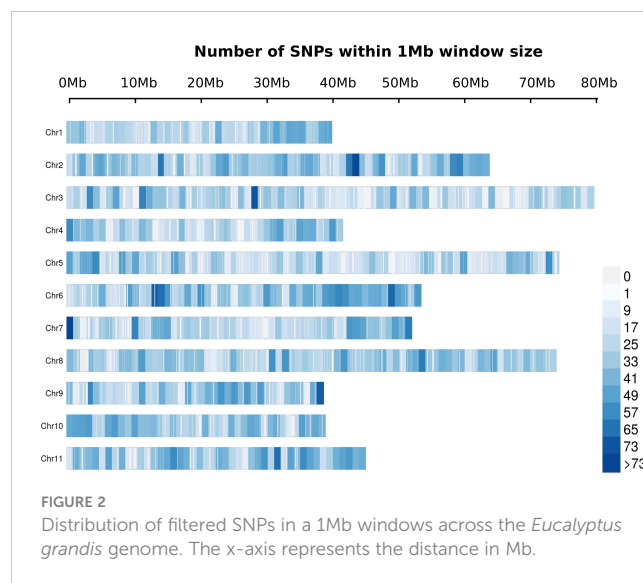
TABLE 1 Summary of assessed traits.

Trait	Age of measurement <sup>a</sup>	Abbreviation	No. Obs	Mean (sd)	Range	$h^2$
Proportion of adult foliage	14	ADFO_14	3,853	9.76 (15.78)	0-90	0.68
	21	ADFO_21	3,580	16.99 (19.18)	0-90	0.66
	26	ADFO_26	3,447	29.14 (20.38)	0-90	0.77
Severity of leaf spots	14	SEV_14	3,711	10.53 (3.79)	0-40	0.71
Defoliation	14	DEF_14	3,711	31.62 (9.44)	0-70	0.49
	21	DEF_21	3,581	52.73 (10.86)	10-90	0.33

<sup>a</sup> in months.

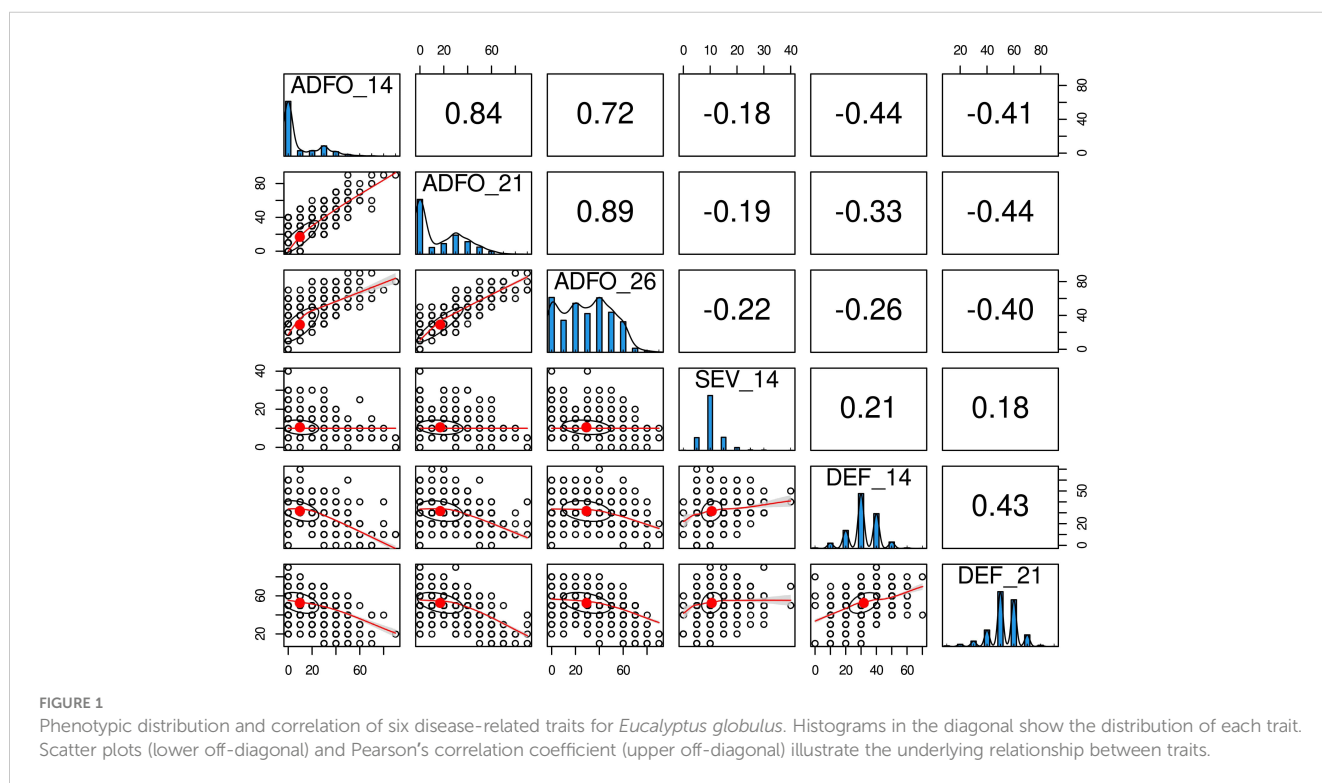
large number of associations for all traits (Table 2). Most of these were apparent false positives or spurious associations, by ignoring the multiple levels of relatedness between individuals. To control the population structure, the three first principal components were included in the model (model P). Although the number of associations slightly decreased, a high number of spurious signals were detected, due to family relatedness not being accounted for. The quantile-quantile plot (QQ-plot) obtained for the naive and model P, showed a great deviation for the observed and expected p-values, reflecting the inadequacy of these models (Supplementary Figure 1).

When the random effects captured for the G matrix were included in the association models (models K and K+P), more p-values follow the uniform diagonal line in the QQ-plots (Supplementary Figure 1). Compared with the naive and model P, these models properly control false positive associations resulting from the within population family structure. As a result, including the G matrix resulted in a drastic reduction in the number of significant SNP associations. No significant associations were detected for SEV and DEF after correcting for multiple testing (Bonferroni and FDR at 5%) (Table 2). All the significant associations were detected for ADFO at the three measurement age. We found a total of 13 SNP-trait associations (8 unique SNPs) for the K model and 11 associations (6 unique SNPs) for the K + P model considering a FDR threshold of 5%, and 6 were also significant after the more stringent Bonferroni correction (5%), in both models. Moreover, both models did not show any difference, because all the SNPs detected in the K + P model were also significant in the K model. Thus, for this breeding population, the



genetic matrix can account for genetic structure, with a negligible effect of the adjustment of principal components.

For ADFO, the significant SNPs were distributed in chromosomes 2, 3 and 11 (Supplementary Figure 2). The higher number of significant SNPs associated was detected at the first age of measurement (14 months). It is not unexpected that the four and unique SNPs detected as significantly associated at the 21 and 26 months respectively, were also significant associated in the previous age measurement. Similar results were obtained using the observed



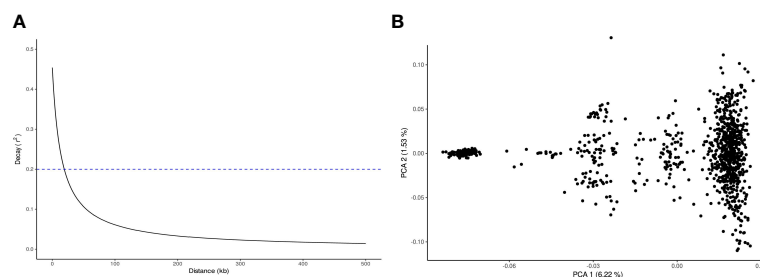


FIGURE 3

Population structure and linkage disequilibrium (LD) decay for the *Eucalyptus globulus* breeding population. (A) Linkage disequilibrium (LD) decay estimated by  $r^2$  (y-axis) plotted against physical distance in Kb (x-axis). Dashed line at  $r^2 = 0.2$  indicates the frequently used threshold of usable LD. (B) Principal Component Analysis (PCA) of 961 individual of the breeding population. PCA plot defined by first and second eigenvectors.

variable and the transformed variable to normal scores for all Single-SNP models tested, as well as for GBLUP-GWAS and ssGWAS models (results not shown).

### 3.3.2 GBLUP-GWAS model

The GBLUP-GWAS model was performed for each variable to evaluate the equivalence with the Single-SNP model that accounts for genetic relationships in the  $G$  matrix (model K). For disease damage traits (SEV and DEF), no significant associations were declared using the multiple test correction (Bonferroni and FDR at 5%) (Table 2). Conversely, for ADFO the GBLUP-GWAS model identified 16 SNP-trait associations (8 unique SNPs) at the three measurement ages at the most permissive level of FDR at 5%. Half of these associations were considered significant after Bonferroni correction (5%).

Of these 16 SNP-trait associations, 12 were previously identified by the Single-SNP model K and four were novel SNPs. The seven SNPs associated with the proportion of adult foliage at 14 months were also identified in the Single-SNP approach (model K) at the same measurement age (Figure 4). Similarly, all the SNPs associated at 21 and 26 months for the Single-SNP model K were detected in the GBLUP-GWAS model.

Associations for the proportion of adult foliage at 14 months were detected on chromosomes 2 (1 SNP), 3 (3 SNPs) and 11 (3 SNPs) (Supplementary Figure 3). The same regions in chromosomes 2, 3 and 11 were detected at 21 months, with six SNPs detected in both measurement ages. Two SNPs, one in

chromosome 3 and the other in chromosome 11, were found associated with ADFO at 26 months. These SNPs were associated with this escape to disease trait in the three measurement ages.

### 3.3.3 ssGWAS model

The number of significant SNP markers detected in the Single-step GBLUP association approach for response and escape to disease traits was larger than in the Single-SNP (model K and K + P) or GBLUP methods, which indicate a better performance of ssGWAS model (Table 2). Considering the FDR multiple test correction at 5%, we detected 66 SNP-trait associations and 38 unique SNPs for ADFO by the three measurement age. This approach identified 12 significant SNP associations for DEF at 21 months (Supplementary Table 1; Supplementary Figure 4). After the more conservative Bonferroni correction at 5%, 19 and one associations continue to be significant with ADFO and DEF traits, respectively. Once again, we found no significant association for SEV trait.

For ADFO, significant SNPs were associated across all chromosomes, except for chromosomes 1, 4 and 10. However, the most significant SNPs were identified in chromosomes 3 and 11 in the three measurement ages, indicating the presence of few genomic regions controlling the precocity of adult foliage change (Figure 5).

The first peak identified on chromosome 3 spans a region of ~ 2 Mb, defined in accordance with SNPs overlapping at the three measurements ages. The high number of significant SNPs was associated with the first age of measurement (14 months), namely

TABLE 2 Number of significant Single Nucleotide Polymorphism (SNP) associations for disease traits for all GWAS models evaluated.

Trait	Single-SNP models				GBLUP-GWAS	ssGWAS
	naive	model P	model K	model K + P		
ADFO_14	3,372 (260)	3,282 (244)	8 (3)	6 (3)	7 (5)	26 (8)
ADFO_21	3,139 (193)	3,081 (190)	4 (2)	4 (2)	7 (2)	25 (9)
ADFO_26	2,837 (165)	2,797 (158)	1 (1)	1 (1)	2 (1)	15 (2)
SEV_14	104 (6)	97 (4)	0 (0)	0 (0)	0 (0)	0 (0)
DEF_14	227 (2)	222 (2)	0 (0)	0 (0)	0 (0)	0 (0)
DEF_21	372 (25)	390 (29)	0 (0)	0 (0)	0 (0)	12 (1)

Significant associations using FDR ( $\alpha = 0.05$ ) correction. In parentheses number of significant associations using Bonferroni correction ( $\alpha = 0.05$ ).

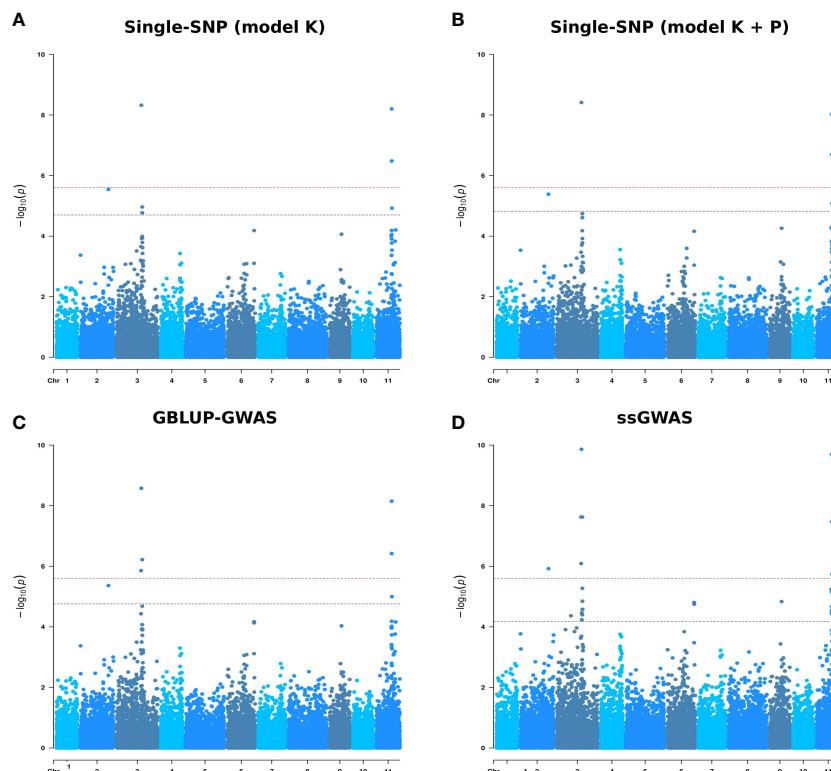


FIGURE 4

Manhattan plots for proportion of adult foliage measured at 14 months (ADFO\_14) using four genome-wide association models. (A) Single-SNP model adjusted for kinship matrix (model K). (B) Single-SNP model adjusted for kinship matrix and population structure (model K + P). (C) GBLUP-GWAS model. (D) Single-step GBLUP association model. The x-axis represents SNP positions on the 11 *Eucalyptus grandis* chromosomes and the y-axis  $-\log_{10}(p)$  values from genotypic associations. The red horizontal line indicates the Bonferroni threshold ( $\alpha = 0.05$ ) and the blue horizontal line indicates a false discovery rate (FDR) at 5% threshold.

nine SNPs located in this region, the most significant with a  $-\log_{10}(p)$  value of 9.86 (position 47 949 673). Inside this area, the six SNPs significantly associated at 21 months, were also detected at the measurement age of 14 months. Additionally, the five SNPs found at the 26 months, were also detected in the two previous measurement ages. The SNPs detected in this region explained the 1.14, 1.64 and 0.44 of the total genetic variance for each age at measurement of 14, 21 and 26 months, respectively.

A small region of  $\sim 200$  kb was also identified in chromosome 11 significantly associated with this trait. This region encompasses 7, 5 and 3 SNPs for measurement at 14, 21 and 26 months, respectively. Once again, the higher number of SNPs was identified at the first measurement age (14 months), and all the markers identified at 21 and 26 months were also detected in the previous measurement age. In this region, the SNP in the position 29 564 447 was the most significant, with  $-\log_{10}(p)$  values of 9.69, 11.07 and 9.36 for 14, 21 and 26 months, respectively. The SNPs inside this region represented 1.14, 0.94 and 0.73 of the genetic total variance at the three measurement ages, respectively.

Additionally, other significant signals were identified in chromosome 2, with one SNP (position 54 236 535) detected at the conservative threshold of Bonferroni correction at 14 and 21 months, and with the FDR threshold at 26 months. For chromosome 6, two SNPs located 100 kb apart were identified with the less stringent FDR correction for the three measurement ages (Figure 5).

The ssGWAS model was able to detect 12 SNPs in chromosomes 3, 6, 8 and 11 associated with the defoliation measured at 21 months. However, no significant associations were detected for the previous measurement age (14 months). As expected by the correlation between defoliation and the proportion of adult foliage traits, six common SNPs between these traits at the same age (21 months) were observed. Thus, two SNPs identified in chromosome 3 and four in chromosome 11, correspond to the significant regions detected for ADFO.

### 3.4 Candidate genes for precocity of vegetative phase change

Of the 38 unique significant SNPs identified for heteroblasty, two of them did not have any associated genes. A total of 133 genes were associated with the rest of the significant SNPs, of which 131 were protein coding. Two key *miRNA* genes, *miR156.5* and *miR157.4*, were found in chromosome 3 and 11, respectively (Table 3). The significant SNPs associated with these genes were at 21 kb and 25 kb for *miR156.5* and *miR157.4*, respectively. None of the *miR156* and *miR157* target proteins were found.

Out of the 131 protein coding genes, 20 could not be assigned a specific functional annotation. In total 95 different functions were found (Supplementary Table 2). A total of 11 proteins related to



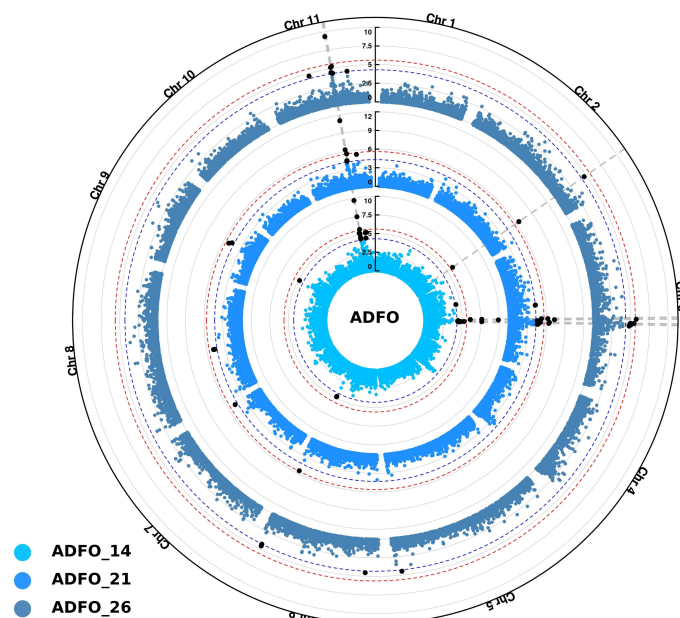


FIGURE 5

Circular Manhattan plots for proportion of adult foliage (ADFO) using Single-step GBLUP association model (ssGWAS). Inner, middle and outer layers represent the measurement ages at 14, 21 and 26 months, respectively. The Bonferroni ( $\alpha = 0.05$ ) and false discovery rate (FDR) threshold at 5% are represented by the red and blue circle in each layer, respectively. Vertical dashed lines highlight common SNPs identified among the three measurement ages. Significant SNPs identified after multiple test corrections are in black.

resistance to biotic stress encoded by 22 different genes were the most interesting ones: 26S proteasome non-ATPase regulatory subunit 9 (Eucgr.C02542), dirigent protein 22 (Eucgr.C02472), disease resistance protein RPV1-like (Eucgr.C02626, Eucgr.C02627, Eucgr.F04311, Eucgr.K02295), endochitinase EP3-like (Eucgr.K02166), F-box/LRR-repeat protein 3 (Eucgr.K02937), (-)-germacrene D synthase (Eucgr.C02474), glutathione transferase GST 23 (Eucgr.K01652, Eucgr.K02248), LRR receptor-like serine/threonine-protein kinase GSO1 (Eucgr.K02824), MAP kinase 4 substrate 1 - MKS1 (Eucgr.C02621), TMV resistance protein N-like (Eucgr.C02111, Eucgr.C02112, Eucgr.C02623, Eucgr.C02624, Eucgr.C02638, Eucgr.C02618, Eucgr.F04310), WRKY transcription factor WRKY76 (Eucgr.C02659, Eucgr.F04317) (Table 3).

## 4 Discussion

In a GWA study, for any population of a given genetic background, the ability to detect associations is a function of several parameters that include the phenotypic variance, the heritability and the complexity of the trait. In the current study, we show significant phenotypic variance for resistance to MLD for all the traits assessed in a breeding population of *E. globulus*. The heritabilities were high ( $h^2 > 0.65$ ) for ADFO and SEV, whereas moderate values were observed for DEF. These estimates were similar or higher to those reported in previous studies carried out for equivalent resistance or damage traits of *Teratosphaeria* fungal diseases in *E. globulus*. For example, heritabilities for severity to *T. nubilosa* and *T. cryptica* ranging from 0.12 to 0.60 (Dungey et al., 1997; Carnegie and Ades, 2005; Milgate et al., 2005; Costa e Silva

et al., 2013; Hamilton et al., 2013) whereas values between 0.43 to 0.74 have been reported for proportion of adult foliage (Balmelli et al., 2013). Regarding the genetic complexity of the MLD resistant traits in *E. globulus*, two major genomic regions which explained a large proportion of the phenotypic variance have been previously identified, and an oligogenic or major gene control proposed for these traits (Freeman et al., 2008). However, recent studies show that resistance mechanisms for other diseases in *Eucalyptus* involve multiple interacting loci of variable effect, according to a quasi-infinite model (Butler et al., 2016; Resende et al., 2017). Specifically for heteroblastic transition, although major genes have been identified in various species, we still have a rather limited understanding of the number of loci involved in the regulatory pathway of this process (Hudson et al., 2014). Thus, for the disease-related traits analyzed here a quantitative nature is expected, which means that the phenotypic variation observed can be dependent on more than one major QTLs and many significant marker-trait associations would be identified.

In a first approximation to the genetic architecture to MLD resistance, we applied the widespread mixed model approach proposed by Yu et al. (2006). We used the PCA scores to control for population stratification and the genomic relationship matrix ( $G$ ) to control for close family relatedness. As expected, when the background genetic structure is not properly accounted for in the model, a high rate of false positive marker-trait associations was observed (Supplementary Figure 1). Our results, in line with previous association studies in *Eucalyptus*, demonstrated that the inclusion of family relatedness correction in association models drastically reduces the number of false-positive associations (Thavamanikumar et al., 2014; Müller et al., 2019). Thus, for

TABLE 3 Details of genes related with biotic stress resistance linked to the most significant Single Nucleotide Polymorphism (SNP) associated with precocity of vegetative change (proportion of adult foliage - ADFO) in *Eucalyptus globulus* identified with the ssGWAS model.

Annotation	SNP Name	Chr.	Position	Putative candidate genes	Age <sup>a</sup>
miR156 precursor	EuBR03s50801384	3	50801384	EgrMIR156.5 <sup>b</sup>	21, 26
miR157 precursor	EuBR11s29630035	11	29630035	EgrMIR157.4 <sup>b</sup>	14
TMV resistance protein N-like	EuBR03s38174146	3	38174146	Eucgr.C02111	21
	EuBR03s38174146	3	38174146	Eucgr.C02112	21
	EuBR03s49848389	3	49848389	Eucgr.C02618	14
	EuBR03s49848389/ EuBR03s49849249	3	49848389/ 49849249	Eucgr.C02623	14/14, 21, 26
	EuBR03s49848389/ EuBR03s49849249	3	49848389/ 49849249	Eucgr.C02624	14/14, 21, 26
	EuBR03s50231590	3	50231590	Eucgr.C02638	14, 21, 26
	EuBR06s51721432	6	51721432	Eucgr.F04310	21, 26
Disease resistance protein RPV1-like	EuBR03s49848389/ EuBR03s49849249	3	49848389/ 49849249	Eucgr.C02626	14/14, 21, 26
	EuBR03s49848389/ EuBR03s49849249	3	49848389/ 49849249	Eucgr.C02627	14/14, 21, 26
	EuBR06s51721432	6	51721432	Eucgr.F04311	21, 26
	EuBR11s30114988	11	30114988	Eucgr.K02295	14, 21
(-)-germacrene D synthase	EuBR03s47296521	3	47296521	Eucgr.C02474	14, 21
26S proteasome non-ATPase regulatory subunit 9	EuBR03s48526965/ EuBR03s48528495	3	48526965/ 48528495	Eucgr.C02542	14, 21/14
	EuBR09s22726692	9	22726692	Eucgr.I01173	14, 21
Dirigent protein 22 DIR22	EuBR03s47296521	3	47296521	Eucgr.C02472	14, 21
Endochitinase EP3-like	EuBR11s28810660	11	28810660	Eucgr.K02166	14
F-box/LRR-repeat protein 3	EuBR11s37401760	11	37401760	Eucgr.K02937	14
Glutathione transferase GST 23	EuBR11s20065344	11	20065344	Eucgr.K01652	26
	EuBR11s29646703	11	29646703	Eucgr.K02248	14, 21, 26
LRR receptor-like serine/threonine-protein kinase GSO1	EuBR11s36275775/ EuBR11s36290348	11	36275775/ 36290348	Eucgr.K02824	14/14, 21
Protein MKS1	EuBR03s49848389/ EuBR03s49849249	3	49848389/ 49849249	Eucgr.C02621	14/14, 21, 26
WRKY transcription factor WRKY76	EuBR03s50778072	3	50778072	Eucgr.C02659	21
	EuBR06s51841342	6	51841342	Eucgr.F04317	14

<sup>a</sup>Age of measurement (in months) at which the SNP was identified.

<sup>b</sup>identified by Hudson et al., 2014.

breeding populations consisting of families with cryptic relatedness, such as open-pollinated populations, the family structure (**G** matrix) efficiently captures the variations caused by genetic structure. The adjustment by using the PCA scores may be not adequate, and would be only necessary in populations with high levels of subrace differentiation or with different geographic origins (Cappa et al., 2013).

In the standard GWAS approach, each marker is tested independently for an association to the trait, included as a fixed effect in the model. For this method, two main drawbacks have been identified. First, the magnitude of the effects by the marker-trait

associations identified are largely overestimated, caused mainly by the limited sample sizes and/or the use of the same data set for discovery and parameter estimation (Beavis, 1998). Second, the effect of one marker is estimated ignoring the rest of the genome, causing confusion about the number and location and also contributing to the upward bias of estimated effects of loci identified (Kemper et al., 2012). The alternative approach that fits all markers simultaneously as a random effect, GBLUP-GWAS, is becoming popular due to having much in common with genomic selection (VanRaden, 2008). In our study, the GBLUP-GWAS and

the Single-SNP approach (model K) show similar results in the marker-trait associations identified, where 75% of the markers identified in the GBLUP model were previously identified in the Single-SNP model. When the SNP effects for this common SNPs were compared, our results agree with previous results that show an upwardly biased estimation under a fixed single-marker model (Kemper et al., 2012).

For GWA studies, the use of more phenotypic and genotypic data improves power and resolution (Tibbs Cortes et al., 2021). Thus, those ssGWAS methods that include the phenotypes of all genotyped and non-genotyped individuals as well as the pedigree information, have been shown to provide more consistent solutions and increased accuracy than the standard GWAS approach (Wang et al., 2012). Although this approach was successfully applied in GWA studies in domestic animals (Misztal et al., 2021), to our knowledge, this is the first report of ssGWAS applied to forest tree species. In this study, the ssGWAS approach identified 66 and 12 significant SNP-trait associations for the ADFO and DEF traits, respectively. The significant SNPs identified here are robust since the ssGWAS approach has shown, through simulations by Mancin et al. (2021), to control for false positives successfully. Additionally, our results show a general agreement between the Single-SNP and GBLUP strategies, with the same genomic regions identified with all models. However, the ssGWAS demonstrated superior performance by detecting new significant SNPs in chromosomes that had not been identified in the other models tested and resulted in clear peaks with higher  $-\log_{10}(p\text{-values})$ . Therefore, the ssGWAS is a promising method to the dissection of complex traits in populations when only a fraction of the population is genotyped, taking advantage of the available phenotypic information of individuals non-genotyped.

Most studies addressing disease resistance in forest trees have mainly assessed the disease's severity as the presence of leaf spots and/or the necrotic lesions present on the leaves. The high heritability found for severity of leaf spots (SEV) in agreement with previous reports, broadens the probability of detecting a gene of large effect. However, large heritability does not imply a direct relationship with the number, effect or proportion of the genetic variance of the genomic regions identified (Visscher et al., 2008). Our outcome can be explained by the low phenotypic variability of this trait in this population. Freeman et al. (2008) showed that the number of resistance QTLs detected in three breeding populations of *E. globulus* was strongly correlated with the variability within each population. In this study we also evaluated the degree of defoliation (DEF) as a descriptor of the impact of the MLD infection. For this trait, assessed at 21 months of age, we found 12 significant associations, although only one was significant after Bonferroni correction (Table 1). The low number of significant associations can be partially explained by the moderately-low heritability of this trait, reflecting the superior performance of the ssGWAS approach relative to the standard GWAS approaches.

The vegetative phase change in plants involves many morphological changes, well differentiated in many tree species as *E. globulus*. The mechanism to the juvenile to adult transition is under a strong genetic control with the same regulatory mechanisms in both annual herbaceous plants and perennial trees (Wang et al., 2011). Although the understanding of the genetic

control of this transition remains limited, it is mediated by environmental and endogenous signals (Poethig, 2010). In *E. globulus*, precocious vegetative change was evidenced in response to abiotic stress (drought, salt and high winds) in a natural coastal population (Jordan et al., 2000). Also, it is proposed as a mechanism involved in disease resistance, anticipating the timing of vegetative change to the resistant adult foliage. For precocity of vegetative change, measured as the proportion of adult foliage in this study, the ssGWAS model was able to identify the highest number of significant trait-SNP associations in this breeding population. For the three measurement ages, the same SNPs were identified as the markers that explained the highest genetic variation, which allowed identify two main significant regions in chromosome 3 and 11. Within those regions, we identified the *miR156.5* and *miR157.4* respectively, both of which target DNA-binding transcription factors that regulates the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. The *miR156* has been recognized as a master regulator of vegetative change in plants, promoting juvenility due to high expression in seedlings that decrease during development (Poethig, 2010; Manuela and Xu, 2020). Our results support similar reports from bi-parental QTLs mapping studies that identified *miR156.5* as the candidate gene causing difference in the timing of phase change between precocious and normal natural populations of *E. globulus* (Hudson et al., 2014). Furthermore, emerging studies suggest that *miR156* is probably an integral component of the *miRNA* response to all environmental stresses in plants (Zhao et al., 2012; Jeyakumar et al., 2020). Specifically for fungal pathogen defense response, *miR156* targets the plant disease resistance proteins (NBS-LRR proteins) in *Populus trichocarpa* infected with stem canker pathogen (Zhao et al., 2012). Additionally, precursors of *miR156* have also been reported located near the main QTL identified for *Puccinia* rust resistance (*Ppr1*), suggesting a main role in the genetic control of plant resistance (Butler et al., 2016).

We also identified various genes proposed to be involved in fungal pathogen resistance through different mechanisms. For example, the 26S proteasome non-ATPase regulatory subunit 9 is part of the ubiquitin/26S proteasome system, known to be involved in almost every step of the defense mechanisms in plants, regardless of the type of pathogen, regulating key cellular processes through protein degradation (Dielen et al., 2010). The dirigent protein 22 (DIR22) is part of a gene family involved in plant defense by their role in the dynamic reorganization of the cell wall through the formation of lignans and production of defense-related compounds (Yadav et al., 2021). Numerous studies have shown that the expression levels of DIRs genes are enhanced during the late stages of fungal infection (e.g., *Colletotrichum gleosporioides*, *Botrytis cinerea*, *Fusarium oxysporum*) in various plant species, increasing antifungal responses (Arasan et al., 2013; Borges et al., 2013; Reboledo et al., 2015). Glutathione transferase GST23 belong to a super-family of multifunctional proteins in plants, being one of their most important roles the detoxification of fungal toxins in the response against biotic stresses (Gardiner et al., 2010). The expression of GSTs is markedly induced during fungal infection, leading to enhanced resistance to the pathogen (Ahn et al., 2016). The (-)-germacrene D synthase is an enzyme involved in the volatile

terpenoid biosynthesis. It has been reported that the downregulation of these genes alters monoterpene levels leading to resistance against biotic stresses at a global level, including defense responses to fungus infection (Rodríguez et al., 2014). Lastly, the endochitinase EP3-like genes encode a protein involved in catalyzing the hydrolytic cleavage in chitin, the prominent wall component of fungus, and are induced by pathogen attack and other biotic stresses (Nagpure et al., 2014).

Finally, we identified genes involved in pathogen defenses' regulation and signaling mechanisms. MAP kinase 4 substrate 1 (MKS1) has been reported to be pivotal in signaling basal defense responses (Andreasson et al., 2005), whereas the WRKY transcription factor WRKY76 is involved in the regulation of the plant defense signaling pathway (Rushton et al., 2010). In addition, many plant disease resistance proteins (NBS-LRR proteins), which trigger signal transduction leading to the induction of defense responses (Andersen et al., 2018), were found. For instance, the disease resistance protein RPV1 is known to confer resistance to oomycete and fungal pathogens in cultivated grapevines (*Vitis vinifera*) (Gessler et al., 2011). LRR receptor-like serine/threonine-protein kinase GSO1 was reported to be related to the resistance of *Nicotiana tabacum* to the oomycete *Phytophthora parasitica* (Dang et al., 2019). TMV resistance protein N-like was reported to be involved in the resistance in potato to the fungus *Synchytrium endobioticum* (Hehl et al., 1999), and in peanut to two serious fungal foliar diseases (Dang et al., 2021). F-Box/LRR repeat protein was reported to be activated by fungal pathogens (*Cladosporium fulvum*, *Puccinia striiformis*) in different plant species in order to regulate defense responses (Yin et al., 2018).

## 5 Conclusions

In the present study, we carried out GWAS for traits related to response and escape to MLD in a breeding population of *E. globulus* only partially genotyped, and assessed the advantage of including phenotypes from relatives without genotypes in a single-step procedure. We compare several GWAS models that differ in the correction for population structure, the statistical approach to fit the markers and the impact of including combined pedigree, genomic and phenotypic information. Several SNP-trait associations were detected for the precocity of vegetative phase change (proportion of adult foliage). The same SNPs were identified at different measurement ages for the different strategies, providing validation of the results for these specific loci. Additionally, our results suggest that ssGWAS is a powerful model in detecting regions associated to the escape to MLD, showing a greater number of significant SNP associations than Single-SNP and GBLUP-GWAS models. Associations were detected near the *miRNAs* that regulate the timing of vegetative phase change, consistent with results from other studies about heteroblastic transition in trees and annual species. Thus, our research supports the hypothesis that this is a complex trait determined by a large number of genes with diverse biological functions, that can also include a mechanism of response to biotic stress, such as fungal resistance. Although the identified putative candidate genes will require future validations, they provide a better

understanding of the genetic architecture of vegetative change in *Eucalyptus*. Overall, our results represent a foundational step in marker-assisted selection for a tree breeding program. For example, significant markers or validated genes identified by GWAS can be incorporated into prediction models, to accelerate breeding progress. However, since this is an association discovery study, validation in independent populations is necessary.

## Data availability statement

The phenotypic and genotypic data analyzed in this study are not publicly available. The data are available upon reasonable request to the corresponding author and with permission of INIA. Requests to access these datasets should be directed to GB, [gbalmelli@inia.org.uy](mailto:gbalmelli@inia.org.uy).

## Author contributions

MQ, IA, and GB contributed to conception and design of the study. GB designed the field experiment, collected the phenotypic data and provided the marker data. MQ performed the statistical analysis under the supervision of IA. FG and CS performed the candidate gene analysis and wrote this section of the manuscript. MQ wrote the final version of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Instituto Nacional de Investigación Agropecuaria (INIA). MQ received a postdoctoral fellowship from INIA.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1124768/full#supplementary-material>

## References

- Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., and Misztal, I. (2019). Frequentist p-values for large-scale single step genome-wide association, with an application to birth weight in American Angus cattle. *Genet. Sel. Evol.* 51, 1–8. doi: 10.1186/s12711-019-0469-3
- Aguilar, I., Misztal, I., Johnson, D., Legarra, A., Tsuruta, S., and Lawlor, T. (2010). Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi: 10.3168/jds.2009-2730
- Aguilar, I., Misztal, I., Tsuruta, S., Legarra, A., and Wang, H. (2014). “PREGSF90 – POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs,” in *Proceedings, 10th World Congr. Genet. Appl. to Livest. Prod.*, Vancouver, Canada. 90–92.
- Aguilar, I., Tsuruta, S., Masuda, Y., Lourenco, D. A. L., Legarra, A., and Misztal, I. (2018). “BLUPF90 suite of programs for animal breeding with focus on genomics,” in *Proceedings of the World Congress on Genetics Applied to Livestock Production*, Auckland, Australia. 751.
- Ahn, S. Y., Kim, S. A., and Yun, H. K. (2016). Glutathione s-transferase genes differently expressed by pathogen-infection in vitis flexuosa. *Plant Breed. Biotechnol.* 4, 61–70. doi: 10.9787/PBB.2016.4.1.61
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Andersen, E. J., Ali, S., Byamukama, E., Yen, Y., and Nepal, M. P. (2018). Disease resistance mechanisms in plants. *Genes* 9, 339. doi: 10.3390/genes9070339
- Andreasson, E., Jenkins, T., Brodersen, P., Thorgrimsen, S., Petersen, N. H., Zhu, S., et al. (2005). The map kinase substrate mks1 is a regulator of plant defense responses. *EMBO J.* 24, 2579–2589. doi: 10.1038/sj.emboj.7600737
- Arasan, S. K. T., Park, J.-I., Ahmed, N. U., Jung, H.-J., Hur, Y., Kang, K.-K., et al. (2013). Characterization and expression analysis of dirigent family genes related to stresses in brassica. *Plant Physiol. Bioch.* 67, 144–153. doi: 10.1016/j.plaphy.2013.02.030
- Balmelli, G., Simeto, S., Marroni, V., Altier, N., and Diez, J. J. (2014). Genetic variation for resistance to *Mycosphaerella* leaf disease and *Eucalyptus* rust on *Eucalyptus globulus* in Uruguay. *Australas. Plant Pathol.* 43, 97–107. doi: 10.1007/s13313-013-0254-7
- Balmelli, G., Simeto, S., Torres, D., Castillo, A., Altier, N., and Diez, J. J. (2013). Susceptibility to *Teratosphaeria nubulosa* and precocity of vegetative phase change in *Eucalyptus globulus* and *E. maidenii* (Myrtaceae). *Aust. J. Bot.* 61, 583–591. doi: 10.1071/BT13225
- Balmelli, G., Simeto, S., Torres, D., Hirigoyen, A., Castillo, A., Altier, N., et al. (2016). Impact of *Teratosphaeria nubulosa* over tree growth and survival of *Eucalyptus globulus* and *eucalyptus maidenii* in Uruguay. *New For.* 47, 829–843. doi: 10.1007/s11056-016-9547-3
- Beavis, W. (1998). *QTL analyses: power, precision, and accuracy* (Boca Raton, FL, USA: CRC Publishing), 145–162. doi: 10.1201/9780429117770-10
- Borges, A. F., Ferreira, R. B., and Monteiro, S. (2013). Transcriptomic changes following the compatible interaction *Vitis vinifera*–*crisphe* necator. paving the way towards an enantioselective role in plant defence modulation. *Plant Physiol. Bioch.* 68, 71–80. doi: 10.1016/j.plaphy.2013.03.024
- Butler, J. B., Freeman, J. S., Vaillancourt, R. E., Potts, B. M., Glen, M., Lee, D. J., et al. (2016). Evidence for different QTL underlying the immune and hypersensitive responses of *Eucalyptus globulus* to the rust pathogen *Puccinia psidii*. *Tree Genet. Genomes* 12, 39. doi: 10.1007/s11295-016-0987-x
- Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, 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 Sci.* 284, 9–15. doi: 10.1016/j.plantsci.2019.03.017
- Cappa, E. P., El-Kassaby, Y. A., Garcia, M. N., Acuña, C., Borralho, N. M., Grattapaglia, D., et al. (2013). Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS One* 8. doi: 10.1371/journal.pone.0081267
- Cappa, E. P., and Varona, L. (2013). An assessor-specific Bayesian multi-threshold mixed model for analyzing ordered categorical traits in tree breeding. *Tree Genet. Genomes* 9, 1423–1434. doi: 10.1007/s11295-013-0648-2
- Carnegie, A. J., and Ades, P. K. (2003). *Mycosphaerella* leaf disease reduces growth of plantation-grown *Eucalyptus globulus*. *Aust. For.* 66, 113–119. doi: 10.1080/00049158.2003.10674900
- Carnegie, A. J., and Ades, P. K. (2005). Variation in *Eucalyptus globulus* LABILL and *E. nitens* DEAN and MAIDEN in susceptibility of adult foliage to disease caused by *Mycosphaerella cryptica* (COOKE) HANSF. *Silvae Genet.* 54, 174–184. doi: 10.1515/sg-2005-0026
- Costa e Silva, J., Potts, B. M., and Tilyard, P. (2013). Stability of genetic effects across clonal and seedling populations of *Eucalyptus globulus* with common parentage. *For. Ecol. Manage.* 291, 427–435. doi: 10.1016/j.foreco.2012.11.005
- Covarrubias-Pazarán, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11, 1–15. doi: 10.1371/journal.pone.0156744
- Dang, P. M., Lamb, M. C., and Chen, C. Y. (2021). Association of differentially expressed r-gene candidates with leaf spot resistance in peanut (*Arachis hypogaea* L.). *Mol. Biol. Rep.* 48, 323–334. doi: 10.1007/s11033-020-06049-3
- Dang, J., Wang, J., Yang, Y., Shang, W., Guo, Q., and Liang, G. (2019). Resistance of *Nicotiana tabacum* to *Phytophthora parasitica* var. *nicotianae* race 0 is enhanced by the addition of *N. plumbaginifolia* chromosome 9 with a slight effect on host genomic expression. *Crop Sci.* 59, 2667–2678. doi: 10.2135/cropsci2019.01.0005
- Dielen, A.-S., Badaoui, S., Candresse, T., and German-Retana, S. (2010). The ubiquitin/26S proteasome system in plant–pathogen interactions: a never-ending hide-and-seek game. *Mol. Plant Pathol.* 11, 293–308. doi: 10.1111/j.1364-3703.2009.00596.x
- Du, Q., Lu, W., Quan, M., Xiao, L., Song, F., Li, P., et al. (2018). Genome-wide association studies to improve wood properties: challenges and prospects. *Front. Plant Sci.* 871. doi: 10.3389/fpls.2018.01912
- Dungey, H. S., Potts, B. M., Carnegie, A. J., and Ades, P. K. (1997). *Mycosphaerella* leaf disease: genetic variation in damage to *Eucalyptus nitens*, *Eucalyptus globulus*, and their F1 hybrid. *Can. J. For. Res.* 759, 750–759. doi: 10.1139/x96-210
- Freeman, J. S., Potts, B. M., and Vaillancourt, R. E. (2008). Few mendelian genes underlie the quantitative response of a forest tree, *Eucalyptus globulus*, to a natural fungal epidemic. *Genetics* 178, 563–571. doi: 10.1534/genetics.107.081414
- Gardiner, S. A., Boddú, J., Berthiller, F., Hametner, C., Stupar, R. M., Adam, G., et al. (2010). Transcriptome analysis of the barley–deoxynivalenol interaction: evidence for a role of glutathione in deoxynivalenol detoxification. *Mol. Plant-Microbe Interact.* 23, 962–976. doi: 10.1094/MPMI-23-7-0962
- Gessler, C., Pertot, I., and Perazzolli, M. (2011). *Plasmopara viticola*: a review of knowledge on downy mildew of grapevine and effective disease management. *Phytopathol. Mediterr.* 50, 3–44. doi: 10.14601/Phytopathol\_Mediterr-9360
- Gualdrón Duarte, J. L., Cantet, R. J., Bates, R. O., Ernst, C. W., Raney, N. E., and Steibel, J. P. (2014). Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations. *BMC Bioinf.* 15, 1–11. doi: 10.1186/1471-2105-15-246
- Hamilton, M. G., Williams, D. R., Tilyard, P. A., Pinkard, E. A., Wardlaw, T. J., Glen, M., et al. (2013). A latitudinal cline in disease resistance of a host tree. *Hereditas* 110, 372–379. doi: 10.1038/hdy.2012.106
- Hehl, R., Faurie, E., Hesselbach, J., Salamini, F., Whitham, S., Baker, B., et al. (1999). Tmv resistance gene n homologues are linked to *Synchytrium endobioticum* resistance in potato. *Theor. Appl. Genet.* 98, 379–386. doi: 10.1007/s001220051083
- Hudson, C. J., Freeman, J. S., Jones, R. C., Potts, B. M., Wong, M. M., Weller, J. L., et al. (2014). Genetic control of heterochrony in *Eucalyptus globulus*. *G3 Genes Genomes Genet.* 4, 1235–1245. doi: 10.1534/g3.114.011916
- Hunter, G. C., Crous, P. W., Carnegie, A. J., and Wingfield, M. J. (2009). *Teratosphaeria nubulosa*, a serious leaf disease pathogen of eucalyptus spp. in native and introduced areas. *Mol. Plant Pathol.* 10, 1–14. doi: 10.1111/j.1364-3703.2008.00516.x
- Jeyakumar, J. M. J., Ali, A., Wang, W. M., and Thiruvengadam, M. (2020). Characterizing the role of the *miR156*-SPL network in plant development and stress response. *Plants* 9, 1–15. doi: 10.3390/plants9091206
- Jordan, G. J., Potts, B. M., Chalmers, P., and Wiltshire, R. J. (2000). Quantitative genetic evidence that the timing of vegetative phase change in *Eucalyptus globulus* ssp. *globulus* is an adaptive trait. *Aust. J. Bot.* 48, 561–567. doi: 10.1071/BT99038
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–354. doi: 10.1038/ng.548
- Kemper, K., Hans, D., Peter, V., and Goddard, M. (2012). Comparing linkage and association analyses in sheep points to a better way of doing GWAS. *Genet. Res. Camb.* 94, 191–203. doi: 10.1017/S0016672312000365
- Mancin, E., Lourenco, D., Bermann, M., Mantovani, R., and Misztal, I. (2021). Accounting for population structure and phenotypes from relatives in association mapping for farm animals: a simulation study. *Front. Genet.* 12. doi: 10.3389/fgenet.2021.642065
- Manuela, D., and Xu, M. (2020). Juvenile leaves or adult leaves: determinants for vegetative phase change in flowering plants. *Int. J. Mol. Sci.* 21, 1–16. doi: 10.3390/ijms21249753
- Marroni, F., Pinosio, S., Zaina, G., Fogolari, F., Felice, N., Cattonaro, F., et al. (2011). Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. *Tree Genet. Genomes* 7, 1011–1023. doi: 10.1007/s11295-011-0391-5
- Masuda, Y., Legarra, A., Aguilar, I., and Misztal, I. (2019). Efficient quality control methods for genomic and pedigree data used in routine genomic evaluation. *J. Anim. Sci.* 97, 50–51. doi: 10.1093/jas/skz258.101
- Milgate, A. W., Potts, B. M., Joyce, K., Mohammed, C., and Vaillancourt, R. E. (2005). Genetic variation in *Eucalyptus globulus* for susceptibility to *Mycosphaerella nubulosa* and its association with tree growth. *Australas. Plant Pathol.* 34, 11–18. doi: 10.1071/AP04073

- Misztal, I., Aguilar, I., Lourenco, D., Ma, L., Steibel, J. P., and Toro, M. (2021). Emerging issues in genomic selection. *J. Anim. Sci.* 99, 1–14. doi: 10.1093/jas/skab092
- Mohammed, C., Wardlaw, T., Smith, A., Pinkard, E., Battaglia, M., Glen, M., et al. (2003). *Mycosphaerella* leaf diseases of temperate eucalypts around the southern pacific rim. *New Zeal. J. For. Sci.* 33, 362–372.
- Müller, B. S., de Almeida Filho, J. E., Lima, B. M., Garcia, C. C., Missiaggia, A., Aguiar, A. M., et al. (2019). Independent and joint-GWAS for growth traits in *Eucalyptus* by assembling genome-wide data for 3373 individuals across four breeding populations. *New Phytol.* 221, 818–833. doi: 10.1111/nph.15449
- Nagpure, A., Choudhary, B., and Gupta, R. K. (2014). Chitinases: in agriculture and human healthcare. *Cr. Rev. Biotech.* 34, 215–232. doi: 10.3109/07388551.2013.790874
- Poethig, R. S. (2010). The past, present, and future of vegetative phase change. *Plant Physiol.* 154, 541–544. doi: 10.1104/pp.110.161620
- Quezada, M., Aguilar, I., and Balmelli, G. (2022). Genomic breeding values prediction including populational selfing rate in an open-pollinated *Eucalyptus globulus* breeding population. *Tree Genet. Genomes* 18, 10. doi: 10.1007/s11295-021-01534-7
- R Core Team (2022). *R: a language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).
- Reboleo, G., Del Campo, R., Alvarez, A., Montesano, M., Mara, H., and Ponce de León, I. (2015). *Physcomitrella patens* activates defense responses against the pathogen *Colletotrichum gloeosporioides*. *Int. J. Mol. Sci.* 16, 22280–22298. doi: 10.3390/ijms160922280
- Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, O. B., 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. doi: 10.1111/nph.14266
- Rodríguez, A., Shimada, T., Cervera, M., Alquézar, B., Gadea, J., Gómez-Cadenas, A., et al. (2014). Terpene down-regulation triggers defense responses in transgenic orange leading to resistance against fungal pathogens. *Plant Physiol.* 164, 321–339. doi: 10.1104/pp.113.224279
- Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). Wrky transcription factors. *Trends Plant Sci.* 15, 247–258. doi: 10.1016/j.tplants.2010.02.006
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., et al. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 50, D20–D26. doi: 10.1093/nar/gkab1112
- 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 Phytol.* 206, 1527–1540. doi: 10.1111/nph.13322
- Simeto, S., Balmelli, G., and Pérez, C. (2020). “Diseases of *Eucalyptus* plantations in Uruguay: current state and management alternatives,” in *For. pest dis. manag. lat. am.* Ed. S. A. Estay (Springer, Cham: Springer Nature Switzerland AG 2020), 123–144.
- Smith, A. H., Wardlaw, T. J., Pinkard, E. A., Ratkowsky, D., and Mohammed, C. L. (2017). Impacts of *Teratosphaeria* leaf disease on plantation *Eucalyptus globulus* productivity. *For. Pathol.* 47, 1–9. doi: 10.1111/efp.12310
- Thavamanikumar, S., McManus, L. J., Ades, P. K., Bossinger, G., Stackpole, D. J., Kerr, R., 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, 1661–1678. doi: 10.1007/s11295-014-0787-0
- Tibbits, W. N., Boomsma, D. B., and Jarvis, S. (1997). “Distribution, biology, genetics, and improvement programs for *Eucalyptus globulus* and *e. nitens* around the world,” in *Proceedings of the 24th biennial southern tree improvement conferences* Orlando, Florida, USA. 81–95.
- Tibbs Cortes, L., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *Plant Genome*, 1–17. doi: 10.1002/tpg2.20077
- Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Software* 3, 731. doi: 10.21105/joss.00731
- Valenzuela, C. E., Ballesta, P., Ahmar, S., Fiaz, S., Heidari, P., Maldonado, C., et al. (2021). Haplotype-and SNP-based GWAS for growth and wood quality traits in *Eucalyptus cladocalyx* trees under arid conditions. *Plants* 10, 1–17. doi: 10.3390/plants10010148
- VanRaden, P. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era – concepts and misconceptions. *Nat. Rev. Genet.* 9, 255–266. doi: 10.1038/nrg2322
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb.)* 94, 73–83. doi: 10.1017/S0016672312000274
- Wang, J. W., Park, M. Y., Wang, L. J., Koo, Y., Chen, X. Y., Weigel, D., et al. (2011). *MiRNA* control of vegetative phase change in trees. *PLoS Genet.* 7, 21–25. doi: 10.1371/journal.pgen.1002012
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis* (New York: Springer-Verlag).
- Yadav, V., Wang, Z., Yang, X., Wei, C., Changqing, X., and Zhang, X. (2021). Comparative analysis, characterization and evolutionary study of dirigent gene family in cucurbitaceae and expression of novel dirigent peptide against powdery mildew stress. *Genes* 12, 326. doi: 10.3390/genes12030326
- Yin, J.-I., Fang, Z.-W., Sun, C., Zhang, P., Zhang, X., Lu, C., et al. (2018). Rapid identification of a stripe rust resistant gene in a space-induced wheat mutant using specific locus amplified fragment (slaf) sequencing. *Sci. Rep.* 8, 1–9. doi: 10.1038/s41598-018-21489-5
- Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinf* 19, 619–628. doi: 10.1016/j.gpb.2020.10.007
- Yong, W. T. L., Ades, P. K., Runa, F. A., Bossinger, G., Sandhu, K. S., Potts, B. M., et al. (2021). Genome-wide association study of myrtle rust (*Austropuccinia psidii*) resistance in *eucalyptus obliqua* (subgenus *eucalyptus*). *Tree Genet. Genomes* 17, 31. doi: 10.1007/s11295-021-01511-0
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhao, J. P., Jiang, X. L., Zhang, B. Y., and Su, X. H. (2012). Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. *PLoS One* 7, 1–13. doi: 10.1371/journal.pone.0044968