

Article

# Genomic Predictions Using Low-Density SNP Markers, Pedigree and GWAS Information: A Case Study with the Non-Model Species *Eucalyptus cladocalyx*

Paulina Ballesta <sup>1</sup>, David Bush <sup>2</sup> , Fabyano Fonseca Silva <sup>3</sup> and Freddy Mora <sup>1,\*</sup> <sup>1</sup> Institute of Biological Sciences, University of Talca, 2 Norte 685, Talca 3460000, Chile; pballesta@utalca.cl<sup>2</sup> CSIRO–Australian Tree Seed Centre, Acton 2601, Australia; David.Bush@csiro.au<sup>3</sup> Department of Animal Science, Universidade Federal de Viçosa, Viçosa 36570-900, Brazil; fabyanofonseca@ufv.br

\* Correspondence: fmora@utalca.cl

Received: 25 November 2019; Accepted: 9 January 2020; Published: 13 January 2020



**Abstract:** High-throughput genotyping techniques have enabled large-scale genomic analysis to precisely predict complex traits in many plant species. However, not all species can be well represented in commercial SNP (single nucleotide polymorphism) arrays. In this study, a high-density SNP array (60 K) developed for commercial *Eucalyptus* was used to genotype a breeding population of *Eucalyptus cladocalyx*, yielding only ~3.9 K informative SNPs. Traditional Bayesian genomic models were investigated to predict flowering, stem quality and growth traits by considering the following effects: (i) polygenic background and all informative markers (GS model) and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model). The estimates of pedigree-based heritability and genomic heritability varied from 0.08 to 0.34 and 0.002 to 0.5, respectively, whereas the predictive ability varied from 0.19 (GS) and 0.45 (GSq). The GSq approach outperformed GS models in terms of predictive ability when the proportion of the variance explained by the significant marker-trait associations was higher than those explained by the polygenic background and non-significant markers. This approach can be particularly useful for plant/tree species poorly represented in the high-density SNP arrays, developed for economically important species, or when high-density marker panels are not available.

**Keywords:** Bayesian models; deviance information criterion; marker-trait associations; predictive ability

## 1. Introduction

A major focus of modern quantitative genetics is on assessing the association between polymorphic markers with phenotypic variations of complex traits. In this sense, genotype–phenotype studies for quantitative traits at the genome level usually require high-density genetic marker panels, i.e., a large number of markers throughout the genome and large population sizes to obtain sufficient power and prediction resolution [1,2]. The development of several genotyping platforms through high-density single nucleotide polymorphism (SNP) arrays, such as genotyping-by-sequencing (GBS) or SNP chips, has enabled the identification of quantitative trait loci (QTL) for different target traits in various plant species [3–6]. Silva-Junior et al. [7], for instance, developed a genome-wide SNP chip for multiple species of *Eucalyptus*, which has been effective for genomic studies in a wide variety of economically important eucalypt species and their hybrids, including *Eucalyptus grandis*, *Eucalyptus urophylla*, *Eucalyptus nitens* and *Eucalyptus globulus* [8–12]. However, despite the versatility of this SNP

array, it does not perform as well in terms of genome coverage or number of available SNPs for species which are more-distantly related to those for which the chip was developed [13].

According to Pryce et al. [14], the use of low-density marker panels will inevitably affect the precision of QTL detection in genome-wide association studies (GWAS) and the accuracy of genomic prediction of target traits to some degree. On the other hand, Müller et al. [15] found that prediction models using a low-density marker panel (a subset of ~5000 SNPs) provided predictive abilities almost equivalent to using all available SNPs in *Eucalyptus* spp., however, they concluded that it is not yet clear whether the use of smaller SNP subsets is warranted for the long-term implementation of genomic selection in *Eucalyptus*, an aspect that remains still unknown. Marker panels could be considered as either low- or high-density depending on the genome size, the extent of linkage disequilibrium (LD) and the traits of interest. Larger genomes and rapid breakdown of LD would imply that a higher density of SNP loci would be needed to detect QTL. However, many authors quite arbitrarily describe panels as high or low density such that comparisons among organisms should be carefully made. In cattle (genome size 3 Gb), for instance, a DNA array of 50 K SNPs is considered a low-density panel [14], while in *Eucalyptus nitens* (640 Mb) the 60 K SNP chip was considered as a high-density marker panel by Suontama et al. [11]. However, LD in the domesticated bovine genome, though it decays rapidly [16], is more extensive than in nearly-wild eucalypts, thus comparison of SNP density is further complicated. Müller et al. [15] considered the subset of ~5 K SNPs that they used as a low-density marker panel in *Eucalyptus benthamii* and *Eucalyptus pellita*. Notably, low-density SNP chips are considered as a way to reduce the cost of high-density SNP panels in animal breeding and would enable cost-effective implementation of genomic studies [17,18]. In accordance with this, recently Silva et al. [19] used linkage (LA) and linkage disequilibrium (LDA) analyses (termed 'LALDA') for low density-based genomic selection (GS) purposes in animal breeding. In a Bayesian framework, the authors evaluated several GS models and verified the slight superiority of the LALDA models in comparison to traditional LDA models, concluding that the best performance evidenced by the LALDA approach can be due mainly to the small number of markers used, since it enabled to exploit relevant genomic regions that were not directly considered in the LDA, in which this extra information may have contributed to the improvement of the model performance. Similarly, other studies have explored the benefits of including fixed-effect covariates tagging peak genome-wide association study (GWAS) signals [20]. Based on their results, the authors conclude that the universal implementation of GS + GWAS for predicting the breeding values of all possible traits should be investigated on a trait-by-trait basis. On the other hand, Bernardo [21] determined whether explicitly modeling the effects of known major genes affects the response to genomic selection and showed that specifying a fixed effect for a single major gene was never disadvantageous except with a gene explains <10% of genetic variance. In that case, it should be included as a covariate in the traditional ridge-regression best linear unbiased prediction (RR-BLUP) model. Therefore, the objectives of this study were (i) to investigate the possible benefits of including marker-trait associations (from a GWAS analysis) and pedigree information into traditional GS model to predict complex traits in trees of *E. cladocalyx* using low-density SNP markers, and (ii) to assess the efficiency of Bayesian whole-genome regression models (Bayes A, Bayes B, Bayes C $\pi$  and Bayesian Ridge Regression) in terms of predictive ability of complex traits and goodness-of-fit measures in the presence of QTL-genotype effects obtained from a marker-trait association analysis.

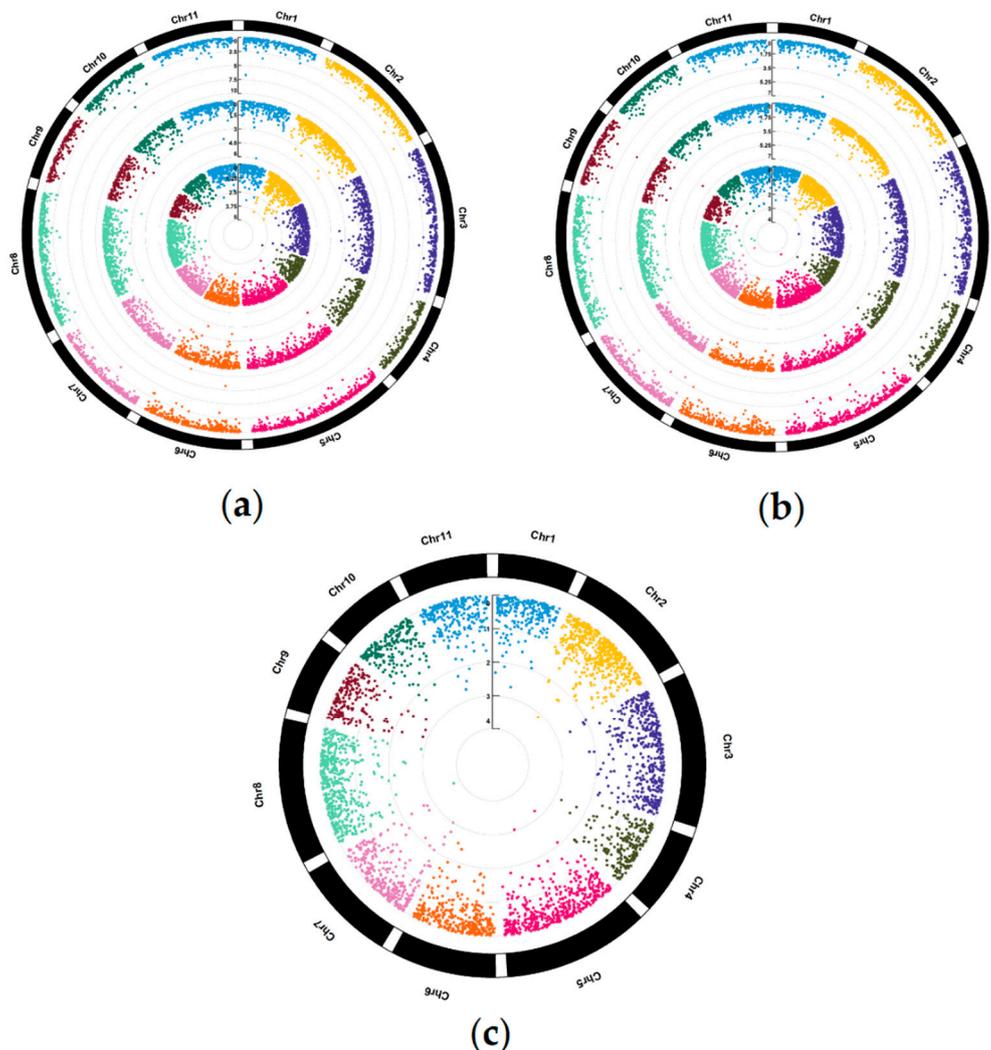
## 2. Results

### 2.1. SNP Data and Comparison of Genomic Prediction Models

In the present study, we genotyped a breeding population of *E. cladocalyx* using the 60 K SNP chip, yielding a subset of only ~3.9 K informative SNPs, due to the low number of polymorphic loci found at locations where they might have been expected. The ~3.9 K SNPs (~6% of the total SNP array) that were retained after filtering for minor allele frequency (MAF) and missing data were located in all eleven chromosomes of *Eucalyptus*, with an average of ~353 SNPs per chromosome, a density of 6

SNPs per 1 Mb and distributed with an average distance between SNPs of 11,600 bp (Table S1 and Figure S1). The genetic diversity of the genotyped population (evaluated with the 3879 SNPs) was 0.28 (in terms of expected heterozygosity) and the observed heterozygosity was 0.22. According to the population structure analysis, the studied population was strongly differentiated in three groups, with the following pairwise  $F_{ST}$  values:  $F_{ST} 1 = 0.086$ ,  $F_{ST} 2 = 0.28$  and  $F_{ST} 3 = 0.25$ .

The association analysis identified a total of 90 significant marker-trait associations (MTAs), which were also distributed across all 11 chromosomes, of which 11, 16, 5, 26, 10, 5 and 17 MTAs were identified for total tree height (HT), diameter at breast height (DBH), stem straightness (STR), slenderness index (SLD), wood density (WD), flowering intensity (FI) and first bifurcation height (BHT), respectively (Figure 1). The MTAs explained a relatively low proportion of the total phenotypic variation, with values of 2% to 4%, 3% to 6%, 3% to 4%, 3% to 5%, 2% to 10%, 3% to 7% and 3% to 4% for HT, DBH, STR, BHT, SLD, WD and FI, respectively. The SNPs involved in these MTAs were posteriorly considered as relationship matrices in the prediction models that include the QTL-genotype effects (GSq model).



**Figure 1.** Manhattan plot for (a) growth-related traits (tree height, diameter at breast height and slenderness index; displayed from inside to outside), (b) stem quality traits (stem straightness, wood density and first bifurcation height; displayed from inside to outside) and (c) flowering intensity.

The prediction models that include the MTA information (GSq) and traditional genomic prediction models (GS) were compared in terms of goodness of fit through the Deviance Information Criterion

(DIC) and predictive ability (PA). The DIC and PA values of all fitted models are shown in Tables 1 and 2, respectively. GSq models outperformed the GS model in terms of PA when more than 10 significant MTAs were included in GSq predictions (i.e., the following traits HT, DBH, SLD, BHT and WD), whose values ranged between 0.19–0.39 (GS) and 0–24–0.45 (GSq). This result was consistent with the goodness-of-fit measures in most of cases. For HT, GS model outperformed the GSq model for Bayes B (BB), according to the  $\Delta$ DIC value, while the GSq model was significantly superior in Bayes A (BA), Bayes C (BC) and Bayesian Ridge Regression (BRR) methods ( $\Delta$ DIC > 10). The GSq approach had a higher predictive ability than the GS approach for all Bayesian models. For DBH, in contrast, the best performance (in terms of goodness-of-fit and PA) was obtained by GSq models based on any Bayesian prediction method (BA, BB, BC or BRR; with  $\Delta$ DIC > 50). In addition, the predictive ability of DBH based on GSq model was two times higher than those based on traditional GS. For STR, GS models presented significantly lower DIC values than GSq models for BC and BRR ( $\Delta$ DIC > 10), while the PA values for both approaches were similar. For SLD, goodness-of-fit measures for GSq models were better than traditional models based on any Bayesian genomic model ( $\Delta$ DIC > 20). The predictive ability of SLD varied between 0.20–0.21 and 0.31–0.32 for GS and GSq models, respectively. For WD, the most of GSq models had better goodness-of-fit measures compared with GS models ( $\Delta$ DIC > 5) for all Bayesian methods. The PAs of WD ranged between 0.27 and 0.43, which were higher for all GSq models than those based on the GS approach. For FI, GS models presented lower DIC values than GSq models for BA, BB, BC and BRR prediction methods ( $\Delta$ DIC > 5). Predictive ability values for FI based on both approaches were similar and varied from 0.23 to 0.25. For BHT, there was a strong superiority of the GSq models over GS in all Bayesian methods in terms of goodness-of-fit measures ( $\Delta$ DIC > 60). Moreover, the PA of BHT varied between 0.19–0.20 and 0.38–0.39 for GS and GSq models, respectively. Consistently with  $\Delta$ DIC values, the predictive ability of BHT based on GSq models was two times higher than those based on GS models.

**Table 1.** Deviance information criterion (DIC) of genomic prediction in *Eucalyptus cladocalyx* based on (i) polygenic background (pedigree information) and all informative markers (GS model) and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model).

Trait/Model	Bayes A	Bayes B	Bayes C	BRR <sup>b</sup>
Tree height				
GS	1968.9	1959.5	1968.6	1965.4
GSq	1951.2	1971.3	1941.3	1941.2
$\Delta$ DIC <sup>a</sup>	17.7 **	11.8 **	27.3 **	24.2 **
Diameter at breast height				
GS	2556.5	2544.7	2539.8	2538.2
GSq	2490.4	2480.7	2480.2	2473.3
$\Delta$ DIC	66.1 **	64.0 **	59.6 **	64.9 **
Stem straightness				
GS	947.7	941.4	932.7	935.3
GSq	947.0	944.7	947.2	946.2
$\Delta$ DIC	0.7	3.3	14.5 **	10.8 **
Slenderness index				
GS	4302.5	4299.5	4294.3	4290.5
GSq	4268.1	4268.5	4264.3	4261.3
$\Delta$ DIC	34.4 **	31.0 **	30.0 **	29.2 **
Wood density				
GS	2094.3	2082.4	2101.8	2042.0
GSq	2067.1	2075.9	2067.6	2070.3
$\Delta$ DIC	27.2 **	6.5 *	34.3 **	28.4 **

Table 1. Cont.

Trait/Model	Bayes A	Bayes B	Bayes C	BRR <sup>b</sup>
Flowering intensity				
GS	1293.9	1301.2	1282.3	1285.9
GSq	1306.0	1309.7	1301.3	1295.2
$\Delta$ DIC	12.1 **	8.4 *	19.0 **	9.4 *
First bifurcation height				
GS	1491.9	1490.6	1487.3	1485.8
GSq	1426.6	1424.9	1424.9	1423.5
$\Delta$ DIC	65.3 **	65.7 **	62.4 **	62.3 **

<sup>a</sup> Difference between DIC values of GSq and GS models. <sup>b</sup> Bayesian Ridge Regression. \* Substantial statistical difference between GSq and GS models. \*\* Strong evidence of statistical difference between GSq and GS models.

**Table 2.** Predictive ability (PA) of all studied traits in *Eucalyptus cladocalyx* according to (i) polygenic background (pedigree information) and all informative markers (GS model) and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model). The PA values for each method correspond to the mean of PA values for 20-folds of cross-validation.

Trait/Model	Bayes A	Bayes B	Bayes C	BRR <sup>a</sup>	$\bar{X}_{PA}$ <sup>b</sup>
Tree height					
GS	0.33	0.32	0.33	0.34	0.33
GSq	0.45	0.44	0.44	0.45	0.44
Diameter at breast height					
GS	0.21	0.23	0.22	0.22	0.22
GSq	0.41	0.41	0.41	0.42	0.41
Stem straightness					
GS	0.39	0.39	0.39	0.39	0.39
GSq	0.40	0.40	0.40	0.39	0.40
Slenderness index					
GS	0.20	0.20	0.21	0.21	0.21
GSq	0.32	0.32	0.31	0.31	0.32
Wood density					
GS	0.27	0.27	0.27	0.28	0.27
GSq	0.43	0.43	0.43	0.43	0.43
Flowering intensity					
GS	0.25	0.25	0.24	0.23	0.24
GSq	0.25	0.25	0.25	0.24	0.25
First bifurcation height					
GS	0.19	0.20	0.20	0.19	0.19
GSq	0.38	0.38	0.39	0.39	0.38

<sup>a</sup> Bayesian Ridge Regression. <sup>b</sup> Corresponds to the average of PA values.

## 2.2. Heritability Estimates

The heritability estimates of the studied traits based on all genomic prediction models are shown in Table 3. The estimates of pedigree-based heritability ( $\hat{h}_a^2$ ) for GS were higher than GSq for all traits, whereas the values of genomic heritability ( $\hat{h}_m^2$  and  $\hat{h}_q^2$ ) were dependent on the Bayesian prediction

method for both models (GS and GSq). Based on the best fitted models (in terms of  $\Delta$ DIC values), the heritability estimates of HT varied between 0.13 and 0.21 ( $\hat{h}_a^2$ ), 0.14 and 0.45 ( $\hat{h}_m^2$ ), and 0.29 and 0.34 ( $\hat{h}_q^2$ ). Based on the models with the lowest DIC values, the heritability estimates of DBH based on pedigree and MTAs ( $\hat{h}_q^2$ ) were similar, which ranged between 0.05 and 0.17, while the heritability estimates based on SNPs ( $\hat{h}_m^2$ ) varied between 0.4 and 0.45. In the case of STR, the estimates of pedigree-based heritability for GS models varied between 0.16 and 0.23, while the estimates of genomic heritability ranged between 0.3 and 0.32. For SLD, the heritability estimates based on pedigree information varied between 0.08 and 0.10. The heritability estimates based on MTAs varied between 0.33 and 0.39, while the heritability estimates based on SNPs ( $\hat{h}_m^2$ ) varied between 0.02 and 0.17. The heritability estimates of WD based on pedigree information (GSq models) ranged between 0.14 and 0.17. The heritability estimates of no-significant QTLs by MTAs analysis varied between 0.12 and 0.24, and those for significant QTLs ranged between 0.27 and 0.31. The heritability estimates of FI based on pedigree with a better goodness-of-fit measure (i.e., the GS model) varied between 0.27 and 0.34, while the genomic heritability estimates ranged between 0.07 and 0.29. Based on the models with the lowest DIC values (i.e., all GSq models), the pedigree-based heritability of BHT was 0.08 in the context of all Bayesian methods (BA, BB, BC and BRR). The heritability estimates based on non-significant SNPs by MTAs analysis varied between 0.04 and 0.13, while those based on the significant QTL varied between 0.41 and 0.45.

**Table 3.** Estimates of heritability of the studied traits for each Bayesian genomic model and effect (i) polygenic background (pedigree information) and all informative markers (GS model), and (ii) polygenic background, QTL-genotype effects (determined by GWAS) and SNP markers that were not associated with any trait (GSq model).  $\hat{h}_a^2$  corresponds to the pedigree-based estimated heritability.  $\hat{h}_m^2$  is the heritability estimate based on a set of markers that were not found to be significantly associated with a trait (GSq) or all SNP markers (GS),  $\hat{h}_q^2$  represents the heritability estimates based on a set of SNPs significantly associated with a trait.

Trait/Model	Bayes A			Bayes B			Bayes C			BRR <sup>a</sup>		
	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$	$\hat{h}_a^2$	$\hat{h}_m^2$	$\hat{h}_q^2$
Tree height												
GS	0.28	0.24	-	0.21	0.45	-	0.22	0.40	-	0.29	0.24	-
GSq	0.16	0.14	0.32	0.18	0.12	0.29	0.13	0.27	0.29	0.14	0.17	0.34
Diameter at breast height												
GS	0.20	0.14	-	0.15	0.37	-	0.15	0.39	-	0.19	0.24	-
GSq	0.11	0.05	0.44	0.09	0.15	0.42	0.09	0.17	0.40	0.10	0.11	0.45
Stem straightness												
GS	0.23	0.32	-	0.18	0.31	-	0.16	0.30	-	0.21	0.30	-
GSq	0.18	0.18	0.01	0.14	0.37	0.01	0.14	0.32	0.01	0.18	0.19	0.013
Slenderness index												
GS	0.19	0.12	-	0.16	0.27	-	0.15	0.33	-	0.18	0.21	-
GSq	0.09	0.02	0.39	0.08	0.05	0.36	0.08	0.17	0.33	0.09	0.10	0.35
Wood density												
GS	0.25	0.28	-	0.18	0.50	-	0.19	0.45	-	0.21	0.42	-
GSq	0.17	0.13	0.31	0.17	0.18	0.27	0.14	0.24	0.27	0.17	0.12	0.31
Flowering intensity												
GS	0.34	0.07	-	0.32	0.10	-	0.27	0.29	-	0.33	0.13	-
GSq	0.30	0.06	0.00	0.29	0.06	0.00	0.27	0.20	0.00	0.31	0.10	0.002
First bifurcation height												
GS	0.20	0.05	-	0.19	0.12	-	0.16	0.27	-	0.19	0.14	-
GSq	0.08	0.04	0.44	0.08	0.11	0.42	0.08	0.13	0.41	0.08	0.06	0.45

<sup>a</sup> Bayesian Ridge Regression.

### 3. Discussion

#### 3.1. Marker-Trait Associations for All Studied Traits

*Eucalyptus cladocalyx* is not a close relative of other eucalypts. Brooker [22] placed *E. cladocalyx* in the monophyletic section *Sejunctae* [23]. As the marker panel we used had been developed for the most widely-planted species which all fall within sections *Maidenaria*, *Exsertaria* and *Latoangulatae*, a lower rate of cross-species amplification of SSR (Simple Sequence Repeats) can be expected in *E. cladocalyx* [24] due to differentiation among widely-distant sections. Despite the low availability of SNP markers, the genetic diversity values were similar to other previous studies of natural populations of *E. cladocalyx* [23,25]. The strong genetic differentiation in three clusters had also been previously reported [23,25,26].

Ninety significant marker-trait associations (MTAs) were detected for the seven target traits, which were subsequently used for the GSq approach. In accordance with this, several studies have previously identified genomic regions explaining part of the phenotypic variation of growth-related traits in *E. cladocalyx* (e.g., Ballesta et al. [25], Arriagada et al. [26], Maldonado et al. [27], Valenzuela et al. [28]). For HT, DBH and SLD, the MTAs were mainly located on chromosomes Chr2 (5 and 4 MTAs for HT and DBH, respectively), Chr6 (3 and 6 MTAs for DBH and SLD, respectively), Chr8 (4 MTAs for SLD) and Chr10 (4 MTAs for SLD). In agreement with these results, Maldonado et al. [27] detected one QTL, based on Simple Sequence Repeat (SSR) markers, located on linkage group LG6 explaining up to 27% of the phenotypic variation of DBH. In addition, Arriagada et al. [26] reported SSR markers associated with HT and DBH, located on the linkage groups LG6, LG8 and LG10, explaining up to 23% of the phenotypic variation.

Flowering components are target traits in breeding programs of some species of *Eucalyptus*, as the flowers provide a reliable source for honey production [26,29]. In dry regions of Chile and South Africa, prolific flowering from forest plantations of *E. cladocalyx* and other eucalypts is particularly advantageous for the supply of honey [30]. In the present study, almost all associations (4/5 MTAs) for flowering intensity were located on chromosome Chr2, which is in accordance with Missiaggia et al. [31], who reported a major QTL located on chromosome Chr2 (*Eef1*) controlling the early flowering in *Eucalyptus grandis*. In *E. cladocalyx*, previous studies have reported that the flowering intensity and early flowering have a positive genetic correlation and common QTLs controlling the phenotypic variation of both traits [29,32]. Interestingly, according to linkage disequilibrium analyses, only two significant SNPs (MTAs for FI) were in disequilibrium, which covered a genomic region of 17,849 bp. For WD, BHT and STR, the MTAs were mainly located on chromosomes Chr2 (WD), Chr5 (BHT) and Chr8 (STR, WD and BHT). In accordance with this, Valenzuela et al. [28] detected a QTL located on Chr2 explaining 8% of the phenotypic variation of WD in *E. cladocalyx*.

#### 3.2. Comparison between Genomic Prediction Models

Several studies have explored the potential of the selection based on genomic tools in forest species [33–36], including *E. cladocalyx* [12,37,38]. According to the results, GSq models outperformed traditional GS models in terms of predictive ability when at least ten significant marker-trait associations were included in GSq. In addition, another important finding of this study was that the GS and GSq models that include the pedigree information (i.e., pedigree information as a relationship matrix), outperformed the model based solely on SNP marker effects (in terms of goodness-of-fit), revealing the importance of polygenic effects in the prediction model based on low-density markers; an aspect emphasized by Silva et al. [19]. For instance, the predictive ability of flowering intensity based on only the SNP marker panel was three times lower than those based on GS or GSq.

According to De Los Campos et al. [39], Bayes A, Bayes B, Bayes C and BRR methods can improve the predictive ability in genome-based evaluations, but these prediction models could have overfitting problems when the ratio of number of markers and individuals is over 50 [40]. To overcome this, the use of genomic relationship matrices between individuals into genomic prediction models could

beneficially capture general information and reduce the dimensionality problem [41,42], while exploring regions in linkage disequilibrium with QTLs [43]. Interestingly, we found that the GSq models had a better fit than GS models and, at the same time increased the predictive ability of BHT, DBH, SLD and WD. Notably, these benefits were only detected for the prediction of traits with greater than ten MTAs, while for the traits with a lower number of MTAs (i.e., HT, FI and STR), the GS models had a better fit than GSq models. The predictive ability for STR and FI was similar between GSq and GS models. It is worth mentioning that the MTAs detected by the classical linkage analysis explained relatively low values of the total phenotypic variation; which is in accordance with the genetic architecture of quantitative traits. According to Silva et al. [19], the superior performance of the models that include QTL information compared with GS model can be due to GSq model exploiting relevant regions not directly considered in GS model, improving the performance of the prediction model. Additionally, previous studies have reported that a pre-selection of SNPs or the use of genome-wide association analyses to identify and rank markers could increase predictive ability [38,44,45].

Although the analytical assumptions differ among studied Bayesian genomic models, the predictive ability of the studied traits was not severely different among them. In accordance with our findings, several studies have reported that the predictive ability did not differ between methods in forest tree species, especially for growth and wood quality related traits [15,33,36,46–48]. In the present study, the main differences in predictive ability values were observed between the GSq and GS methods, so that the superiority of GSq (or GS) to predict the studied traits was conserved for any Bayesian genomic model.

### 3.3. Heritability Estimates

According to the goodness-of-fit measures, the GSq model outperformed the GS model for BHT, DBH, SLD and WD, whereas GS offered better model fit compared with GSq for STR and FI. Additionally, for HT, DBH, SLD, WD and BHT, the genomic heritability estimates based on MTAs were higher than the estimates based on SNPs (not significantly associated with a trait). On the other hand, the genomic heritability estimates based on SNP markers (not associated) considering BB and BC method, were higher than those based on BA and BRR methods (for all traits, except HT). BB and BC models involve variable selection procedures, which favor the selection based on major effect markers/genes [49]. Notably, other studies have confirmed that some whole-regression methods could overestimate the heritability values [50], and therefore, these findings should be interpreted with caution. Overall, all target traits had pedigree-based heritability estimates from low to moderate ( $h^2 = 0.08–0.34$ ), which are in accordance with the range usually expected for forest tree growth, flowering and stem quality related traits, including *E. cladocalyx* [28,29,51–54].

The heritability estimates based on MTAs for STR and FI were lower than those based on pedigree information or non-significant SNPs from the MTA analysis. In tree species, several studies had reported genomic regions explaining a high percent of phenotypic variation of STR and FI. For example, Arriagada et al. [26] reported one QTL explaining up to 15% of the total variation of STR in *E. cladocalyx*. In a meta-analysis, Hall et al. [55] confirmed that phenological traits are highly heritable and controlled by key genomic regions in trees. In fact, Missiaggia et al. [31] reported a key region on chromosome Chr2 (*Eef1*) controlling the early flowering in *E. grandis*. In this context, low-density SNP panels could limit the probability to detect key regions explaining the phenotypic variation of these traits. In addition, these results are supported by the fact that the GS model outperformed the GSq model for STR and FI (in terms of goodness-of-fit), which means that the total variation of STR and FI is better explained by the classic prediction model (i.e.,  $\sigma_a^2$  and  $\sigma_m^2$ ).

## 4. Materials and Methods

### 4.1. Plant Material and Phenotypic Evaluation

A genomic selection study was performed in a long-term open-pollinated progeny trial comprising 49 families of *E. cladocalyx* established in 2001 in northern Chile (locality of Los Vilos; 31°54' S; 71°27' W; 167 m.a.s.l.). The climate is classified as predominantly arid, according to the De Martonne aridity index [26]. Trees were arranged in a randomized complete block design with 30 blocks and single-tree plots (total of 1470 trees). Trees were planted at 2 m spacing within rows and 3 m between rows (~1667 trees ha<sup>-1</sup>). The following quantitative traits were measured in 17 years-old trees: diameter at breast height (DBH), total tree height (HT), first bifurcation height (BHT), wood density (WD), stem straightness (STR) and slenderness index (SLD). The BHT was rated on a scale of five levels, in which a value of 1 is assigned to trees with a loss of the central axis in the first fifth of the tree's height, a value of 2 indicates that a loss of the central axis occurs in the second fifth of the tree, a value of 3 indicates that a loss of the central axis occurs in the third fifth of the tree, a value of 4 implies a loss of the central axis in the fourth fifth of the tree's height, and a value of 5 implies that a loss of the central axis in the last fifth of the tree height or does not show loss of the apical axis (modified scale by Bush et al. [54]). WD was measured indirectly according to Valenzuela et al. [28]. STR was measured in the first two-thirds of the total height of tree and was considered as ordinal variables with four levels [53], in which a value of 0 if the stem was strongly twisted, a value of 1 if the stem presents moderate levels of curvature, a value of 2 if the stem was slightly curved, and 3 if the stem was completely straight. The SLD was calculated as the ratio between HT (m) and DBH (m). Additionally, flowering intensity (FI) was measured using a scale that ranged from 0 to 3, in 18 years-old trees, according to Arriagada et al. [26], where a value of 0 means absence of flowers, buds and/or capsules, a value of 1 sparse flowers on a small part of crown, a value of 2 if the flowers/capsules/buds were covering the half of the crown, and a value of 3 for trees with numerous flowers on the whole crown.

### 4.2. DNA Extraction and Tree Genotyping

DNA for tree genotyping was isolated from leaf tissues of 480 individuals according to Porebsky et al. [56] and Doyle and Doyle [57]. On average, 10 individuals per family were randomly selected to be genotyped using the Illumina Infinium EUChip60K SNP array [7]. The SNP data were filtered for SNP call rate score >0.7 and minor allele frequency (MAF) >0.05. The missing data were imputed using the LD-kNNi method in Tassel 5.2 [58]. Linkage disequilibrium between marker pairs (MTAs at the same chromosome) was calculated using TASSEL version 5.2.

### 4.3. Genomic Prediction Models

The following four Bayesian whole-genome regression models were used for the estimation of SNP marker effects, variance components and genomic heritability: Bayes A ([59]; BA), Bayes B ([59]; BB), Bayes C $\pi$  ([60]; BC) and Bayesian Ridge Regression ([61]; BRR). Due to the low density of markers found in this studied population, the following two approaches were used to predict the studied traits: traditional genomic prediction model (GS), considering all informative markers (~3.8 K SNP) from the commercial SNP panel, and a combination of traditional GS and QTL information (GSq). The GS model is defined as:

$$\mathbf{y}^* = \mathbf{1}\mu + \sum_{i=1}^m \mathbf{x}_i \mathbf{m}_i + \mathbf{Z}\mathbf{a} + \boldsymbol{\varepsilon} \quad (1)$$

where  $\mathbf{y}^*$  is the vector of phenotypic records pre-corrected for the effects of block and genetic structure [62,63]. Respectively,  $\mathbf{1}$  and  $\mu$  are vectors of ones and overall mean.  $\mathbf{m}_i$  corresponds to the additive genetic effect of the  $i$ -th marker, with  $m$  as the number of markers.  $\mathbf{x}_i$  is the incidence vector of each marker (codified as: AA = 0, AB = 1 and BB = 2).  $\mathbf{a}$  corresponds to polygenic effects, with  $\mathbf{a} \sim \mathbf{N}(0, \sigma_a^2 \mathbf{A})$ , and  $\mathbf{Z}$  is the incidence matrix related to polygenic effects. The coefficients of relationship matrix

were adjusted according to Bush et al. [54]. Finally,  $\epsilon$  is the residual vector,  $\epsilon \sim \mathbf{N}(0, \hat{\sigma}_e^2 \mathbf{I}_n)$ . The genomic predictions were performed using BGLR package in R [64]. Specifically, 1,000,000 iterations of Markov Chain Monte Carlo simulations were used in all genomic prediction models, with a burn-in period of 100,000.

The GSq model consisted of two steps, in which the first step involves a marker-trait association (MTA) analysis for QTL detection according to Silva et al. [19]. The MTA analysis was conducted using the following mixed linear model:

$$\mathbf{y}^* = \mathbf{S}\mathbf{a} + \mathbf{Q}\mathbf{v} + \mathbf{Z}\mathbf{u} + \epsilon \quad (2)$$

where  $\mathbf{y}^*$  is a vector of adjusted phenotypic observations (by the block effect).  $\mathbf{S}$ ,  $\mathbf{Q}$  and  $\mathbf{Z}$  correspond to the incidence matrices for  $\mathbf{a}$ ,  $\mathbf{v}$  and  $\mathbf{u}$ , respectively.  $\mathbf{a}$  and  $\mathbf{v}$  are the vectors of SNP effects (fixed) and population structure effects (fixed).  $\mathbf{u}$  corresponds to a vector of polygenic effects (random), and  $\epsilon$  is the residual vector. The variances of  $\mathbf{u}$  and  $\epsilon$  are  $\text{Var}(\mathbf{u}) = 2\mathbf{K}\sigma_g^2$  and  $\text{Var}(\epsilon) = \mathbf{R}\sigma_e^2$ , respectively, where  $\mathbf{K}$  is the kinship coefficient matrix, which was estimated using the program TASSEL [58]. For this analysis, the significance threshold of 0.001 and a false discovery rate (FDR) <10% were applied to test for significant MTAs according to Uchiyama et al. [65]. The GWAS analysis was carried out using rrBLUP package in R v. 3.5 [66]. The population structure was assessed by Bayesian model-based clustering in STRUCTURE software v. 2.3.4 [67]. In addition, the SNP data was used to estimate the genetic diversity of the genotyped population using GenAlex v.6.5 [68].

The second step of the GSq approach consists of the compilation of traditional GS model (Equation (1)) and QTL-genotype effect ( $\mathbf{q}$ ) from the association analysis, which is expressed as:

$$\mathbf{y}^* = \mathbf{1}\mu + \sum_{i=1}^m \mathbf{x}_i \mathbf{m}_i + \mathbf{Z}\mathbf{a} + \mathbf{Z}\mathbf{q} + \epsilon \quad (3)$$

where  $\mathbf{y}^*$  is the vector of phenotypic records pre-corrected for the effects of block and genetic structure, and  $\mathbf{1}$  and  $\mu$  are vectors of ones and overall mean, respectively.  $\mathbf{m}_i$  corresponds to the additive genetic effect of the  $i$ -th marker that were not found to be significantly associated with a trait.  $\mathbf{x}_i$  is the incidence vector of each marker (AA = 0, AB = 1 and BB = 2). The variance of  $m$  ( $\hat{\sigma}_m^2$ ) depends on the Bayesian model implemented.  $\mathbf{Z}$  is the incidence matrix of polygenic ( $\mathbf{a}$ ) and QTL-genotype ( $\mathbf{q}$ ) effects assuming  $\mathbf{a} \sim \mathbf{N}(0, \hat{\sigma}_a^2 \mathbf{A})$  and  $\mathbf{q} \sim \mathbf{N}(0, \hat{\sigma}_q^2 \mathbf{Q})$ , respectively. The  $\mathbf{Q}$  is a covariance matrix, in which the elements are the probabilities that individuals are identical by descent based on significant SNP markers according to the MTA analysis.

#### 4.4. Heritability Estimates

For the GSq model, the heritability estimates were obtained as follows:

$$\hat{h}_a^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_m^2 + \hat{\sigma}_q^2 + \hat{\sigma}_e^2} \quad (4)$$

$$\hat{h}_m^2 = \frac{\hat{\sigma}_m^2}{\hat{\sigma}_a^2 + \hat{\sigma}_m^2 + \hat{\sigma}_q^2 + \hat{\sigma}_e^2} \quad (5)$$

$$\hat{h}_q^2 = \frac{\hat{\sigma}_q^2}{\hat{\sigma}_a^2 + \hat{\sigma}_m^2 + \hat{\sigma}_q^2 + \hat{\sigma}_e^2} \quad (6)$$

where  $\hat{h}_a^2$ ,  $\hat{h}_m^2$  and  $\hat{h}_q^2$  correspond to heritability estimates based on pedigree information, a set of markers that were not found to be significantly associated with a trait and a set of SNPs significantly associated with a trait (MTAs), respectively.  $\hat{\sigma}_a^2$  is the variance due to the additive polygenic effect,  $\hat{\sigma}_m^2$  corresponds

to the marker effect variance,  $\hat{\sigma}_q^2$  is the variance of markers significantly associated with a trait, and  $\hat{\sigma}_e^2$  is the residual variance.

For traditional GS model, the term  $\hat{h}_m^2$  correspond to the heritability estimates based on all SNPs ( $n = 3879$ ). In the cases of BC and BRR methods,  $\hat{\sigma}_m^2$  was calculated as:  $\hat{\sigma}_m^2 = 2\hat{\sigma}_{SNP}^2 \sum_{i=1}^n \hat{p}_i(1 - \hat{p}_i)$ , in which  $\hat{\sigma}_{SNP}^2$  corresponds to a common variance for SNP markers and  $\hat{p}_i$  is the MAF of the  $i$ -th marker. The  $\hat{\sigma}_m^2$  term for BA and BB models was estimated as  $\hat{\sigma}_m^2 = 2 \sum_{i=1}^n \hat{p}_i(1 - \hat{p}_i)\hat{\sigma}_{SNP_i}^2$ , in which  $\hat{\sigma}_{SNP_i}^2$  corresponds to the variance due to the  $i$ -th marker.

#### 4.5. Comparison between Genomic Prediction Models

The Bayesian prediction models were compared in terms of goodness-of-fit, by using Deviance Information Criterion (DIC) [69], and the predictive ability (PA). The DIC is defined by the following expression:

$$DIC = \bar{D} + pD \quad (7)$$

where  $D$  is a Bayesian measure of model fit, which is defined as the posterior expectation of the deviance ( $\bar{D} = E_{\theta/y}[-2 \ln f(y/\theta)]$ );  $pD$  is the effective number of parameters. A  $DIC$  difference of more than 10 between two competitive models (GSq and GS models) was considered to be supported against a model with higher  $DIC$ ; a  $DIC$  difference between 5 and 10 was considered as substantial difference between models, and a difference less than 5 was considered as not significant.

The predictive ability of each model was calculated as the correlation between the pre-corrected observations ( $\mathbf{y}^*$ ) from validation dataset and the estimated breeding value ( $\hat{\mathbf{y}}^*$ ). A total of 20-fold cross-validation was used to evaluate the predictive ability of all models. The  $\hat{\mathbf{y}}^*$  for GS and GSq models was calculated as  $\hat{\mathbf{y}}^* = \sum_{i=1}^n \mathbf{x}_i \hat{m}_i + \mathbf{Z}\hat{\mathbf{u}}$  and  $\hat{\mathbf{y}}^* = \sum_{i=1}^n \mathbf{x}_i \hat{m}_i + \mathbf{Z}\hat{\mathbf{u}} + \mathbf{Z}\hat{\mathbf{q}}$ , respectively.

## 5. Conclusions

In this study, we evaluated the performance of Bayesian genomic models that include the genetic background (pedigree) and QTL information from GWAS analysis, which was specially implemented in the context of a low-density SNP markers. Importantly, predictive abilities were superior in GSq when the proportion of the variance explained by the significant marker-trait associations (MTAs) was higher than those explained by the polygenic background and SNP markers (that were not found to be significantly associated with a trait). Therefore, we emphasized and hypothesized that both the number of associations and/or the percentage of variation explained by the MTAs are determinants in the effectiveness of the GSq method. This approach can be particularly useful for plant/tree species poorly represented in the high-density SNP arrays, developed for economically important species or when high-density marker panels are not available.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2223-7747/9/1/99/s1>, Figure S1: Ideogram representing the SNP density in a *Eucalyptus cladocalyx* population genotyped by the 60K SNP array, Table S1: Summary of the single nucleotide polymorphism (SNP) density in *Eucalyptus cladocalyx*.

**Author Contributions:** Conceptualization, F.M. and P.B.; methodology, P.B., F.F.S. and F.M.; software, P.B.; validation, F.M. and D.B.; formal analysis, F.M. and P.B.; investigation, F.M. and P.B.; resources, F.M.; data curation, F.M.; writing—original draft preparation, F.M. and P.B.; writing—review and editing, F.M., D.B. and F.F.S.; supervision, F.M., D.B. and F.F.S.; project administration, F.M.; funding acquisition, F.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by FONDECYT (grant number 1170695) and Hacienda Agrícola Caracas—Los Vilos, Chile.

**Acknowledgments:** The authors thank FONDECYT (grant number 1170695) and Eduardo Collantes. Paulina Ballesta thanks CONICYT-PCHA/Doctorado Nacional/2016-21160624.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Viana, J.; Pereira, H.D.; Mundim, G.B.; Piepho, H.-P.; Fonseca e Silva, F. Efficiency of genomic prediction of non-assessed single crosses. *Heredity* **2018**, *120*, 283–295. [[CrossRef](#)] [[PubMed](#)]
2. Singh, D.; Wang, X.; Kumar, U.; Gao, L.; Noor, M.; Imtiaz, M.; Singh, R.P.; Poland, J. High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. *Front. Plant Sci.* **2019**, *10*, 394. [[CrossRef](#)] [[PubMed](#)]
3. Contreras-Soto, R.I.; Mora, F.; De Oliveira, M.A.R.; Higashi, W.; Scapim, C.A.; Schuster, I. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS ONE* **2017**, *12*, e0171105. [[CrossRef](#)]
4. Maldonado, C.; Mora, F.; Scapim, C.A.; Coan, M. Genome-wide haplotype-based association analysis of key traits of plant lodging and architecture of maize identifies major determinants for leaf angle: hapLA4. *PLoS ONE* **2019**, *14*, e0212925. [[CrossRef](#)] [[PubMed](#)]
5. Senhorinho, H.J.C.; Coan, M.M.D.; Marino, T.P.; Kuki, M.C.; Barth-Pinto, R.S.; Scapim, C.A.; Holland, J.B. Genomic-Wide Association Study of Popping Expansion in Tropical Popcorn and Field Corn Germplasm. *Crop Sci.* **2019**, *59*, 2007–2019. [[CrossRef](#)]
6. Mafra, G.S.; Do Amaral Júnior, A.T.; Almeida, F.J.E.D.; Vivas, M.; Araújo Diniz-Santos, P.H.; Saltires-Santos, J.; Ferreira-Pena, G.; De Lima, V.J.; Kamphorst, S.H.; De Oliveira, F.T.; et al. SNP-based mixed model association of growth-and yield-related traits in popcorn. *PLoS ONE* **2019**, *14*, e0218552. [[CrossRef](#)]
7. Silva-Junior, O.B.; Faria, D.A.; Grattapaglia, D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytol.* **2015**, *206*, 1527–1540. [[CrossRef](#)]
8. Torres-Dini, D.; Nunes, A.C.P.; Aguiar, A.; Nikichuk, N.; Centurión, C.; Cabrera, M.; Moraes, M.L.T.; Resende, M.D.V.; Sebbenn, A.M. Clonal selection of *Eucalyptus grandis* x *Eucalyptus globulus* for productivity, adaptability, and stability, using SNP markers. *Silvae Genet* **2016**, *65*, 30–38. [[CrossRef](#)]
9. Klápště, J.; Suontama, M.; Telfer, E.; Graham, N.; Low, C.; Stovold, T.; McKinley, R.; Dumgey, H. Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of *Eucalyptus nitens*. *PLoS ONE* **2017**, *12*, e0185137. [[CrossRef](#)]
10. Durán, R.; Isik, F.; Zapata-Valenzuela, J.; Balocchi, C.; Valenzuela, S. Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genet Genomes* **2017**, *13*, 74. [[CrossRef](#)]
11. Suontama, M.; Klápště, J.; Telfer, E.; Graham, N.; Stovold, T.; Low, C.; McKinley, R.; Dumgey, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* **2019**, *122*, 370. [[CrossRef](#)] [[PubMed](#)]
12. Ballesta, P.; Maldonado, C.; Pérez-Rodríguez, P.; Mora, F. SNP and Haplotype-Based Genomic Selection of Quantitative Traits in *Eucalyptus globulus*. *Plants* **2019**, *8*, 331. [[CrossRef](#)] [[PubMed](#)]
13. Aguirre, N.C.; Filippi, C.V.; Zaina, G.; Rivas, J.G.; Acuña, C.V.; Villalba, P.V.; García, M.N.; González, S.; Rivarola, M.; Maetinez, M.C. Optimizing ddRADseq in non-model species: A case study in *Eucalyptus dunnii* Maiden. *Agronomy* **2019**, *9*, 484. [[CrossRef](#)]
14. Pryce, J.E.; Johnston, J.; Hayes, B.J.; Sahana, G.; Weigel, K.A.; McParland, S.; Spurlock, D.; Krattenmacher, N.; Spelman, R.J.; Wall, E.; et al. Imputation of genotypes from low density (50,000 markers) to high density (700,000 markers) of cows from research herds in Europe, North America, and Australasia using 2 reference populations. *J. Dairy Sci.* **2014**, *97*, 1799–1811. [[CrossRef](#)] [[PubMed](#)]
15. Müller, B.S.; Neves, L.G.; De Almeida, F.J.E.; Resende, M.F.R., Jr.; Muñoz, P.R.; Dos Santos, P.E.T.; Paludzyszyn Filho, E.; Kirst, M.; Grattapaglia, D. Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genomes* **2017**, *18*, 524.
16. Porto-Neto, L.R.; Kijas, J.W.; Reverter, A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet. Sel. Evol.* **2014**, *46*, 22. [[CrossRef](#)]
17. Bolormaa, S.; Gore, K.; van der Werf, J.H.J.; Hayes, B.J.; Daetwyler, H.D. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. *Anim. Genet.* **2015**, *46*, 544–556. [[CrossRef](#)]

18. Wu, X.-L.; Xu, J.; Feng, G.; Wiggans, G.R.; Taylor, J.F.; He, J.; Qian, C.; Qiu, J.; Simpson, B.; Walker, J.; et al. Optimal Design of Low-Density SNP Arrays for Genomic Prediction: Algorithm and Applications. *PLoS ONE* **2016**, *11*, e0161719. [[CrossRef](#)]
19. Silva, F.F.; Jerez, E.A.Z.; De Resende, M.D.V.; Soriano-Viana, M.; Ferreira-Azevedo, C.; Lopes, P.S.; Nascimento, M.; Oliveira de Lima, R.; Facioni-Guimaraes, S.E. Bayesian model combining linkage and linkage disequilibrium analysis for low density-based genomic selection in animal breeding. *J. Appl. Anim. Res.* **2018**, *46*, 873–878. [[CrossRef](#)]
20. Rice, B.; Lipka, A.E. Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *Plant Genome* **2019**, *12*. [[CrossRef](#)]
21. Bernardo, R. Genomewide selection when major genes are known. *Crop Sci.* **2014**, *54*, 68–75. [[CrossRef](#)]
22. Brooker, M.I.H. A new classification of the genus *Eucalyptus* L'Her. (Myrtaceae). *Aust. Syst. Bot.* **2000**, *13*, 79–148. [[CrossRef](#)]
23. McDonald, M.W.; Rawlins, M.; Butchet, P.A.; Bell, C. Regional divergence and inbreeding in *Eucalyptus cladocalyx* (Myrtaceae). *Aust. J. Bot.* **2003**, *51*, 393–403. [[CrossRef](#)]
24. Mora, F.; Arriagada, O.; Ballesta, P.; Ruiz, E. Genetic diversity and population structure of a drought-tolerant species of *Eucalyptus*, using microsatellite markers. *J. Plant Biochem. Biotechnol.* **2017**, *26*, 274–281. [[CrossRef](#)]
25. Ballesta, P.; Mora, F.; Contreras-Soto, R.I.; Ruiz, E.; Perret, S. Analysis of the genetic diversity of *Eucalyptus cladocalyx* (sugar gum) using ISSR markers. *Acta Sci. Agron.* **2015**, *37*, 133–140. [[CrossRef](#)]
26. Arriagada, O.; Do Amaral Junior, A.T.; Mora, F. Thirteen years under arid conditions: Exploring marker-trait associations in *Eucalyptus cladocalyx* for complex traits related to flowering, stem form and growth. *Breed. Sci.* **2018**, *68*, 367–374. [[CrossRef](#)]
27. Maldonado, C.; Contreras-Soto, R.I.; Gerhardt, I.F.S.; Do Amaral Júnior, A.T.; Mora, F. Stable marker-trait associations for growth across different ages in *Eucalyptus cladocalyx* through the use of microsatellites. *Sci. For.* **2018**, *46*, 367–376. [[CrossRef](#)]
28. Valenzuela, C.E.; Ballesta, P.; Maldonado, C.; Baettig, R.; Arriagada, O.; Mafra, G.S.; Mora, F. Bayesian Mapping Reveals Large-Effect Pleiotropic QTLs for Wood Density and Slenderness Index in 17-Year-Old Trees of *Eucalyptus cladocalyx*. *Forests* **2019**, *10*, 241. [[CrossRef](#)]
29. Cané-Retamales, C.; Mora, F.; Vargas-Reeve, F.; Perret, S.; Contreras-Soto, R. Bayesian threshold analysis of breeding values, genetic correlation and heritability of flowering intensity in *Eucalyptus cladocalyx* under arid conditions. *Euphytica* **2011**, *178*, 177–183. [[CrossRef](#)]
30. De Lange, W.J.; Veldtman, R.; Allsopp, M.H. Valuation of pollinator forage services provided by *Eucalyptus cladocalyx*. *J. Environ. Manag.* **2013**, *125*, 12–18. [[CrossRef](#)]
31. Missiaggia, A.A.; Piacuzzi, A.L.; Grattapaglia, D. Genetic mapping of Eef1, a major effect QTL for early flowering in *Eucalyptus grandis*. *Tree Genet. Genomes* **2005**, *1*, 79. [[CrossRef](#)]
32. Contreras-Soto, R.; Ballesta, P.; Ruiz, E.; Mora, F. Identification of ISSR markers linked to flowering traits in a representative sample of *Eucalyptus cladocalyx*. *J. For. Res.* **2016**, *27*, 239–245. [[CrossRef](#)]
33. Ratcliffe, B.; El-Dien, O.G.; Klápště, J.; Porth, I.; Chen, C.; Jaquish, B.; El-Kasaby, Y.A. A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity* **2015**, *115*, 547–555. [[CrossRef](#)] [[PubMed](#)]
34. Gamal-Dien, O.; Ratcliffe, B.; Klápště, J.; Chen, C.; Porth, I.; El-Kasaby, Y.A. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genom.* **2015**, *16*, 370.
35. Lenz, P.R.; Beaulieu, J.; Mansfield, S.D.; Clément, S.; Despons, M.; Bousquet, J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genom.* **2017**, *18*, 335. [[CrossRef](#)]
36. Chen, Z.Q.; Baison, J.; Pan, J.; Karlsson, B.; Andersson, B.; Westin, J.; García-Gil, M.R.; Wu, H.X. Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BMC Genom.* **2018**, *19*, 946. [[CrossRef](#)]
37. Bush, D.; Thumma, B. Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. *Tree Genet. Genomes* **2013**, *9*, 741–752. [[CrossRef](#)]
38. Ballesta, P.; Serra, N.; Guerra, F.; Hasbún, R.; Mora, F. Genomic prediction of growth and stem quality traits in *Eucalyptus globulus* Labill at its southernmost distribution limit in Chile. *Forests* **2018**, *9*, 779. [[CrossRef](#)]
39. De los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P.L. Whole-Genome regression and prediction methods applied to plant and animal breeding. *Genetics* **2013**, *193*, 327–345. [[CrossRef](#)]

40. González-Recio, O.; Rosa, G.J.; Gianola, D. Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest. Sci.* **2014**, *166*, 217–231. [[CrossRef](#)]
41. Solberg, T.R.; Sonesson, A.K.; Woolliams, J.A.; Meuwissen, T.H.E. Reducing dimensionality for prediction of genome-wide breeding values. *Genet. Sel. Evol.* **2009**, *41*, 29. [[CrossRef](#)] [[PubMed](#)]
42. Du, C.; Wei, J.; Wang, S.; Jia, Z. Genomic selection using principal component regression. *Heredity* **2018**, *121*, 12–23. [[CrossRef](#)] [[PubMed](#)]
43. Habier, D.; Fernando, R.L.; Dekkers, J.C. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **2007**, *177*, 2389–2397. [[CrossRef](#)] [[PubMed](#)]
44. Macciotta, N.P.; Gaspa, G.; Steri, R.; Pieramati, C.; Carnier, P.; Dimauro, C. Pre-selection of most significant SNPs for the estimation of genomic breeding values. *BMC Proc.* **2009**, *3*, 14. [[CrossRef](#)] [[PubMed](#)]
45. Arojju, S.K.; Conaghan, P.; Barth, S.; Milbourne, D.; Casler, M.D.; Hodkinson, T.R.; Michel, T.; Byrne, S.L. Genomic prediction of crown rust resistance in *Lolium perenne*. *BMC Genet.* **2018**, *19*, 35. [[CrossRef](#)]
46. Resende, M.F., Jr.; Muñoz, P.; Resende, M.D.; Garrick, D.J.; Fernando, R.L.; Davis, M.J.; Jokela, E.J.; Martin, T.A.; Peter, G.F.; Kirst, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* **2012**, *190*, 1503–1510. [[CrossRef](#)]
47. Beaulieu, J.; Doerksen, T.K.; MacKay, J.; Rainville, A.; Bousquet, J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genom.* **2014**, *15*, 1048. [[CrossRef](#)]
48. Thistlethwaite, F.R.; Ratcliffe, B.; Klápště, J.; Porth, I.; Chen, C.; Stoehr, M.U.; El-Kassaby, Y.A. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genom.* **2017**, *18*, 930. [[CrossRef](#)]
49. Wolfe, M.D.; Del Carpio, D.P.; Alabi, O.; Ezenwaka, L.C.; Ikeogu, U.N.; Kayondo, I.S.; Lozano, R.; Okeke, U.G.; Ozimati, A.A.; Williams, E.; et al. Prospects for genomic selection in cassava breeding. *Plant Genome* **2017**, *10*, 1–9. [[CrossRef](#)]
50. De los Campos, G.; Sorensen, D.; Gianola, D. Genomic heritability: What is it? *PLoS Genet.* **2015**, *11*, e1005048. [[CrossRef](#)]
51. Mora, F.; Gleadow, R.; Perret, S.; Scapim, C.A. Genetic variation for early flowering, survival and growth in sugar gum (*Eucalyptus cladocalyx* F. Muell) in southern Atacama Desert. *Euphytica* **2009**, *169*, 335–344. [[CrossRef](#)]
52. Bush, D.; McCarthy, K.; Meder, R. Genetic variation of natural durability traits in *Eucalyptus cladocalyx* (sugar gum). *Ann. For. Sci.* **2011**, *68*, 1057. [[CrossRef](#)]
53. Vargas-Reeve, F.; Mora, F.; Perret, S.; Scapim, C.A. Heritability of stem straightness and genetic correlations in *Eucalyptus cladocalyx* in the semi-arid region of Chile. *Crop Breed. Appl. Biotechnol.* **2013**, *13*, 107–112. [[CrossRef](#)]
54. Bush, D.; Kain, D.; Kanowski, P.; Matheson, C. Genetic parameter estimates informed by a marker-based pedigree: A case study with *Eucalyptus cladocalyx* in southern Australia. *Tree Genet. Genomes* **2015**, *11*, 798. [[CrossRef](#)]
55. Hall, D.; Hallingbäck, H.R.; Wu, H.X. Estimation of number and size of QTL effects in forest tree traits. *Tree Genet. Genomes* **2016**, *12*, 110. [[CrossRef](#)]
56. Porebski, S.; Bailey, L.G.; Baum, B.R. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **1997**, *15*, 8–15. [[CrossRef](#)]
57. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.
58. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]
59. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829.
60. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [[CrossRef](#)]
61. Gianola, D. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* **2013**, *194*, 573–596. [[CrossRef](#)] [[PubMed](#)]
62. Asoro, F.G.; Newell, M.A.; Beavis, W.D.; Scott, M.P.; Jannink, J.-L. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* **2011**, *4*, 132–144. [[CrossRef](#)]

63. Lorenz, A.J.; Smith, K.P.; Jannink, J.L. Potential and optimization of genomic selection for *Fusarium* head blight resistance in six-row barley. *Crop Sci.* **2012**, *52*, 1609–1621. [[CrossRef](#)]
64. Pérez, P.; De Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [[CrossRef](#)] [[PubMed](#)]
65. Uchiyama, K.; Iwata, H.; Moriguchi, Y.; Ujino-Ihara, T.; Ueno, S.; Taguchi, Y.; Tsuboruma, M.; Mishima, K.; Iki, T.; Watanabe, A.; et al. Demonstration of genome-wide association studies for identifying markers for wood property and male strobili traits in *Cryptomeria japonica*. *PLoS ONE* **2013**, *8*, e79866. [[CrossRef](#)] [[PubMed](#)]
66. Endelman, J.B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **2011**, *4*, 250–255. [[CrossRef](#)]
67. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
68. Peakall, R.O.D.; Smouse, P.E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [[CrossRef](#)]
69. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc.* **2002**, *64*, 583–639. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).