plants

MDPI

Article

SNP and Haplotype-Based Genomic Selection of Quantitative Traits in *Eucalyptus globulus*

Paulina Ballesta [1], Carlos Maldonado [1], Paulino Pérez-Rodríguez [2] and Freddy Mora [1,*]

[1] Institute of Biological Sciences, University of Talca, 2 Norte 685, Talca 3460000, Chile
[2] Colegio de Postgraduados, Statistics and Computer Sciences, Montecillos, Edo. de México 56230, Mexico
* Correspondence: fmora@utalca.cl; Tel.: +56-71-2200280

Received: 25 July 2019; Accepted: 3 September 2019; Published: 5 September 2019

Abstract: *Eucalyptus globulus* (Labill.) is one of the most important cultivated eucalypts in temperate and subtropical regions and has been successfully subjected to intensive breeding. In this study, Bayesian genomic models that include the effects of haplotype and single nucleotide polymorphisms (SNP) were assessed to predict quantitative traits related to wood quality and tree growth in a 6-year-old breeding population. To this end, the following markers were considered: (a) ~14 K SNP markers (SNP), (b) ~3 K haplotypes (HAP), and (c) haplotypes and SNPs that were not assigned to a haplotype (HAP-SNP). Predictive ability values (PA) were dependent on the genomic prediction models and markers. On average, Bayesian ridge regression (BRR) and Bayes C had the highest PA for the majority of traits. Notably, genomic models that included the haplotype effect (either HAP or HAP-SNP) significantly increased the PA of low-heritability traits. For instance, BRR based on HAP had the highest PA (0.58) for stem straightness. Consistently, the heritability estimates from genomic models were higher than the pedigree-based estimates for these traits. The results provide additional perspectives for the implementation of genomic selection in *Eucalyptus* breeding programs, which could be especially beneficial for improving traits with low heritability.

Keywords: genomic prediction; haplotype blocks; predictive ability; Bayesian models

1. Introduction

The number of breeding programs that use the principles of genomic selection (GS) has increased considerably in recent years. Over the last decade, several investigations have illustrated how to incorporate the principles of genomic prediction in the genetic improvement programs of complex traits [1–10]. In this sense, the use of GS principles has been facilitated by the development of genotyping techniques such as genotyping by sequencing (GBS) and DNA arrays (or chips) of high density in various crops including self-pollinated plants such as soybean; wheat; barley; rice [10–13]; and outcrossing plants such as fruit trees, forest trees, and corn, among others [14–17]. Single nucleotide polymorphic markers (SNP) have been a powerful tool in breeding programs for different agricultural crops [18–23]. SNP markers have multiple applications in plants, including positional cloning, whole genome association studies, the mapping of quantitative trait loci (QTL), and the determination of genetic relationships between individuals.

The non-random association between two or more loci due to a low probability of recombination between them (linkage disequilibrium: LD) in a given population constitutes haplotypes [24,25], which correspond to sets of genomic regions within a chromosome that tend to be inherited together [26]. In this context, GS can be implemented not only using individual SNPs but also haplotypes, or a combination of both, using haplotypes in conjunction with SNPs not assigned to a haplotype. For example, Cuyabano et al. [27] presented a haplotype approach for genomic prediction using high-density data in dairy cattle as an alternative to individual marker methods, demonstrating

that the haplotypes improved prediction accuracy compared to an individual SNP. On the other hand, Calus et al. [28] determined that the inclusion of haplotypes in genomic prediction models was beneficial for low-heritability traits. Matias et al. [29] found that the use of haplotypes in the prediction of complex traits of maize increased the predictive ability by 20%. From these and other studies, it appears that the haplotype approach emerges as a methodological variant that can improve not only the predictive abilities but also the precision in the detection of genomic regions in association studies [30–33]. Moreover, given that a haplotype is defined as a set of nearby SNPs, which are in strong LD [19], this analytical approach would take the natural dependence that exists between SNPs into account, which becomes more relevant when considering high-density DNA arrays.

Several studies have emphasized the use of haplotypes (from high-density DNA arrays) to estimate the predictive ability of different GS models [27,34–38]. One of the advantages of using haplotypes in GS is the ability to detect and include mutations as genomic information. According to Curtis et al. [39], when mutations have occurred, it is possible that the frequencies of the alleles remain (almost) unchanged. However, when analyzing haplotypes, mutations at different loci tend to cause significant changes in haplotype frequencies. Therefore, a QTL that is not in complete LD with an individual marker can be in complete LD with a particular haplotype [27]. Additionally, the use of haplotypes reduces the degrees of freedom in the models of prediction or genomic association (reduction of dimensionality), which contributes to greater precision in the detection of QTL [40]. It should be noted that there are few studies that have evaluated the combined use of GS and haplotypes in plants [29,41,42]. In particular, these methods have been implemented in predominantly self-pollinated (autogamous) species—for example, soybean and wheat [41,42]—in which extensive LD values can be found in their genomes, which favors the identification of haplotypes, while in outcrossing species (allogamous) such as *Eucalyptus* and most species of forest interest, LD usually decays at short genomic distances [43,44], which allows the identification of smaller haplotype blocks and those conformed by a smaller number of alleles.

The objectives of this study were to (i) determine and characterize haplotype blocks in *Eucalyptus globulus* from a high-quality SNP array (EUChip60K), (ii) investigate the possible benefits of using haplotype information to predict complex traits in trees, and (iii) estimate genomic parameters (e.g., genomic heritability and genetic gains) using Bayesian whole-regression models (Bayesian Ridge Regression, Bayes A, Bayes B, Bayes C and Bayesian Least Absolute Shrinkage and Selection Operator) that include the haplotype/SNP effect on a 6-year-old breeding population of *E. globulus*.

2. Results

2.1. Haplotype Block Construction

The final dataset used for the haplotype block construction and genomic predictions consisted of 14,422 SNPs. An average of 1356 SNPs per chromosome and an average frequency of one SNP every 4000 bp were found. Chromosome 8 contained 12.5% (1811 SNPs) of the total SNPs, while chromosome 4 presented only 6.2% (893 SNPs). Total values of 1137 haplotype blocks and 3279 haplotypes (Table 1) were identified in all chromosomes of the species under study. In total, 14.5% of the total SNPs (2092 SNPs) were grouped into haplotype blocks. The largest number of haplotype blocks was determined by combinations of SNPs located on chromosome 8 (n = 152), while a smaller amount was constructed by SNPs located on chromosome 4 (n = 71). Of the total blocks formed by the chromosomes, 300 haplotypes were obtained (on average) per chromosome, while each block had three haplotype variants on average. The number of SNPs within each haplotype block varied from two to 12. The smallest blocks had an extension of 36 bp, while the largest haplotypes had a length of 482 kbp (chromosome 8). About 24% of the total haplotype blocks had a size greater than 10 kbp, while 2% had a size above 100 kbp. Particularly, on chromosomes 4 and 9, no haplotype blocks with an extension higher than 100 kbp were detected. The genome-wide average linkage disequilibrium decay estimated across all chromosomes

is shown in Figure S1 (Supplementary Material). On a genome-wide average, LD decayed within ~10–12 Kbp to a level below r^2 = 0.14 (critical r^2 value).

Table 1. Summary of information on haplotypes and haplotype blocks determined in a breeding population of *E. globulus*. Ch corresponds to the chromosome number and single nucleotide polymorphic markers (SNPs) to the number of SNPs detected; HAP-Blocks is the number of haplotype blocks constructed; HAPs is the number of haplotypes; Max (kbp) corresponds to the maximum size (in kbp) for the haplotype blocks; Min (bp) corresponds to the minimum size (in bp) for the haplotype blocks; and Min (SNPs) and Max (SNPs) correspond to the maximum and minimum number of SNPs forming the haplotype blocks, respectively.

Ch	SNPs	HAP-Blocks	HAPs	Max (kb)	Min (bp)	Max (SNPs)	Min (SNPs)
1	924	75	219	381	61	6	2
2	1766	121	370	357	36	6	2
3	1587	99	299	123	30	11	2
4	893	71	207	31	31	5	2
5	1500	83	238	279	49	8	2
6	1474	144	407	343	63	6	2
7	1220	87	248	356	121	5	2
8	1811	152	418	482	70	6	2
9	946	89	249	94	34	5	2
10	1065	103	295	318	49	10	2
11	1236	113	329	250	75	12	2
Total	14,422	1137	3279	-	-	-	-
Mean	1311	103	298	274	54	7	2

2.2. Estimates of Genetic Parameters Based on Pedigree

Estimates of variance components, heritability and genetic gains based on pedigree information are shown in Table 2. The heritability estimates varied from 0.04 to 0.46. Wood density (WD) was the most heritable trait (h_a^2 = 0.46), while diameter at breast height (DBH), stem straightness (ST), and branch quality (BQ) had the lowest heritability estimates in this breeding population (h_a^2 = 0.04, 0.06, and 0.05, respectively). Additionally, the estimated genetic gains based on pedigree information were 7.7%, 2.9%, 4.4%, 3.9% and 9.7% for tree height (HT), DBH, ST, BQ and WD, respectively, considering an intensity of selection of ~10% (10.06%; n = 65).

2.3. Prediction Based on Genomic Data

The genomic prediction methods were compared using the average values of marginal posterior distributions of each estimated parameter. Genomic heritability, genetic gains and predictive ability (PA) values, obtained for each trait and prediction model, are shown in Tables 2 and 3 and Figure 1. PA values were dependent on the genomic prediction models and marker type (SNP markers, haplotypes (HAP) and haplotypes in conjunction with SNPs that were not assigned to a haplotype (HAP-SNP)). PA values for HT varied between 0.19 and 0.44. For Bayes B (BB) and Bayesian Ridge Regression (BRR) methods, the PA values based on SNP, HAP and HAP-SNP were not statistically different from each other, but Bayes A (BA), Bayesian Least Absolute Shrinkage and Selection Operator (Bayesian LASSO or BL) and Bayes C (BC) methods based on the three markers (SNP, HAP and HAP-SNP) showed significant differences in terms of PA values. Based on the comparison among the assessed models, BC had the highest predictive ability (PA = 0.44 (SNP) and 0.38 (HAP-SNP)) for HT in most cases, and this method consistently gave one of the highest values for genomic heritability ($\hat{h}_g^2$ = 0.36) and genetic gain (GG = 8%). In addition, the genomic heritability based on this method (either SNP or HAP-SNP) was statistically higher than the pedigree-based heritability ($\hat{h}_a^2$ = 0.15 [0.01–0.28]). On the other hand, the genetic gains for HT based on pedigree information and BC methods were not statistically different.

For DBH, PA values ranged between 0.17 and 0.46. The PA varied significantly among models based on SNP, HAP and HAP-SNP. In general, the models based on SNP had the highest PA

values. Among the models based on SNP, BRR had the highest predictive ability for DBH (PA = 0.45), while HAP-SNP-based BC had the highest predictive ability (PA = 0.46). The genomic heritability estimates based on BC and BRR ranged between 0.26–0.32 and 0.12–0.16, respectively, and were significantly higher than the pedigree-based heritability ($\hat{h}_a^2$ = 0.04 [<0.01–0.10]). Consistently, the genetic gains based on the selection with these genomic models were statistically higher than those based on pedigree information (GG = 2.9% [1.4–4.5]).

The predictive ability of ST varied from 0.20 to 0.58. On the other hand, the PA values of ST were not statistically different among markers (SNP, HAP or HAP-SNP) in most cases. BB and BC had the highest PA among the models based on SNP (PA = 0.52 and 0.54, respectively) and HAP-SNP (PA = 0.52 for BB and BC), while BRR had the highest predictive ability (PA = 0.58) among the HAP-based models. Consistently, all genomic heritability estimates based on BB, BC and BRR ($\hat{h}_g^2$ = 0.15–0.34) were higher than the pedigree-based heritability ($\hat{h}_a^2$ = 0.06 [<0.01–0.14]). In addition, only BC and BRR had genetic gain estimates statistically higher than those based on pedigree information for ST.

The predictive ability of BQ ranged between 0.06 and 0.33, which did not vary significantly between models based on SNP, HAP or HAP-SNP in most cases. Among the models based on SNP, BC had the highest predictive ability of BQ (PA = 0.31). For the HAP-based models, the highest PA was obtained by BC (PA = 0.28). In accordance with this, BC and BRR had the highest PA of BQ among the HAP-SNP-based models (PA = 0.31 and 0.33, respectively). All genomic heritability estimates based on BC and BRR ($\hat{h}_g^2$ = 0.12–0.29) were higher than the pedigree-based heritability estimate ($\hat{h}_a^2$ = 0.05 [<0.01–0.11]). However, only BRR had genetic gain estimates statistically higher than those based on pedigree information.

The predictive ability of WD ranged between 0.24 and 0.46, which varied significantly between models based on SNP, HAP and HAP-SNP in most cases. Among SNP-based models, BC had the highest predictive ability of WD (PA = 0.46). For HAP-based models, the highest predictive ability was obtained by BC and BRR (PA = 0.41 and 0.39, respectively). In addition, for HAP-SNP-based models, BRR had the highest predictive ability (PA = 0.44). The heritability of WD based on any genomic model ($\hat{h}_g^2$ = 0.04–0.26) was statistically lower than the pedigree-based heritability ($\hat{h}_a^2$ = 0.46 [0.22–0.69]). Consistently, the genetic gains for WD based on any genomic prediction model did not exceed 5% and were statistically lower than those based on pedigree information.

The regressions between pedigree-based Estimated Breeding Values (EBVs) and genomic-based EBVs (GEBVs considering models with the highest predictive ability for each trait) are shown in Figure 2. All coefficients of determination (R^2) between EBVs and GEBVs for all traits were above 0.98.

Table 2. Estimates of pedigree-based heritability ($\hat{h}_a^2$), genomic heritability ($\hat{h}_g^2$) and genetic gain (GG; percentage) for each method of prediction based on pedigree (PBP), SNP markers (SNP), haplotype (HAP), and haplotypes and SNPs that were not assigned to a haplotype (HAP-SNP). BA, BB, BC, BL, and BRR correspond to Bayes A, Bayes B, Bayes C, Bayesian Least Absolute Shrinkage, and Selection Operator and Bayesian Ridge Regression, respectively.

Trait/Model	Pedigree		SNP		HAP		HAP-SNP	
	$\hat{h}_a^2$ [CR]	GG [CR]	$\hat{h}_g^2$	GG	$\hat{h}_g^2$	GG	$\hat{h}_g^2$	GG
Tree height								
PBP	0.15 [0.01–0.28]	7.7 [5.4–10.3]	-	-	-	-	-	-
BA	-	-	0.11	5.6	0.06	4.2 *	0.10	5.2 *
BB	-	-	0.27	6.0	0.11	3.6 *	0.29 *	6.3
BC	-	-	0.36 *	7.9	0.28	6.6	0.36 *	7.8
BL	-	-	0.07	4.2 *	0.04	3.1 *	0.06	3.6 *
BRR	-	-	0.19	8.7	0.14	7.6	0.20	8.6
Diameter at breast height								
PBP	0.04 [<0.01–0.10]	2.9 [1.4–4.5]	-	-	-	-	-	-
BA	-	-	0.08	5.0 *	0.04	3.8	0.07	4.2
BB	-	-	0.19 *	4.9 *	0.09	3.4	0.14 *	3.3
BC	-	-	0.31 *	7.2 *	0.26 *	6.6 *	0.32 *	7.2 *
BL	-	-	0.05	3.6	0.05	4.1	0.05	3.4
BRR	-	-	0.16 *	8.3 *	0.12 *	7.8 *	0.16 *	8.2 *
Stem straightness								
PBP	0.06 [<0.01–0.14]	4.4 [1.9–7.1]	-	-	-	-	-	-
BA	-	-	0.10	4.4	0.09	4.7	0.09	4.1
BB	-	-	0.26 *	5.7	0.18 *	4.7	0.28 *	5.5
BC	-	-	0.34 *	7.1	0.30 *	7.5 *	0.34 *	7.2 *
BL	-	-	0.05	2.7	0.04	2.8	0.07	3.2
BRR	-	-	0.18 *	7.6 *	0.15 *	7.9 *	0.20 *	8.0 *
Branch quality								
PBP	0.05 [<0.01–0.11]	3.9 [1.4–6.0]	-	-	-	-	-	-
BA	-	-	0.04	2.2	0.04	2.7	0.05	2.3
BB	-	-	0.12 *	2.4	0.10	2.7	0.08	1.8
BC	-	-	0.29 *	5.0	0.25 *	5.4	0.29 *	4.9
BL	-	-	0.03	2.0	0.03	2.3	0.04	2.0
BRR	-	-	0.15 *	6.2 *	0.12 *	6.4 *	0.15 *	6.1 *
Wood density								
PBP	0.46 [0.22–0.69]	9.7 [7.5–12]	-	-	-	-	-	-
BA	-	-	0.07 *	2.0 *	0.05 *	2.0 *	0.08 *	2.1 *
BB	-	-	0.17 *	2.2 *	0.12 *	1.9 *	0.16 *	2.1 *
BC	-	-	0.34	3.2 *	0.26	3.0 *	0.33	3.1 *
BL	-	-	0.06 *	1.7 *	0.04 *	1.7 *	0.06 *	1.8 *
BRR	-	-	0.16 *	3.6 *	0.12 *	3.4 *	0.17 *	3.5 *

Numbers with asterisks are statically different from pedigree-based estimates (90% Bayesian credible sets). CR: 90% credible region from marginal posterior distributions.

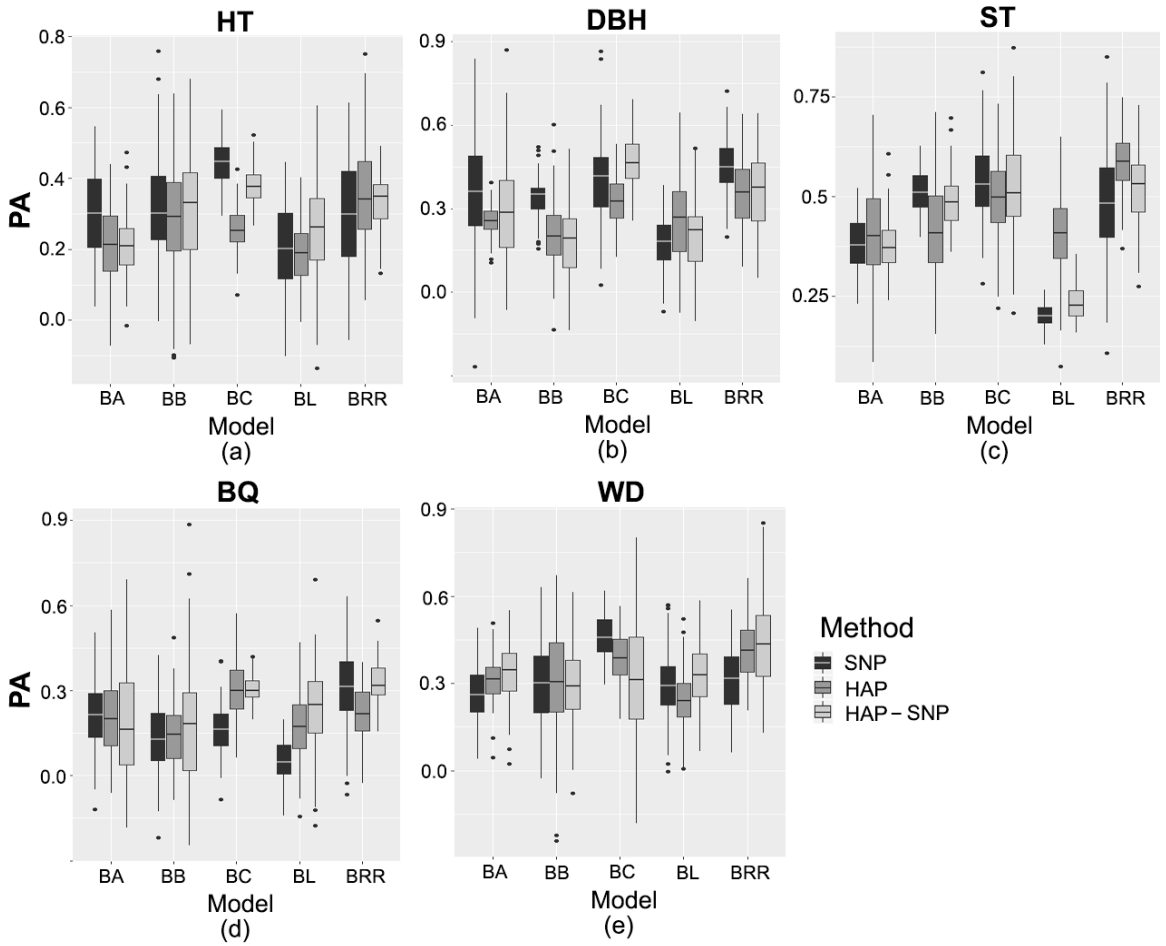

Figure 1. Predictive ability (PA) of (**a**) tree height (HT), (**b**) diameter at breast height (DBH), (**c**) stem straightness (ST), (**d**) branch quality (BQ), and (**e**) wood density (WD). Models based on SNP markers (SNP), haplotypes (HAP), and haplotypes with SNPs that were not assigned to a haplotype (HAP-SNP) are represented by black, dark gray and light gray bars, respectively. BA, BB, BC, BL, and BRR correspond to Bayes A, Bayes B, Bayes C, Bayesian Least Absolute Shrinkage and Selection Operator, and Bayesian Ridge Regression, respectively. Each box-plot represents the distribution of PA values for 100 cycles of cross-validation.

Table 3. Estimates of predictive ability (average of 100 cross-validation cycles) of Bayesian models based on SNPs (SNP), haplotypes (HAP) and haplotypes with SNPs that were not assigned to a haplotype (HAP-SNP) for each studied trait.

Trait/Markers	Genomic Model				
	BA	**BB**	**BC**	**BL**	**BRR**
Tree height					
SNP	0.31 bA	0.32 bA	**0.44** aA	0.21 cB	0.30 bA
HAP	0.21 cdB	0.28 bA	0.25 bcC	0.19 dB	0.35 aA
HAP-SNP	0.21 dB	0.31 bA	0.38 aB	0.26 cA	0.33 bA
Diameter at breast height					
SNP	0.35 bA	0.34 bA	0.39 bB	0.17 cB	0.45 aA
HAP	0.26 bcB	0.21 cB	0.33 aC	0.26 bA	0.36 aB
HAP-SNP	0.28 cAB	0.19 dB	**0.46** aA	0.20 dB	0.37 bB
Stem straightness					
SNP	0.38 cA	0.52 abA	0.54 aA	0.20 dC	0.48 bB
HAP	0.40 cA	0.42 cB	0.50 bA	0.40 cA	**0.58** aA
HAP-SNP	0.38 bA	0.49 aA	0.52 aA	0.23 cB	0.52 aB
Branch quality					
SNP	0.22 bA	0.13 cA	0.16 cB	0.06 dC	0.31 aA
HAP	0.20 bA	0.14 cA	0.28 aA	0.17 bcB	0.22 bB
HAP-SNP	0.19 bA	0.18 bA	0.31 aA	0.24 bA	0.33 aA
Wood density					
SNP	0.26 bB	0.30 bA	**0.46** aA	0.29 bA	0.32 bB
HAP	0.32 bA	0.31 bA	0.39 aB	0.24 cB	0.41 aA
HAP-SNP	0.34 bA	0.29 bA	0.32 bC	0.33 bA	0.44 aA

BA: Bayes A; BB: Bayes B; BC: Bayes C; BL: Bayesian Least Absolute Shrinkage and Selection Operator; BRR: Bayesian Ridge Regression. Statistical significance between different genomic models (BA, BB, BC, BL and BRR) is noted by lowercase letters, while that between different markers (SNP, HAP and HAP-SNP) is shown by upper case letters. Different letters show the statistical significance at $p < 0.01$ using the Tukey–Kramer test. Numbers in bold show the highest PA estimates considering both approaches: genomic models and marker types (SNP, HAP or HAP-SNP).

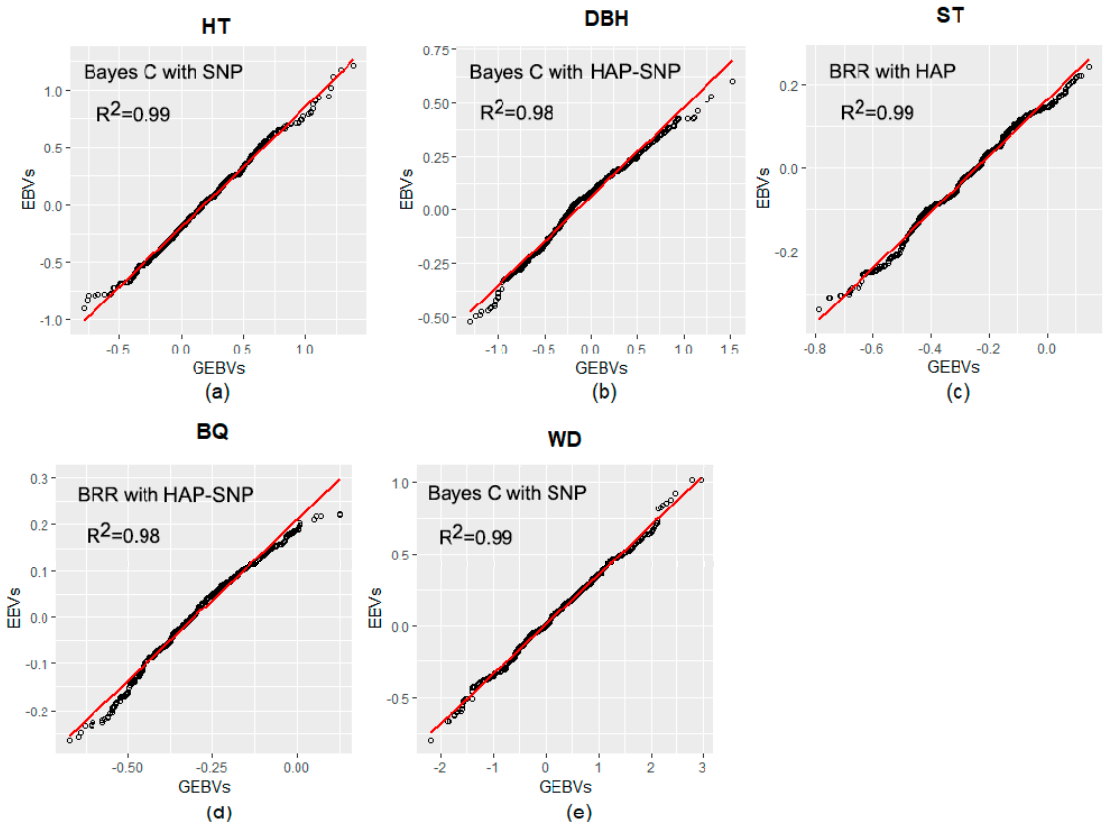

Figure 2. Linear regression plots relating estimated breeding values (pedigree-based Estimated Breeding Values; EBVs) and genomic-based EBVs (GEBVs). (**a**) EBVs and GEBVs for tree height (HT); (**b**) EBVs and GEBVs for diameter at breast height (DBH); (**c**) EBVs and GEBVs for stem straightness (ST); (**d**) EBVs and GEBVs for branch quality (BQ); and (**e**) EBVs and GEBVs for wood density (WD).

3. Discussion

3.1. Haplotype Blocks Construction

Previously, Durán et al. [45], Thavamanikumar et al. [46] and Cappa et al. [47] determined that the linkage disequilibrium (LD) in natural and controlled populations of *E. globulus* decreases rapidly in the range of 1000–4000 bp. However, in this study, several genomic regions were detected in strong disequilibrium above 10,000 bp. About 24% and 2% of the haplotype blocks formed had an extension over 10,000 bp and 100,000 bp, respectively, with a disequilibrium coefficient (D′) value higher than 0.7. The construction of the haplotype blocks was based on the confidence interval algorithm of Gabriel et al. [48], which establishes that those pairs of SNPs that are in a strong linkage disequilibrium have a D′ value between 0.7 and 0.98, considering a confidence interval of 95%. These D′ values could reveal that, historically, the probability of recombination between both loci is quite low. In this study, more than 1300 haplotype blocks that would meet this condition were detected. Therefore, it is possible to assume that the SNPs grouped in blocks are in strong LD and they are possibly inherited together across generations.

In breeding populations of self-pollinated plants and controlled crosses, it is expected that LD decays over long distances. For example, in soybean and rice, LD can be significant at a distance of 100,000 bp and 250,000 bp between loci, respectively [49], while in the genomes of outcrossing plants, LD is expected to decay at short distances due to the reproductive mechanisms that underlie these species, and their loci tend to be heterozygous [50,51]. In the context of this study, the genotyped individuals were sampled from a breeding population formed by full- and half-sib families, which could increase the probability of the occurrence of regions that form haplotype blocks.

3.2. Performance of Pedigree and Genomic Prediction Models

According to the pedigree and genomic prediction models, the traits studied in this population were found to be weakly to highly heritable. The results indicated that DBH, ST and BQ are under relatively low genetic control. However, previous studies have reported that the heritability of DBH in *Eucalyptus* spp. can be greater than 0.1 in populations from 3 to 15 years of age [52–55]. For DBH, the genomic prediction models based on BC and BRR had higher heritability values than those based on pedigree, with a genetic gain up to ~8%. Previously, Tan et al. [56] determined that the genomic heritability of DBH, based on Ridge Regression Best Linear Unbiased Prediction (BLUP) and Reproducing Kernel Hilbert Spaces, was higher than the pedigree-based heritability estimate in *Eucalyptus* hybrids at 6 years of age. In this study, SNP-based BRR and HAP-SNP-based BC had the highest predictive ability for DBH, but the genomic heritability estimated by BC (HAP-SNP) was two times higher than the BRR model based on SNP markers. In BC, the prior assumptions of all SNP effects have a common variance, and the method assigns a nonnull prior probability for the marker effect to be equal to zero [5,57]. Due to this assumption, Bayes C has been used to identify QTLs with large effects [58]. Similarly, BRR assumes a common variance for all SNP markers, but all SNP effects are shrunk to a similar extent [7]. In contrast, Suontama et al. [59] and Resende et al. [6] reported that the genomic heritability of DBH was not superior to the heritability estimated by BLUP in *Eucalyptus nitens* and hybrids of *Eucalyptus urophylla* × *E. grandis*, respectively. In another study, Müller et al. [60] reported that the genomic heritability of DBH was lower than the estimated heritability by BLUP in *E. benthamii*. Interestingly, Müller et al. [60] performed a prediction for DBH using 13,787 and 10,460 SNPs that were not in LD, which could confirm our hypothesis that the inclusion of SNPs in strong LD (such as a BC model based on HAP-SNP) could be beneficial for genetic parameter estimation.

The heritability of ST has been previously reported as low [54,61], which is consistent with the pedigree-based model and some genomic prediction models assessed in the present study. However, other studies have found that the heritability of ST in *E. globulus* can be moderate [62] and even high ($\hat{h}_a^2 > 0.3$, [53]). Almost all genomic-based heritability estimates of ST were statistically higher than those estimated by the pedigree-based method. Suontama et al. [59] found that the pedigree-based heritability of stem straightness in *E. nitens* increased approximately twice when using marker-based models. In the context of our study, the highest predictive ability for ST was obtained by BC based on SNP, BRR based on HAP, and BRR and BC models based on HAP-SNP. BC and BRR models had higher heritability than the estimate based on pedigree. However, BC models (based on either SNP or HAP) had a higher genomic heritability than other prediction models. There have been conflicting reports on the genetic architecture of stem straightness in trees. For example, Bartholomé et al. [63] detected QTLs that explained up to 5% of the total phenotypic variation of ST in maritime pine. Yang et al. [64] found QTLs that explained up to 15% of the total phenotypic variation of the stem straightness of Pinus hybrids. Additionally, Arriagada et al. [65] reported five QTLs for ST, explaining a total of 6–14% of the total proportion of trait variation in *Eucalyptus cladocalyx*.

BQ has been a scarcely evaluated trait in breeding programs of *E. globulus*. For instance, Callister et al. [53] reported a range in branching quality (measured as branch thickness) of 0–0.16 in full-sib families of E. globulus (at age 3.5). Ballesta et al. [43] reported that the pedigree-based heritability estimate of BQ in *E. globulus* at 4 years of age was less than 0.1. In other tree species, traits related to branch quality have been described to be moderately heritable [66,67]. In the present study, among the models based on SNP and HAP markers, BC had the highest predictive ability of branch quality. On the other hand, BC and BRR had the highest PA values among the models based on HAP-SNP. Interestingly, all genomic heritability estimates based on BC and BRR were higher than the pedigree-based heritability estimates. However, this result should be interpreted with caution due to Bayes C performing variable selection and shrinkage procedures, which could mean that the polygenic background may not be taken into account, favoring a selection based on major effect genes [68]. This could explain why the genetic gains based on BC were lower than those based on BRR.

Several studies have shown that WD is a highly heritable trait in *Eucalyptus* spp. [59,62,69], which is consistent with the estimates from the pedigree-based model and BC based on SNP, HAP and HAP-SNP. None of the methods based on genomic data exhibited higher heritability (and genetic gain) for this trait than the pedigree-based method. In other studies, Suontama et al. [59] reported that genomic-based and pedigree-based heritability values were similar for the basic wood density in *E. nitens*. Consistent with our findings, Resende et al. [6] reported that the genomic heritability of WD is subtly lower than the pedigree-based heritability in hybrids of *E. urophylla* and *E. grandis*. In other tree species, Beaulieu et al. [70] reported that the heritability of the wood density in trees of *Picea glauca* is lower when estimated by genomics than pedigree. As expected, the highest value of genetic gain was obtained for WD, while the lowest value was obtained for DBH. In this sense, the selection based on genomics for traits such as DBH, ST and BQ is highly justified and particularly attractive because of the potential of enhancing selection accuracy for low heritability traits and increasing the genetic gains for these traits [6,71]. On the other hand, Beaulieu et al. [70] stated that one of the reasons why genomic selection might not be as effective for predicting wood density is that pedigree information makes it possible to capture loci that have not been considered in a genomic prediction model, which can be important for the genetic control of the trait. According to our results, the BC model had a higher predictive ability and genomic heritability in most cases, supporting the idea of an oligogenic architecture for WD. Contrarily, Durán et al. [45] reported that BC model had a lower PA than the BL model.

The genomic heritability for HT, estimated by a BC model (SNP and HAP-SNP), was higher than the pedigree-based estimate. However, the HAP-SNP-based BC model had a predictive ability lower than the model based on SNP markers. In accordance with this, Tan et al. [56] determined that the genomic heritability of HT, based on Ridge Regression BLUP and Reproducing Kernel Hilbert Spaces, was superior to the pedigree-based heritability in *Eucalyptus* hybrids. However, the genetic gains estimated using the pedigree and BC methods were not statistically different. Previously, Lenz et al. [72] and Beaulieu et al. [70] reported that the genetic gains based on genomic prediction could be lower than the gains based on the pedigree method for HT in other tree species. As mentioned, the BC model (based on SNP and HAP-SNP) had the highest predictive ability for HT in the most cases. Contrarily, Müller et al. [60] found no differences in the predictive ability of growth traits (HT and DBH) between the BA, BB, BC, BL, and BRR models.

Several studies in animals have emphasized the use of haplotypes (from high-density DNA arrays) to estimate the predictive ability of different GS models [27,35–38]. In plants, Matias et al. [29] demonstrated that the use of haplotypes in prediction studies in maize increased its predictive ability by 20%. It should be noted that in outcrossing plants, such as *Eucalyptus* and other forest species, LD usually decays at short genomic distances, which allows the identification of smaller haplotype blocks and with fewer variants (haplotypes). In this study, we found haplotype blocks formed by SNPs in a strong LD ($D' > 0.7$) with a size above 300 kbp. According to Cuyabano et al. [27], haplotypes that are constructed by SNPs in a LD of $D' > 0.45$ can significantly increase the predictive ability of a genomic selection model. Our results revealed that predictive ability values were mainly dependent on the Bayesian methods assessed (i.e., BA, BB, BC, BL and BRR) more than on the marker type (SNP, HAP or HAP-SNP). However, genomic models that included the haplotype effect (either HAP or HAP-SNP) significantly increased the PA of low-heritability traits. These results, nevertheless, should be interpreted with caution due to the age of the trees, and we therefore emphasize that further studies are needed to evaluate the performance of the genomic models. On the other hand, the development of techniques to select trees at early growth stages may greatly increase the genetic gain per unit time, and thus, substantially accelerate tree-breeding programs.

4. Materials and Methods

4.1. Trial Conditions and Phenotyping

In 2018, 6-year-old trees of *Eucalyptus globulus* from a progeny trial consisting of a mix of half-sib and full-sib families located in the La Poza sector, Purranque, Chile [43], were evaluated according to the following traits (description in Table S1 and Figure S2): wood density (WD), stem straightness (ST), branch quality (BQ), diameter at breast height (DBH), and tree height (HT). This location (40°57′S, 73°30′W; 326 m.a.s.l.) has an Oceanic or Marine climate type with an annual accumulated rainfall of 1282 mm and an average annual temperature of 13 °C. The WD was measured indirectly according to Valenzuela et al. [69]. ST was evaluated in the first 2/3 of the total height of the tree according to an ordinal scale (seven levels). The value 0 corresponds to trees that have a curvature in the first third of the total height of the tree and 6 in the case of trees that could present a slight curvature in the upper third of the total height of the tree without affecting productivity. BQ was evaluated according to different criteria that define quality (diameter, angle and distribution of branches in the tree) by means of an ordinal scale of six levels, in which a value of 0 is assigned to trees with an extreme deficiency in the diameter of branches and any other variable, and a value of 6 corresponds to trees that have an optimal combination of all quality variables without generating loss of productivity. The trees were distributed in a randomized complete block design with 30 blocks, considering single-tree plots (each family is represented by only one tree in each block) and a planting density of 2.5 m between each tree within each block.

4.2. Genotyping, Linkage Disequilibrium and Haplotype Blocks

Genomic DNA was isolated from the leaves of 646 randomly selected individuals of *E. globulus* (~10 individuals per family). The DNA extraction protocol followed the work of Ballesta et al. [43]. Individuals were genotyped using the EUChip60K SNP system (GeneSeek, Lincoln, NE, USA) [17]. The genotyping quality of the samples was evaluated in Genome Studio software (Illumina, San Diego, CA). Monomorphic SNP markers and those with a call rate <90% were removed. Subsequently, those SNPs with a minor allele frequency (MAF) <0.05 were eliminated. A total of 14,442 remaining SNPs was retained for the 646 individuals.

The haplotype blocks were defined according to the confidence interval algorithm developed by Gabriel et al. [48] using the software Haploview v. 4.2 [73]. The pairs of SNPs were considered to be in strong linkage disequilibrium (LD) if the upper limit of the 95% confidence interval of the value of normalized disequilibrium coefficient (D′) was higher than 0.98 and if the lower limit had a minimum value of 0.7. The D′ between A and B loci, was calculated as follows:

$$D'_{AB} = D/D_{MAX} \tag{1}$$

where D is calculated as $D = p_{A_1B_1}p_{A_2B_2} - p_{A_1B_2}p_{A_2B_1}$, and D_{MAX}:

$$D_{MAX} = \begin{cases} -\min\left\{p_{A_1}p_{B_1}, p_{A_2}p_{B_2}\right\}, \text{ when } D < 0 \\ \min\left\{p_{A_1}p_{B_2}, p_{A_2}p_{B_1}\right\}, \text{ when } D \geq 0 \end{cases} \tag{2}$$

The physical positions of each SNP were established according to the consensus map of the genome of *E. grandis* [74]. The extent of LD was also estimated as the squared allele frequency correlation (r^2). The critical r^2 value was calculated according to the method used by Breseghello and Sorells [75].

4.3. Prediction Models Based on Pedigree and Genomic Data

In this study, prediction models based on pedigree and genomic data from an array of SNP markers were used. In the pedigree-based model, individual breeding values were predicted using a Bayesian generalized linear model implemented in the MCMCglmm (Markov Chain Monte Carlo—Generalized

Linear Mixed Model) library [76] of R 3.6.1 [77] This Bayesian analysis was carried out using the following base model:

$$y = X\beta + Za + \varepsilon \tag{3}$$

where y corresponds to the phenotypic data vector, β is the vector of block effects, a is the vector of the additive genetic effects $a \sim N\left(0, A\sigma_a^2\right)$, A corresponds to the matrix of Wright's coefficients (pedigree information), and σ_a^2 is the additive genetic variance. X and Z correspond to the known incidence matrices that relate the observation vector (y) to vectors β and a, respectively, and ε corresponds to the vector of residual effects, $\varepsilon \sim N\left(0, I\sigma_\varepsilon^2\right)$, where I is an identity matrix, and σ_ε^2 is the residual variance. The Bayesian models were run with 1,000,000 iterations, a burn-in period of 100,000 and a thin of 50.

The prediction models based on SNPs/haplotypes were the following: Bayesian Least Absolute Shrinkage and Selection Operator (Bayesian LASSO or BL, [78,79]), Bayesian Ridge Regression (BRR, [7]), Bayes A (BA, [1]), Bayes B (BB, [1]) and Bayes Cπ (BC, [5]). All whole-regression models can be expressed in matrix form as follows:

$$y = X\beta + Zm + \varepsilon \tag{4}$$

where y corresponds to the phenotypic data vector, β is the vector of block effects, m corresponds to marker effects (SNPs and/or haplotypes), and depending on the model, different prior distributions are assigned; for instance, a double exponential distribution in the case of BL and a Gaussian distribution in the case of BRR, among others (see Refs. [7,57]). ε corresponds to the vector of residuals, $\varepsilon \sim N\left(0, I\sigma_\varepsilon^2\right)$. X and Z correspond to the incidence matrices that relate the observation vector (y) to vectors β and m, respectively.

The matrix of SNP markers or haplotypes was coded by the numbers 0, 1 and 2. In the case of SNP markers, 0 represents the homozygous genotype of the allele with the lowest frequency for the *i*-th marker ($i = 1, \ldots, n$), 1 represents the heterozygous genotype for the *i*-th marker, and 2 represents the homozygous genotype of the allele with the highest frequency for the *i*-th marker. In the case of haplotypes, since one haplotype block can have more than two allelic variants, the values of 0, 1 and 2 represent the number of copies for each variant (haplotype), in which a value of 0 was assigned for those individuals who did not present any copy of the *j*-th haplotype and *i*-th block, a value of 1 was assigned if they presented a copy of the *j*-th haplotype and *i*-th block, and a value of 2 was assigned if they presented two copies of the *j*-th haplotype and *i*-th block [27]. The models based on SNP markers, haplotypes and haplotypes in conjunction with SNPs (that were not assigned to a haplotype) are identified by SNP, HAP and HAP-SNP, respectively.

The BL method assumes that the marker effects are distributed a priori according to a double exponential (DE), $p\left(m_i|\lambda, \sigma_\varepsilon^2\right) = DE\left(m_i|0, \lambda, \sigma_\varepsilon^2\right)$, where λ corresponds to a regularization parameter. The distribution of DE generates a strong contraction (close to zero) to estimate the effects of the markers. BRR is a Bayesian method based on the fact that model regressors (SNPs and/or haplotypes) have a common variance (σ_m^2); those regressors with the same allelic frequency explain the same proportion of the additive variance and have the same contraction effect [7]. The marker effect (m_i) is distributed as follows: $m_i\left|\sigma_m^2\right. \sim N\left(0, \sigma_m^2\right)$; and the common variance $p(\sigma_m^2)$~scaled inverse Chi-squared ($\sigma_m^2\left|\upsilon_m, S_m\right. \sim \chi^{-2}(\upsilon_m, S_m)$), with degree of freedom and scale parameters υ_m and S_m, respectively. In BA, the marginal distribution of marker effects is a scaled-t density, in which, for computational convenience, this density is implemented as an infinite mixture of scaled-normal densities (see Ref. [57]). The variance of each marker is assumed to be distributed scaled inverse Chi-squared. BB uses a mixed distribution with a mass at zero, such that the prior distribution of the effects of the all markers is given by [51]

$$m_i|\sigma_{m_i}^2, \pi = \begin{cases} 0 & \text{with probability } \pi \\ N\left(0, \sigma_{m_i}^2\right) & \text{with probability } 1-\pi \end{cases} \tag{5}$$

A scaled inverse Chi-square prior distribution $\chi^{-2}(\upsilon_m, S_m)$ is assumed for $\sigma^2_{m_i}$ ($i = 1, \ldots, n$), which is equal for all markers. In BC, all markers are considered to have a common variance (σ^2_m) and promote the selection of variables such as Bayes B. The marker effects are assumed to be $m_i \middle| \sigma^2_{m_i} \sim N\left(0, \sigma^2_m\right)$ with a probability of $1 - \pi = 0$.

All the genomic-based Bayesian methods were implemented in the library BGLR (Bayesian Generalized Linear Regression) [57] of R 3.6.1 [77]. Variance components were estimated with a total of 1,000,000 iterations, a burn-in period of 100,000, and a thin of 50. The predictive ability (PA) of each model was measured as the correlation between the Genomic Estimated Breeding Values (GEBVs) obtained by Equation (2) and genomic breeding values predicted by cross-validation, which considered 90% of the individuals as the training population and the remaining 10% as the validation population. The PA was reported as the average of correlation coefficients for 100 cycles of the cross-validation.

4.4. Heritability and Genetic Gain

The prediction methods (genomic and pedigree-based) were also compared by the values of heritability and genetic gains. In the case of the pedigree-based method [80], the heritability ($\hat{h}^2_a$) in a narrow sense was calculated as follows:

$$\hat{h}^2_a = \frac{\hat{\sigma}^2_a}{\hat{\sigma}^2_a + \hat{\sigma}^2_\varepsilon} \tag{6}$$

where $\hat{\sigma}^2_a$ and $\hat{\sigma}^2_\varepsilon$ correspond to the additive genetic and residual variances, respectively. In the case of Bayesian genomic prediction models (SNP/haplotypes), genomic heritability ($\hat{h}^2_g$), genomic variance ($\hat{\sigma}^2_g$) and the residual variance ($\hat{\sigma}^2_\varepsilon$) were estimated using the marginal posterior distributions of each estimated parameter [81–83]. The genomic variance was estimated for each model as follows:

For BRR and BC:

$$\hat{\sigma}^2_g = 2\,\hat{\sigma}^2_m \sum_{i=1}^{n} p_i(1 - p_i) \tag{7}$$

For BA and BB:

$$\hat{\sigma}^2_g = 2\sum_{i=1}^{n} p_i(1 - p_i)\hat{\sigma}^2_{m_i} \tag{8}$$

For BL:

$$\hat{\sigma}^2_g = 2\sum_{i=1}^{n} \tau^2_i \hat{\sigma}^2_\varepsilon p_i(1 - p_i) \tag{9}$$

where p_i is the MAF of ith marker, $\hat{\sigma}^2_m$ is the variance of markers, and τ^2_i was assumed to be the exponential of λ, $\tau^2_i \sim \mathrm{Exp}\left(\lambda^2\right)$, in which λ was assumed to belong to Gamma distribution $\lambda^2 \sim \mathrm{G}(\varphi_1, \varphi_2)$. BRR and BC models assume that all markers have the same variance ($\hat{\sigma}^2_m$), while BB and BA models assume a variance for each ith marker ($\hat{\sigma}^2_{m_i}$). BB, BC and BL models assume the selection of variables; however, BL uses regularization parameter λ that directs markers with irrelevant effects close to zero.

The genetic gain (GG) was estimated for each prediction method using the following expression [54,84]:

$$\mathrm{GG} = \frac{\left(\overline{y}_{\mathrm{sel}} - \overline{y}_{\mathrm{pop}}\right)}{\overline{y}_{\mathrm{phe}}} * 100 \tag{10}$$

where $\overline{y}_{\mathrm{sel}}$ and $\overline{y}_{\mathrm{pop}}$ correspond to the estimated posterior mean of the breeding values of selected trees (with a selection intensity of ~ 10.06%) and the estimated posterior population mean of the breeding values, respectively, and $\overline{y}_{\mathrm{phe}}$ is the phenotypic mean for each trait. For ordinal traits, the $\overline{y}_{\mathrm{phe}}$ term was calculated according to Burdon et al. [85].

5. Conclusions

To our knowledge, this study is one of the first to examine the inclusion of haplotypes in genomic selection models of *Eucalyptus*. In general, genomic heritability estimates were higher than those based on pedigree information for most of the studied traits. On the other hand, the predictive ability values were dependent on the genomic prediction models and marker type. On average, the homocedastic methods (BRR and Bayes C) had the highest predictive ability for the majority of traits. Notably, genomic models that included the haplotype effect (either HAP or HAP-SNP) significantly increased the PA of traits with low heritability. The results of this study provide additional perspectives for the implementation of genomic selection in *Eucalyptus* breeding programs, which could be especially beneficial for improving low heritability traits.

Supplementary Materials: The following are available online at http://www.mdpi.com/2223-7747/8/9/331/s1, Figure S1: Genome-wide average linkage disequilibrium (LD) decay plot estimated across all chromosomes of *Eucalyptus* for the studied population. The LD values correspond to the average of the correlation between alleles at two loci (r^2) and the normalized disequilibrium coefficient (D′) for each 1 Kpb. The LD threshold of $r^2 = 0.14$ is indicated with a red line; Figure S2: Distributions and histograms of the studied traits: (a) tree height (HT) in m, (b) diameter at breast height (DBH) in cm, (c) stem straightness (ST), (d) branch quality (BQ) and (e) wood density (WD) in mm; Table S1: Summary of phenotypic information for quantitative traits related to wood quality and tree growth measured in a six-year-old breeding population of *E. globulus*.

Author Contributions: Conceptualization, P.B. and F.M.; methodology, P.B. and F.M.; software, C.M.; validation, P.B., F.M. and C.M.; formal analysis, C.M., F.M. and P.B.; investigation, P.B.; resources, F.M.; data curation, C.M., F.M.; writing original draft preparation, P.B.; writing review and editing, F.M., P.P.-R.; supervision, F.M.; project administration, F.M.; funding acquisition, F.M.

Funding: This research was funded by FONDECYT (grant number 1170695) and Semillas Imperial SpA.

Acknowledgments: The authors thank FONDECYT (grant number 1170695) and Semillas Imperial SpA. Paulina Ballesta thanks CONICYT-PCHA/Doctorado Nacional/2016-21160624. Carlos Maldonado thanks CONICYT-PCHA/Doctorado Nacional/2017-21171466.

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

References

1. 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.
2. Gianola, D. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* **2006**, *173*, 1761–1776. [CrossRef]
3. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
4. Pérez, P.; de los Campos, G.; Crossa, J.; Gianola, D. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* **2010**, *3*, 106–116. [CrossRef]
5. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform.* **2011**, *12*, 186. [CrossRef]
6. Resende, M.D.; Resende, M.F.; Sansaloni, C.P.; Petroli, C.D.; Missiaggia, A.A.; Aguiar, A.M.; Abad, J.M.; Takahashi, E.K.; Rosado, A.M.; Faria, D.A.; et al. Genomic selection for growth and wood quality in *Eucalyptus*: Capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* **2012**, *194*, 116–128. [CrossRef]
7. Gianola, D. Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* **2013**, *194*, 573–596. [CrossRef]
8. Azevedo, C.F.; Silva, F.F.; de Resende, M.D.V.; Lopes, M.S.; Duijvesteijn, N.; Guimarães, S.E.F.; Lopes, P.S.; Kelly, M.J.; Viana, J.M.S.; Knol, E.F. Supervised independent component analysis as an alternative method for genomic selection in pigs. *J. Anim. Breed. Genet.* **2014**, *131*, 452–461. [CrossRef]
9. Azevedo, C.F.; de Resende, M.D.V.; e Silva, F.F.; Viana, J.M.S.; Valente, M.S.F.; Resende, M.F.R.; Muñoz, P. Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genet.* **2015**, *16*, 105. [CrossRef]
10. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE* **2013**, *8*, e54985. [CrossRef]

11. Chen, H.; Xie, W.; He, H.; Yu, H.; Chen, W.; Li, J.; Yu, R.; Yao, Y.; Zhang, W.; He, Y.; et al. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol. Plant* **2014**, *7*, 541–553. [CrossRef]
12. Avni, R.; Nave, M.; Eilam, T.; Sela, H.; Alekperov, C.; Peleg, Z.; Dvorak, J.; Korol, A.; Distelfeld, A. Ultra-dense genetic map of durum wheat× wild emmer wheat developed using the 90K iSelect SNP genotyping assay. *Mol. Breed.* **2014**, *34*, 1549–1562. [CrossRef]
13. Bayer, M.M.; Rapazote-Flores, P.; Ganal, M.; Hedley, P.E.; Macaulay, M.; Plieske, J.; Ramsay, L.; Russell, J.; Shaw, P.D.; Thomas, W.; et al. Development and evaluation of a barley 50k iSelect SNP array. *Front. Plant Sci.* **2017**, *8*, 1792. [CrossRef]
14. Verde, I.; Bassil, N.; Scalabrin, S.; Gilmore, B.; Lawley, C.T.; Gasic, K.; Micheletti, D.; Rosyara, U.R.; Cattonaro, F.; Vendramin, E.; et al. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS ONE* **2012**, *7*, e35668. [CrossRef]
15. Bianco, L.; Cestaro, A.; Sargent, D.J.; Banchi, E.; Derdak, S.; Di Guardo, M.; Salvi, S.; Jansen, J.; Viola, R.; Gut, I.; et al. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus*× *domestica* Borkh). *PLoS ONE* **2014**, *9*, e110377. [CrossRef]
16. Unterseer, S.; Bauer, E.; Haberer, G.; Seidel, M.; Knaak, C.; Ouzunova, M.; Meitinger, T.; Strom, T.M.; Fries, R.; Pausch, H.; et al. A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomes* **2014**, *15*, 823.
17. Silva-Junior, O.B.; Faria, D.A.; Grattapaglia, D. 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]
18. Mora, F.; Quitral, Y.A.; Matus, I.; Russell, J.; Waugh, R.; Del Pozo, A. SNP-based QTL mapping of 15 complex traits in barley under rain-fed and well-watered conditions by a mixed modeling approach. *Front. Plant Sci.* **2016**, *7*, 909. [CrossRef]
19. 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]
20. Rasheed, A.; Hao, Y.; Xia, X.; Khan, A.; Xu, Y.; Varshney, R.K.; He, Z. Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. *Mol. Plant* **2017**, *10*, 1047–1064. [CrossRef]
21. Battenfield, S.D.; Sheridan, J.L.; Silva, L.D.; Miclaus, K.J.; Dreisigacker, S.; Wolfinger, R.D.; Peña, R.J.; Singh, R.P.; Jackson, E.W.; Fritz, A.K.; et al. Breeding-assisted genomics: Applying meta-GWAS for milling and baking quality in CIMMYT wheat breeding program. *PLoS ONE* **2018**, *13*, e0204757. [CrossRef]
22. Li, C.X.; Xu, W.G.; Guo, R.; Zhang, J.Z.; Qi, X.L.; Hu, L.; Zhao, M.Z. Molecular marker assisted breeding and genome composition analysis of Zhengmai 7698, an elite winter wheat cultivar. *Sci. Rep.* **2018**, *8*, 322. [CrossRef]
23. 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]
24. Nordborg, M.; Tavaré, S. Linkage disequilibrium: What history has to tell us. *Trends Genet.* **2002**, *18*, 83–90. [CrossRef]
25. Machiela, M.J.; Chanock, S.J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **2015**, *31*, 3555–3557. [CrossRef]
26. Andersen, J.; Lübberstedt, T. Functional markers in plants. *Trends Plant Sci.* **2003**, *8*, 554–560. [CrossRef]
27. Cuyabano, B.C.; Su, G.; Lund, M.S. Genomic prediction of genetic merit using LD based haplotypes in the Nordic Holstein population. *BMC Genom.* **2014**, *15*, 1171. [CrossRef]
28. Calus, M.P.; Meuwissen, T.H.; Windig, J.J.; Knol, E.F.; Schrooten, C.; Vereijken, A.L.; Veerkamp, R.F. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genet. Sel. Evol.* **2009**, *41*, 11. [CrossRef]
29. Matias, F.I.; Galli, G.; Correia Granato, I.S.; Fritsche-Neto, R. Genomic prediction of autogamous and allogamous plants by SNPs and haplotypes. *Crop Sci.* **2017**, *57*, 2951–2958. [CrossRef]
30. Sun, C.; Wang, B.; Yan, L.; Hu, K.; Liu, S.; Zhou, Y.; Guan, C.; Zhang, Z.; Li, J.; Zhang, J.; et al. Genome-wide association study provides insight into the genetic control of plant height in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* **2016**, *7*, 1102. [CrossRef]

31. Nimmakayala, P.; Abburi, V.L.; Saminathan, T.; Alaparthi, S.B.; Almeida, A.; Davenport, B.; Nadimi, M.; Davidson, J.; Tonapi, K.; Yadav, L.; et al. Genome-wide diversity and association mapping for capsaicinoids and fruit weight in *Capsicum Annuum*, L. *Sci. Rep.* **2016**, *6*, 38081. [CrossRef]
32. Vinholes, P.; Rosado, R.; Roberts, P.; Borém, A.; Schuster, I. Single nucleotide polymorphism-based haplotypes associated with charcoal rot resistance in Brazilian soybean germplasm. *Agron. J.* **2018**, *111*, 182–192. [CrossRef]
33. Nyine, M.; Wang, S.; Kiani, K.; Jordan, K.; Liu, S.; Byrne, P.; Haley, S.; Baenziger, S.; Chao, S.; Bowden, R.; et al. Genotype imputation in winter wheat using first generation haplotype map SNPs improves genome-wide association mapping and genomic prediction of traits. *G3 Genes Genomes Genet.* **2019**, *9*, 125–133. [CrossRef]
34. Calus, M.P.L.; De Roos, A.P.W.; Veerkamp, R.F. Accuracy of genomic selection using different methods to define haplotypes. *Genetics* **2008**, *178*, 553–561. [CrossRef]
35. De Roos, A.P.W.; Schrooten, C.; Druet, T. Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. *J. Dairy Sci.* **2011**, *94*, 4708–4714. [CrossRef]
36. Boichard, D.; Guillaume, F.; Baur, A.; Croiseau, P.; Rossignol, M.N.; Boscher, M.Y.; Druet, T.; Genestout, L.U.C.I.E.; Colleau, J.J.; Journaux, L.; et al. Genomic selection in French dairy cattle. *Anim. Prod. Sci.* **2012**, *52*, 115–120. [CrossRef]
37. Edriss, V.; Fernando, R.L.; Su, G.; Lund, M.S.; Guldbrandtsen, B. The effect of using genealogy-based haplotypes for genomic prediction. *Genet. Sel. Evol.* **2013**, *45*, 5. [CrossRef]
38. Jónás, D.; Ducrocq, V.; Croiseau, P. The combined use of linkage disequilibrium–based haploblocks and allele frequency–based haplotype selection methods enhances genomic evaluation accuracy in dairy cattle. *J. Dairy Sci.* **2017**, *100*, 2905–2908. [CrossRef]
39. Curtis, D.; North, B.V.; Sham, P.C. Use of an artificial neural network to detect association between a disease and multiple marker genotypes. *Ann. Hum. Genet.* **2001**, *65*, 95–107. [CrossRef]
40. Yu, J.; Pressoir, G.; Briggs, W.H.; Bi, I.V.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203. [CrossRef]
41. Jarquín, D.; Kocak, K.; Posadas, L.; Hyma, K.; Jedlicka, J.; Graef, G.; Lorenz, A. Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomes* **2014**, *15*, 740. [CrossRef]
42. Habyarimana, E. Genomic prediction for yield improvement and safeguarding of genetic diversity in CIMMYT spring wheat (*Triticum aestivum* L.). *Aust. J. Crop. Sci.* **2016**, *10*, 127.
43. Ballesta, P.; Serra, N.; Guerra, 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]
44. Thavamanikumar, S.; McManus, L.J.; Ades, P.K.; Bossinger, G.; Stackpole, D.J.; Kerr, R.; Hadjigol, S.; Freeman, J.S.; Vaillancourt, R.E.; Zhu, P.; et al. 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* **2014**, *10*, 1661–1678.
45. 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]
46. Thavamanikumar, S.; McManus, L.J.; Tibbits, J.F.; Bossinger, G. The significance of single nucleotide polymorphisms (SNPs) in *Eucalyptus globulus* breeding programs. *Aust. For.* **2011**, *74*, 23–29. [CrossRef]
47. Cappa, E.P.; El-Kassaby, Y.A.; Garcia, M.N.; Acuña, C.; Borralho, N.M.; Grattapaglia, D.; Poltri, S.N.M. 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* **2013**, *8*, e81267. [CrossRef]
48. Gabriel, S.B.; Schaffner, S.F.; Nguyen, H.; Moore, J.M.; Roy, J.; Blumenstiel, B.; Higgins, J.; DeFelice, M.; Lochner, A.; Faggart, M.; et al. The structure of haplotype blocks in the human genome. *Science* **2002**, *296*, 2225–2229. [CrossRef]
49. Gupta, P.K.; Pawan, S.; Kulwal, P.L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol. Biol.* **2005**, *57*, 461–485. [CrossRef]
50. Fiil, A.; Lenk, I.; Petersen, K.; Jensen, C.S.; Nielsen, K.K.; Schejbel, B.; Andersen, J.R.; Lübberstedt, T. Nucleotide diversity and linkage disequilibrium of nine genes with putative effects on flowering time in perennial ryegrass (*Lolium perenne* L.). *Plant Sci.* **2011**, *180*, 228–237. [CrossRef]

51. Pérez-Rodríguez, P.; Gianola, D.; González-Camacho, J.M.; Crossa, J.; Manès, Y.; Dreisigacker, S. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. *G3 Genes Genomes Genet.* **2012**, *2*, 1595–1605. [CrossRef]
52. Costa e Silva, J.C.; Hardner, C.; Potts, B.M. Genetic variation and parental performance under inbreeding for growth in *Eucalyptus globulus*. *Ann. For. Sci.* **2010**, *67*, 606. [CrossRef]
53. Callister, A.N.; England, N.; Collins, S. Genetic analysis of *Eucalyptus globulus* diameter, straightness, branch size, and forking in Western Australia. *Can. J. For. Res.* **2011**, *41*, 1333–1343. [CrossRef]
54. Mora, F.; Serra, N. Bayesian estimation of genetic parameters for growth, stem straightness, and survival in *Eucalyptus globulus* on an Andean Foothill site. *Tree Genet Genomes* **2014**, *10*, 711–719. [CrossRef]
55. Resende, R.T.; Resende, M.D.V.; Silva, F.F.; Azevedo, C.F.; Takahashi, E.K.; Silva-Junior, O.B.; Grattapaglia, D. Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity* **2017**, *119*, 245. [CrossRef]
56. Tan, B.; Grattapaglia, D.; Martins, G.S.; Ferreira, K.Z.; Sundberg, B.; Ingvarsson, P.K. Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol.* **2017**, *17*, 110. [CrossRef]
57. Pérez, P.; De Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [CrossRef]
58. Van den Berg, I.; Fritz, S.; Boichard, D. QTL fine mapping with Bayes C (π): A simulation study. *Genet. Sel. Evol.* **2013**, *45*, 19. [CrossRef]
59. Suontama, M.; Klápště, J.; Telfer, E.; Graham, N.; Stovold, T.; Low, C.; McKinley, R.; Dungey, H. Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* **2018**, *122*, 370–379. [CrossRef]
60. Müller, B.S.; Neves, L.G.; de Almeida Filho, J.E.; Resende, M.F.; Muñoz, P.R.; dos Santos, P.E.; 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. [CrossRef]
61. Lopez, G.A.; Potts, B.M.; Dutkowski, G.W. Genetic variation and inter-trait correlations in *Eucalyptus globulus* base population trials in Argentina. *For. Genet.* **2002**, *9*, 217–231.
62. Blackburn, D.P.; Hamilton, M.G.; Harwood, C.E.; Baker, T.G.; Potts, B.M. Assessing genetic variation to improve stem straightness in *Eucalyptus globulus*. *Ann. For. Sci.* **2013**, *70*, 461–470. [CrossRef]
63. Bartholomé, J.; Bink, M.C.; van Heerwaarden, J.; Chancerel, E.; Boury, C.; Lesur, I.; Isik, F.; Bouffier, L.; Plomion, C. Linkage and association mapping for two major traits used in the maritime pine breeding program: Height growth and stem straightness. *PLoS ONE* **2016**, *11*, e0165323.
64. Yang, H.; Liu, T.; Xu, B. QTL detection for growth and form traits in three full-sib pedigrees of *Pinus elliottii* var. *elliottii* × *P. caribaea* var. *hondurensis* hybrids. *Tree Genet. Genomes* **2015**, *11*, 130. [CrossRef]
65. Arriagada, O.; Mora, F.; Amaral Junior, A.T. 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]
66. Song, J.; Brendel, O.; Bodénès, C.; Plomion, C.; Kremer, A.; Colin, F. X-ray computed tomography to decipher the genetic architecture of tree branching traits: Oak as a case study. *Tree Genet. Genomes* **2017**, *13*, 5. [CrossRef]
67. Monclus, R.; Leplé, J.C.; Bastien, C.; Bert, P.F.; Villar, M.; Marron, N.; Brignolas, F.; Jorge, V. Integrating genome annotation and QTL position to identify candidate genes for productivity, architecture and water-use efficiency in *Populus* spp. *BMC Plant Biol.* **2012**, *12*, 173. [CrossRef]
68. 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]
69. Valenzuela, C.E.; Ballesta, P.; Maldonado, C.; Baettig, R.; Arriagada, O.; Sousa Mafra, G.; 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]
70. Beaulieu, J.; Doerksen, T.; Clément, S.; MacKay, J.; Bousquet, J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* **2014**, *113*, 343. [CrossRef]

71. Makowsky, R.; Pajewski, N.M.; Klimentidis, Y.C.; Vazquez, A.I.; Duarte, C.W.; Allison, D.B.; de Los Campos, G. Beyond missing heritability: Prediction of complex traits. *PLoS Genet.* **2011**, *7*, e1002051. [CrossRef]
72. Lenz, P.R.; Beaulieu, J.; Mansfield, S.D.; Clément, S.; Desponts, 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 Genome* **2017**, *18*, 335. [CrossRef]
73. Barrett, J.C.; Fry, B.; Maller, J.D.M.J.; Daly, M.J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **2005**, *21*, 263–265. [CrossRef]
74. Myburg, A.A.; Grattapaglia, D.; Tuskan, G.A.; Hellsten, U.; Hayes, R.D.; Grimwood, J.; Jenkins, J.; Lindquist, E.; Tice, H.; Bauer, D.; et al. The genome of *Eucalyptus grandis*. *Nature* **2014**, *510*, 356. [CrossRef]
75. Breseghello, F.; Sorrells, M.E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* **2006**, *172*, 1165–1177. [CrossRef]
76. Hadfield, J.D. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
77. R Core Team. *R: A Language and Environment for Statistical Computing, 3.6.1*; R Foundation for Statistical Computing: Vienna, Austria, 2019.
78. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **1996**, *58*, 267–288. [CrossRef]
79. Legarra, A.; Robert-Granié, C.; Croiseau, P.; Guillaume, F.; Fritz, S. Improved LASSO for genomic selection. *Genet. Res.* **2011**, *93*, 77–87. [CrossRef]
80. Mora, F.; Ballesta, P.; Serra, N. Bayesian analysis of growth, stem straightness and branching quality in full-sib families of *Eucalyptus globulus*. *Bragantia* **2019**, *78*, 1–9. [CrossRef]
81. Torres, L.G.; Rodrigues, M.C.; Lima, N.L.; Trindade, T.F.H.; e Silva, F.F.; Azevedo, C.F.; DeLima, R.O. Multi-trait multi-environment Bayesian model reveals G × E interaction for nitrogen use efficiency components in tropical maize. *PLoS ONE* **2018**, *13*, e0199492. [CrossRef]
82. Volpato, L.; Alves, R.S.; Teodoro, P.E.; de Resende, M.D.V.; Nascimento, M.; Nascimento, A.C.C.; Ludke, W.H.; da Silva, F.L.; Borém, A. Multi-trait multi-environment models in the genetic selection of segregating soybean progeny. *PLoS ONE* **2019**, *14*, e0215315. [CrossRef]
83. Mora, F.; Zúñiga, P.E.; Figueroa, C.R. Genetic variation and trait correlations for fruit weight, firmness and color parameters in wild accessions of *Fragaria chiloensis*. *Agronomy* **2019**, *9*, 506. [CrossRef]
84. Baltunis, B.S.; Huber, D.A.; White, T.L.; Goldfarb, B.; Stelzer, H.E. Genetic gain from selection for rooting ability and early growth in vegetatively propagated clones of loblolly pine. *Tree Genet. Genomes* **2007**, *3*, 227–238. [CrossRef]
85. Burdon, R.D. Short note: Coefficients of variation in variables with bounded scales. *Silvae Genet.* **2008**, *57*, 179–180. [CrossRef]

© 2019 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/).