Tree Genetics & Genomes (2018) 14: 71
https://doi.org/10.1007/s11295-018-1286-5

ORIGINAL ARTICLE

Performance of multi-trait genomic selection for *Eucalyptus robusta* breeding program

Tahina Rambolarimanana[1,2] · **Lolona Ramamonjisoa**[1] · **Daniel Verhaegen**[2,3] · **Jean-Michel Leong Pock Tsy**[2,4] · **Laval Jacquin**[3] · **Tuong-Vi Cao-Hamadou**[3] · **Garel Makouanzi**[5] · **Jean-Marc Bouvet**[2,3]

Received: 29 July 2017 / Revised: 18 March 2018 / Accepted: 14 August 2018 / Published online: 31 August 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

In forest tree genetic improvement, multi-trait genomic selection (GS) may have advantages in improving the accuracy of the genotype estimation and shortening selection cycles. For the breeding of *Eucalyptus robusta*, one of the most exotic planted species in Madagascar, volume at 49 months (V49), total lignin (TL), and holo-cellulose (Holo) were considered. For GS, 2919 single nucleotide polymorphisms (SNP) were used with the genomic best linear unbiased predictor (GBLUP) method, which was as efficient as the reproducing kernel Hilbert space (RKHS) and elastic net methods (EN), but more adapted to multi-trait modeling. The efficiency of individual *I* model, including the genomic data, was much higher than the provenance effect *P* model. For example, with V49, mean goodness-of-fit was: $r_{I\_Full} = 0.79$, $r_{P\_Full} = 0.37$ for *I* and *P*, respectively. The prediction accuracies using the cross-validation procedure were lower for V49: $r_I = 0.29$ $r_P = 0.28$. The genetic gains resulting from the indexes associating (V49, TL) and (V49, Holo) were higher using *I* than for the *P* model; for V49, the relative genetic gain was 37 and 20%, respectively, with 5% of selection intensity. The single-trait approach was as efficient as the multi-trait approach given the weak correlations between V49 and TL or Holo. The *I* model also brings greater diversity: for V49 the number of provenances represented in a selected population was two and three with the *P* model, and 6 and 16 with the *I* model.

Keywords GBLUP · Selection index · Lignin · Cellulose · Genetic diversity · Madagascar

Introduction

The selection methodology has strongly evolved in plants by incorporating molecular information and developing the concept of marker-assisted selection (Bernardo 2008) to optimize ideotype creation (Donald 1968). Molecular markers have been used to define quantitative trait loci (Robbins et al. 2002), but this inefficient approach (Jannink 2010) was supplanted by genomic selection (GS) (Meuwissen et al. 2001), particularly since the introduction of low-cost high-speed sequencing methods.

Genomic selection (GS) method was quickly implemented in animal breeding (Luan 2009; Hayes and Goddard 2010; Meuwissen and Goddard 2010) and more recently in plants (Heffner et al. 2009; Heffner et al. 2010). GS is very beneficial for perennial crops because the length of the selection cycle can be significantly shortened by the selection of individuals at the juvenile stage using genomic markers to predict the genetic value of adults, in forest trees for example in pines (Isik et al. 2016) or *Eucalyptus* (Grattapaglia and Resende 2010; Grattapaglia 2014; Bouvet et al. 2015) or oil palm

---

Communicated by A. A. Myburg

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-018-1286-5) contains supplementary material, which is available to authorized users.

✉ Jean-Marc Bouvet
jean-marc.bouvet@cirad.fr

[1] ESSA, Ecole Supérieure des Sciences Agronomiques, Université d'Antananarivo, Antananarivo, Madagascar

[2] Dispositif de recherche et d'enseignement en partenariat "Forêts et Biodiversité à Madagascar", Cirad-FOFIFA-Université d'Antananarivo, Antananarivo, Madagascar

[3] UMR AGAP, Amélioration Génétique et Adaptation des Plantes Tropicales et Méditerranéennes, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

[4] DRFGRN, Département des Recherches Forestières et de Gestion des Ressources Naturelles, FOFIFA, Antananarivo, Madagascar

[5] ENSAF, Ecole Supérieure des Sciences Agronomiques et Forestières, Université Marien NGOUABI, Brazzaville, République du Congo

(Cros et al. 2015). Simulation and experimental studies in the context of perennial plants showed that GS increased the genetic gain per unit of time (Grattapaglia and Resende 2010; Denis and Bouvet 2011, 2013).

Most GS studies carried out in plant only considered the single-trait approach and very few have addressed the potential of GS with a multi-trait approach. The accuracy of multi-trait genomic selection (MT-GS) using different methods has first been addressed by simulation (Calus and Veerkamp 2011). Jia and Jannink (2012) showed that the predictive accuracy of a trait that has low heritability can be increased by MT-GS when traits correlated with high heritability are available. Similar results were obtained for perennial crops like oil palm (Marchal et al. 2016) and other plants (Guo et al. 2014).

Additional studies are therefore needed in plants to explore the potential of MT-GS in different breeding contexts. Among the numerous perennial crops of interest in this regard, the genus *Eucalyptus* is particularly interesting. Widely planted in the tropical and Mediterranean zones, this genus has contributed significantly to the production of pulpwood and fuel wood, with more than 20 million hectares of plantations (Vigneron and Bouvet 1997), and the development of numerous breeding programs stresses the need to implement new selection methods to optimize it.

We implemented our study with *Eucalyptus robusta*, a species planted in Madagascar for fuel wood and covering about 150,000 ha of plantations managed by smallholders. This species is particularly important since 80% of the population uses its wood for domestic purposes. Two main types of traits are important to consider: volume and two chemical components of wood, namely lignin and cellulose (Higuchi 1997). The place of these traits in selection criteria is critical according to the selection goal. On one side, if the objective is to increase the fuel wood quality, more lignin is recommended, on the other side, if pulp production is the target more cellulose is needed.

Moreover, this species was introduced 100 years ago in Madagascar (Verhaegen et al. 2011); the first results concerning genetic diversity estimated by molecular markers suggested that the seed origin was from a single Australian zone and that the level of relatedness was high within plantations (Daniel Verhaegen, personal communication). One of the best strategies for future plantations would therefore be to put in place a new improvement program using a broader genetic base and innovative breeding methods. The selection method is crucial for establishing the base population of parents that will favorably recombine their genes during subsequent generations. As observed for other *Eucalyptus* species (Hung et al. 2015), multi-trait selection is one option that can lead to Eucalyptus plantations with improved productivity and wood quality.

The main objective of this study was to evaluate the performance of multi-trait genomic selection (MT-GS) in the context of genetic improvement of *Eucalyptus robusta* by (i) analyzing the variability of growth and chemical properties of wood within a provenance trial covering the distribution area of *E. robusta* in Australia, (ii) identifying the best linear mixed model incorporating genomic data for genetic value prediction, and (iii) comparing single- and multi-trait-based selection.

Materials and methods

Population studied

We conducted the present study with a *Eucalyptus robusta* provenance trial located in the central eastern part of Madagascar at the Mahela Forest Station (19 ° 01 ′16.4′ ′S; 48 ° 57′ 31.7 ″E). It was established in 1993 and included 19 Australian provenances sampled in the entire area of natural distribution of the species (see the identity and geographical information of their origin in Table 1). The experimental design consisted of 18 unbalanced blocks of 15 to 19 plots, each plot representing a single provenance. The plots were randomly distributed into the blocks. Originally, the total number of plots was 324 and each plot was planted with 16 (4 × 4) trees from the same provenance (5184 trees in total). After successive thinnings, 1314 trees remained within the trial, representing an approximate selection rate of 25% in each provenance. Among those 1314 individuals, 415 were randomly retained for measurement, representing on average 22 trees per provenance. The trees were genotyped by single nucleotide polymorphism (SNP) markers (see below) and 395 were measured by near infrared spectroscopy to estimate their wood chemical components.

Variables studied

We considered three critical traits for the production of fuel wood or pulpwood: stem volume at 49 months (V49), the total lignin (TL) and the holo-cellulose content (Holo). The volume was calculated as a function of the height (H) and the stem circumference at breast height (C) of trees with the formula: $V = (0.3 \times (H-130) + 130 \times 4 \times \pi \times C) / 1000$. In this case, the volume is expressed in dm^3. The age considered was relatively young, 49 months, to minimize the important environmental effect on growth traits due to wind damage after 60 months. Chemical properties, including TL and Holo, were analyzed at the age of 20 years by near infrared spectroscopy on the 395 trees. For chemical predictions, we used existing near infrared spectroscopy models of multispecies *Eucalyptus* that included samples from this study (Denis et al. 2013).

Table 1 Identity and geographical origin on the *E. robusta* provenances established in the Mahela trial in Madagascar

No. id	No. seed lot	Provenance name	Australian state	Long (E)	Lat (S)	Alt (m)
1	7300500–7300505 N	SF 1004 East Gympie	Queensland	153°00	26°10	3
2	8002760 N	Fraser Island	Queensland	153°02	25°29	5
3	8003233 N	Elliot River, South Bundaberg	Queensland	152°18	25°02	100
4	9309759 N	Coopernook SF	New South Wales	152°37	31°50	60
5	7301083–7301092 N	Bowenia SF 20	Queensland	150°37	22°57	100
6	9309758 N	Bulahdelah Wallingat	New South Wales	152°13	32°24	270
7	9309760 N	Nambucca SF	New South Wales	153°00	30°39	30
8	9309761 N	13 km North Coffs Harbour	New South Wales	153°10	30°14	30
9	9309762 N	Newfoundland SF	New South Wales	153°00	29°58	13
10	9309763 N	48 km South Grafton	New South Wales	152°58	29°50	18
11	9309764 N	12 km West Nambour	Queensland	152°51	26°37	300
12	9309765 N	East South East of Nambour	Queensland	154°04	26°40	5
13	9309766 N	South Grafton	New South Wales	152°57	29°45	80
14	9309767 N	Woodford	Queensland	152°50	26°57	50
15	9309769 N	Myall Lakes NP	New South Wales	152°22	32°27	4
16	9309770 N	North Coffs Harbour	New South Wales	153°09	30°15	9
17	9309771 N	Noosa SF	Queensland	152°58	26°11	58
18	9309772 N	Byfield SF	Queensland	150°38	22°55	38
19	9309773 N	David Low WY Brisbane	Queensland	153°06	26°36	50

Long: Longitude, E: East, S: South

Lat: Latitude, Alt (m): altitude (meters)

Molecular markers

We used SNP markers. This information was then integrated into different models to predict genetic value.

DNA was extracted from the leaves of 415 trees that were genotyped using the DArTseq technology (Wenzl et al. 2004). Of the 30,000 SNPs identified, 2919 were selected on the basis of repeatability, a minor allele frequency higher than 2.5%, and no missing data.

Statistical models associating phenotypes and genotypes

A first approach consisted in detecting the models that best fitted the data. Except for the general mean, all effects were considered random (block, provenance, individual within provenance, and residual). The plot effect was not explicitly included in the model as plots were only represented by two individuals on average.

The different models were tested according to a hierarchical approach. The models were compared according to the Akaike information criterion (Akaike 1974). Three models have shown Akaike information criteria that are very close and stand out from the other models tested (see the comparison of models in supplementary material no. 1a). These are the mixed linear model with the random "block" and "provenance" effects (*P* model), the model with the random "block" and "individual" effects (*I* model) and the model with the random "block", "provenance" and "individual" (*PI* model). Given the small differences between these three models (see the comparison of models in supplementary material no. 1a), we decided to test only the *P* and *I* models in the multi-trait selection. It was important to consider the *P* model in the following analyses as a baseline method, because in Madagascar, provenance selection and the reintroduction of the best provenances remain a very efficient option to get genetic gain.

We then tested three methods of predicting the individual genetic value associating the "block" and the "SNP marker" effects of each individual. These are: elastic net (EN), genomic best linear unbiased predictor (GBLUP) and reproducing kernel Hilbert space (RKHS) (Jacquin et al. 2016). The results showed that these models are very similar in terms of the quality of the prediction (see supplementary material no. 1b). We finally used the GBLUP method, because it performs best in many GS studies (Cros et al. 2015). Moreover, its programming for a multivariate approach is facilitated by ASReml software (Gilmour et al. 2009). Following these preliminary analyses, the two linear mixed models are presented as follows.

The provenance model:

$$\mathbf{y_i} = \mu_\mathbf{i} + \mathbf{X_i}\,\beta_\mathbf{i} + \mathbf{Q_i}\,\lambda_\mathbf{i} + \varepsilon_\mathbf{i} \ (P \text{ model})$$

The individual model:

$$\mathbf{y_i} = \boldsymbol{\mu}_\mathbf{i} + \mathbf{X_i}\ \boldsymbol{\beta}_\mathbf{i} + \mathbf{Z_i}\ \mathbf{u_i} + \varepsilon_\mathbf{i}\ \text{(I model)}$$

where $\mathbf{y}_{i\,(415\times 1)}$ is the vector of phenotypic values for trait, $\boldsymbol{\mu}_i$ is a vector corresponding to the general mean of the trait i, $\boldsymbol{\beta}_{i\,(18\times 1)} \sim N(0, \boldsymbol{I}_d\, \sigma^2_\mathbf{b})$ is the random block effect, $\boldsymbol{\lambda}_{i\,(19\times 1)} \sim N(0, \boldsymbol{I}_d \boldsymbol{\sigma}^2_\mathbf{prov})$, is the vector of random provenance effect, $\mathbf{u}_{i\,(415\times 1)} \sim N(0, \sigma^2_\mathbf{a}\mathbf{A})$ is the vector of random individual additive effect where, $\mathbf{A}$ is the relationship matrix among individuals (estimation is given in the next paragraph), and $\varepsilon_{i\,(415\times 1)} \sim N(0, \boldsymbol{I}_d\ \sigma^2_\mathbf{r})$ is the random residual effect. $\boldsymbol{X}_{i\,(415\times 18)}$, $\boldsymbol{Q}_{i\,(415\times 19)}$, and $\mathbf{Z}_{i\,(415\times 415)}$ are the incidence matrices for "block", "provenance", and "individual" effects, respectively and $\boldsymbol{I}_\mathbf{d}$ is the identity matrix.

The formulation of the single and multi-trait models is given in matrix form as an example for the *I* model:

- Individual model with single-trait selection (*I*-ST) for trait 1:

$$\boldsymbol{y}_1 = \boldsymbol{\mu}_1 + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_1\boldsymbol{u}_1 + \boldsymbol{\varepsilon}_1$$

- Individual model with bivariate selection (*I*-MT2) for traits 1 and 2:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}_1 & 0 \\ 0 & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_1 & 0 \\ 0 & \boldsymbol{Z}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

Random block effect: $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{b(1,1)} & \sigma_{b(1,2)} \\ \sigma_{b(2,1)} & \sigma^2_{b(2,2)} \end{bmatrix} \otimes \boldsymbol{I}_d\right)$

Random individual (additive) effect:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{a(1,1)} & \sigma_{a(1,2)} \\ \sigma_{a(2,1)} & \sigma^2_{a(2,2)} \end{bmatrix} \otimes \mathbf{A}\right)$$

- Combined individual model with trivariate selection (I-MT3) for traits 1, 2 and 3:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}_1 & 0 & 0 \\ 0 & \boldsymbol{X}_2 & 0 \\ 0 & 0 & \boldsymbol{X}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_1 & 0 & 0 \\ 0 & \boldsymbol{Z}_2 & 0 \\ 0 & 0 & \boldsymbol{Z}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \boldsymbol{u}_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

random block effect $\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{b(1,1)} & \sigma_{b(1,2)} & \sigma_{b(1,3)} \\ \sigma_{b(2,1)} & \sigma^2_{b(2,2)} & \sigma_{b(2,3)} \\ \sigma_{b(3,1)} & \sigma_{b(3,2)} & \sigma^2_{b(3,3)} \end{bmatrix} \otimes \boldsymbol{I}_d\right)$

random individual (additive) Effect

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{a(1,1)} & \sigma_{a(1,2)} & \sigma_{a(1,3)} \\ \sigma_{a(2,1)} & \sigma^2_{a(2,2)} & \sigma_{a(2,3)} \\ \sigma_{a(3,1)} & \sigma_{a(3,2)} & \sigma^2_{a(3,3)} \end{bmatrix} \otimes \mathbf{A}\right)$$

For the provenance model, the single, bivariate and trivariate approaches were abbreviated P-ST, P-MT2, and P-MT3, respectively, and the abbreviations I-ST, I-MT2, and I-MT3 for the individual model.

Relationship matrix A

The relationship matrix used in the *I* model was defined on the basis of the relationship coefficients between the $n = 415$ individuals. The relationship coefficient, using the SNP markers, was defined as the probability that two alleles taken at random are identical by state and was estimated using Van Raden's formula (Van Raden 2007, 2008). The relationship matrix $\mathbf{A}$ $[n, n]$ was calculated as follows:

$$\mathbf{A} = \frac{(\mathbf{M}-\mathbf{P})(\mathbf{M}-\mathbf{P})'}{2\sum_1^s p_i(1-p_i)}$$

where $\mathbf{M}$ [n, s] is the matrix corresponding to the genotype for all loci of all individuals. Each column of M is coded 0, 1, and 2 for genotypes AA, Aa, and aa, respectively, representing the number of minor alleles; and "s" is the total number of SNPs (2919 SNPs). The matrix $\mathbf{P}$ [n, s] contains the frequencies (pi) of the second allele at each locus such that the column i of $\mathbf{P}$ is 2**pi**.

Estimation of variance components, heritabilities, and correlations

The variance components were estimated with ASReml-R (Gilmour et al. 2009). The solutions (BLUP) were obtained by solving the mixed linear model equations (Mrode and Thompson 2005). For the two models, the heritabilities at the provenance level h^2_prov and individual level h^2_i were estimated from the variance components.

P model

$$h^2_\text{prov} = \frac{\sigma^2_\text{prov}}{\sigma^2_\text{b} + \sigma^2_\text{prov} + \dfrac{\sigma^2_\text{r}}{N\text{prov}}}$$

I model

$$h^2_\text{i} = \frac{\sigma^2_\text{a}}{\sigma^2_\text{b} + \sigma^2_\text{a} + \sigma^2_\text{r}}$$

where *N*prov is the harmonic mean of the number of individuals per provenance.

The correlations related to genetic effects between two traits (1 and 2) were calculated from variance/covariance estimates with the following formulas:

the additive genetic correlation $\rho_\text{a} = (\text{cov}_{\text{a}(1,2)}) / (\sigma_{\text{a}1}) (\sigma_{\text{a}2})$,
the provenance correlation $\rho_\text{rov} = (\text{cov}_{\text{prov}(1,2)}) / (\sigma_{\text{prov}1}) (\sigma_{\text{prov}2})$,
the residual correlation $\rho_\text{r} = (\text{cov}_{\text{r}(1,2)}) / (\sigma_{\text{r}1}) (\sigma_{\text{r}2})$,
the phenotypic correlation $\rho_\text{p} = (\text{cov}_{\text{p}(1,2)}) / (\sigma_{\text{p}1}) (\sigma_{\text{p}2})$,

where $\mathrm{cov}_{(1,2)}$ is the covariance between traits 1 and 2 and σ_1 and σ_2 are the standard deviations of traits 1 and 2, respectively.

The phenotypic variance $\sigma^2{}_\mathrm{p}$ and covariance cov_p were calculated as follows:

$$\sigma^2{}_\mathrm{p} = \sigma^2{}_\mathrm{b} + \sigma^2{}_\mathrm{prov} + \sigma^2{}_\mathrm{r} \text{ and } \mathrm{cov}_{\mathrm{p}(1,2)} = \mathrm{cov}_{\mathrm{b}(1,2)} + \mathrm{cov}_{\mathrm{prov}(1,2)} + \mathrm{cov}_{\mathrm{r}(1,2)} \text{ for the } P \text{ model}$$

$$\sigma^2{}_\mathrm{p} = \sigma^2{}_\mathrm{b} + \sigma^2{}_\mathrm{a} + \sigma^2{}_\mathrm{r} \text{ and } \mathrm{cov}_{\mathrm{p}(1,2)} = \mathrm{cov}_{\mathrm{b}(1,2)} + \mathrm{cov}_{\mathrm{a}(1,2)} + \mathrm{cov}_{\mathrm{r}(1,2)} \text{ for the } I \text{ model}$$

The CVp, CVa, and CVr are respectively the phenotypic, additive, and the residual coefficients of variation and are defined by the ratio standard deviation to the mean over the population.

Goodness-of-fit of the models

The models were compared by evaluating the goodness-of-fit, using the full data set, i.e., 415 trees. The goodness-of-fit was assessed by the correlation between estimated genetic values and phenotypes.

With the *P* model, we estimated the genetic value at provenance level $\hat{\mathrm{G}}_\mathbf{P}$ (which is the same for all individuals from the same provenance) by the sum of the general mean $\hat{\mu}_\mathrm{i}$ and the BLUPs of the provenance effect $\boldsymbol{Q}_{\boldsymbol{i}}\widehat{\lambda}_\mathrm{i}$.

$$\hat{\mathrm{G}}_\mathrm{P} = \widehat{\mu}_\mathrm{i} + \mathrm{Q}_\mathrm{i}\widehat{\lambda}_i$$

The goodness-of-fit was calculated using the Spearman correlation between $\hat{\mathrm{G}}_\mathrm{P}$ and the observed phenotype **Y**: $\mathrm{r}_{\mathrm{P\_Full}} = \mathrm{cor}\ (\hat{\mathrm{G}}_\mathrm{P}\ ,\mathbf{Y})$.

With the *I* model, we estimated the individual genetic value $\hat{\mathbf{G}}_\mathbf{I}$ by the sum of the general mean $\hat{\mu}_\mathrm{i}$ and the BLUPs of the individual genetic effect (additive) $\boldsymbol{Z_i}\hat{u}_\mathrm{i}$.

$$\hat{\mathrm{G}}_\mathrm{I} = \widehat{\mu}_\mathrm{i} + Z_\mathrm{i}\widehat{U}_i$$

The goodness-of-fit was calculated using the Spearman correlation between $\hat{\mathbf{G}}_\mathbf{I}$ and the observed phenotype **Y**: $\mathrm{r}_{\mathrm{I\_Full}} = \mathrm{cor}\ (\hat{\mathrm{G}}_\mathrm{I}\ ,\mathbf{Y})$.

These correlation coefficients were used to compare the single-trait (ST), bivariate, (MT2), and trivariate (MT3) selection according to the two models (*P* and *I*) leading to the six situations: (P-ST, P-MT2, P-MT3), (I-ST, I-MT2, I-MT3).

Evaluation of model prediction accuracy

The comparison of the models was also addressed by evaluating their ability to predict the genotype with unobserved phenotypic data. The prediction accuracy of the three models associating the multi-trait approach was carried out by cross-validation. We made up the calibration population by randomly selecting two-thirds of the total population and the validation population was composed of the remaining one-third. The calibration population was used to estimate the variance components that were then integrated into the prediction equations of the mixed linear model (Mrode and Thompson 2005) to estimate the genetic values of the population of validation. Previously, the accuracy of the prediction was measured by the Spearman correlation between the predicted genetic and phenotypic values of the validation population for the two models: $\mathrm{r_P} = \mathrm{cor}\ (\hat{\mathrm{G}}_\mathrm{P}\ , \mathbf{Y})$ and $\mathrm{r_I} = \mathrm{Cor}\ (\hat{\mathrm{G}}_\mathrm{I}\ , \mathbf{Y})$. We used 100 cross-validation procedures to estimate the variability of the prediction accuracy.

Estimation of genetic gains using selection index

The genetic gain was calculated using a selection index defined as the linear combination of the standardized genetic values of each trait estimated using the multi-trait approach with the two models, and coefficients (weights) depending on the economic importance given to the different traits (Gallais and Poly 1990). The objective was to provide a selection method combining the volume production and the chemical wood properties. We therefore considered the indexes based on bivariate analyses (V49, TL) and (V49, Holo), which correspond respectively to two breeding objectives: fuelwood and pulpwood production. To test different scenarios, we chose three pairs of weights (0.75, 0.25), (0.50, 0.50), and (0.25, 0.75), the first weight being associated with V49 and the second with TL or Holo. For each pair of traits, the three indexes followed the formulas below:

Index combining V49 and TL

$$\mathrm{Index}_{75V49,25TL} = 0.75\frac{\hat{G}_{V49}}{\sigma\left(\hat{G}_{V49}\right)} + 0.25\frac{\hat{G}_{TL}}{\sigma\left(\hat{G}_{TL}\right)}$$

$$\mathrm{Index}_{50V49,50TL} = 0.50\frac{\hat{G}_{V49}}{\sigma\left(\hat{G}_{V49}\right)} + 0.50\frac{\hat{G}_{TL}}{\sigma\left(\hat{G}_{TL}\right)}$$

$$\mathrm{Index}_{25V49,75TL} = 0.25\frac{\hat{G}_{V49}}{\sigma\left(\hat{G}_{V49}\right)} + 0.75\frac{\hat{G}_{TL}}{\sigma\left(\hat{G}_{TL}\right)}$$

Index combining V49 and Holo

$$\mathrm{Index}_{75V49,25Holo} = 0.75\frac{\widehat{G}_{V49}}{\sigma_{\widehat{\left(G_{V49}\right)}}} + 0.25\frac{\widehat{G}_{Holo}}{\sigma_{\widehat{\left(G_{Holo}\right)}}}$$

$$\mathrm{Index}_{50V49,50Holo} = 0.50\frac{\widehat{G}_{V49}}{\sigma_{\widehat{\left(G_{V49}\right)}}} + 0.50\frac{\widehat{G}_{Holo}}{\sigma_{\widehat{\left(G_{Holo}\right)}}}$$

$$\mathrm{Index}_{25V49,75Holo} = 0.25\frac{\widehat{G}_{V49}}{\sigma_{\widehat{\left(G_{V49}\right)}}} + 0.75\frac{\widehat{G}_{Holo}}{\sigma_{\widehat{\left(G_{Holo}\right)}}}$$

where $\sigma_{(\hat{G}_{V49})}$, $\sigma_{(\hat{G}_{V49})}$ And $\sigma_{\widehat{(G_{Holo})}}$ are the standard deviation of the genetic value of V49, TL, and Holo.

The genetic gain ΔG induced by the index selection for the trait T (V49, TL or Holo) was calculated using the formula:

$$\Delta G = 100*\frac{\overline{T}s-\overline{T}tot}{\overline{T}tot}$$

where $\overline{T}s$ is the mean of the individuals selected for the trait T based on the index ranking, and $\overline{T}tot$ is the population mean.

The genetic gains obtained through the previous index based on the multi-trait approach were compared with genetic gains based on the single-trait approach. We considered three selection intensities that can be used in the context of breeding program in Madagascar, 5, 10 and 15%.

Results

Phenotypic performance of the provenances

Considering the total population, we observed that the mean volume was 104 dm^3 and that trees contained on average 61% Holo and 37% TL (Table 2). Volume had a high variability with a coefficient of variation of 35%, while TL and Holo showed very low variability, with coefficients of variation less than 6%, which is generally observed with this type of variable in *Eucalyptus* (Hein et al. 2012; Makouanzi et al. 2017). Although measured trees resulted from successive thinning,

Table 2 Mean phenotypic performance of the 19 provenances in volume at 49 months (V49), total lignin (TL), and holo-cellulose (Holo) content

			Volume at 49 months (V49 in dm^3)		Total lignin content (TL in %)		Holo-cellulose content (Holo in %)	
No. id	Provenance name	N	Mean	CVp	Mean	CVp	Mean	CVp
1	SF 1004 East Gympie	23	109.24	28.97	37.04	5.35	59.99	4.41
2	Fraser Island	26	123.65	27.93	35.56	5.41	62.33	4.24
3	Elliot River, South Bundaberg	24	122.29	43.14	36.65	5.07	60.43	5.01
4	Coopernook SF	24	124.98	32.04	36.09	5.21	61.27	4.87
5	Bowenia SF 20	24	114.06	30.92	37.29	6.42	59.95	4.79
6	Bulahdelah Wallingat	20	86.78	35.35	35.64	7.55	61.46	7.92
7	Nambucca SF	19	93.69	25.86	36.27	4.07	61.1	3
8	13 km North Coffs Harbour	20	92.29	33.48	36.32	6.42	61.13	4.54
9	Newfoundland SF	22	82.42	31.55	37.28	5.7	59.67	4.73
10	48 km South Grafton	23	112.65	27.69	36.92	3.83	59.69	3.37
11	12 km West Nambour	16	85.76	29.98	35.98	4.91	61.6	4.56
12	East South East of Nambour	20	100.58	38.41	36.29	3.96	60.86	3.83
13	South Grafton	23	104.07	33.23	36.05	4.32	60.98	4.19
14	Woodford	21	108.54	27.08	36.61	2.85	61.22	2.77
15	Myall Lakes NP	23	85.38	33.86	36.12	4.72	60.77	4.33
16	North Coffs Harbour	22	81.59	32.52	36.84	6.13	59.46	5.44
17	Noosa SF	22	117.33	29.04	35.85	4.55	61.95	5.1
18	Byfield SF	22	105.6	46.82	36.94	5.19	60.11	4.05
19	David Low WY Brisbane	21	113.85	26.21	36.04	5.84	61.95	4.66
	General mean	22	104.1	35.15	36.50	5.29	60.84	4.66

N: number of trees measured in each provenance

The total number of trees is 409 trees for V49 and 325 trees for TL and Holo reducing the population from 5184 to 1314, of which 415 were analyzed, the total variability within the trial remained high in line with variability observed in similar populations (Denis et al. 2013).

The results presented for each provenance showed a clear difference in V49 among provenances, varying from 81.59 to 124.98 dm^3. As expected, this variation was less marked for TL, which varied from 35.56 to 37.29%, and Holo, which varied from 59.46 to 62.33%.

Variance components and heritabilities

For the *P* model (Table 3), whatever the trait, the algorithm converged easily leading to the variance component estimation. The results showed the higher variance of the residual effect followed by the variance of the block and that of the provenance effects. This result emphasized the preponderance of individual trees coupled to micro-environmental variability within the experimental design. This distribution of variance among the different effects remained stable using the single- and multi-trait approaches, indicating that covariance between traits did not contribute that much to trait variation. The mean heritability in the single- and multi-trait approaches was higher for V49, $h^2_{prov} = 0.35$ than for wood chemistry traits; $h^2_{prov} = 0.25$ for Holo and $h^2_{prov} = 0.23$ for TL.

With the *I* model (Table 4), the residual variance remained preponderant, followed by the block and the additive variance. Compared with the *P* model (Table 3), we observed a decrease in the residual variance resulting from a partial redistribution into the additive variance. We noted, with the *I* model, in the case of trivariate analysis that the algorithm failed to converge correctly, and therefore to give solutions. We tried to improve the convergence using the factor analysis model with ASReml, but similarly the algorithm failed to converge with trivariate model. We suspect that the important rate of missing values for wood traits explains this non-convergence. We observed a decrease in heritabilities compared with the *P* model, which can be explained by a higher residual variance. As for the *P* model, TL and Holo presented lower heritability than V49, $h^2_i = 0.05$ for TL and Holo, with $h^2_i = 0.16$ for V49, whatever the multi-trait approach (Table 4).

Correlation between traits

The phenotypic and genetic correlations between the three traits were estimated for each model (Table 5). With the *P* model, the correlations between V49/TL and V49/Holo for block, provenance, and residual effects remained moderate to low, with high standard deviations. For example, for the provenance effect, $\rho_{prov} = 0.41$ with a standard deviation (SD) of 0.37, underlining the imprecision of the estimate. The correlations between TL/Holo were high for provenance $\rho_{prov} = -0.92$ and residual $\rho_r = -0.91$ with a low standard deviation of SD = 0.11 and SD = 0.01, respectively, underlining a strong opposition in the expression of these two traits.

With the *I* model, we also found for V49/TL and V49/Holo moderate correlations for block, individual, and residual effects, with high standard deviations. Estimates were on average close to zero.

Goodness-of-fit of the models

With the *P* model, the goodness-of-fit was higher for V49 ($r_{P\_Full} = 0.37$) than for TL (r_{P_Full} ranging from 0.23 to 0.25) and Holo (r_{P_Full} ranging from 0.27 to 0.29) (Table 3). The bi-

Table 3 Variance components, heritability, and goodness-of-fit using the GBLUP method with the single-trait provenance model (P-ST), bivariate provenance model (P-MT2), and trivariate provenance model (P-MT3)

Targeted trait	Associated traits	Model	$\hat{\mu}$	$\hat{\sigma}^2_b$ [SD]	$\hat{\sigma}^2_{prov}$ [SD]	$\hat{\sigma}^2_r$ [SD]	h^2_{prov} [SD]	r_{P_full}
V49	V49	P-ST	102.59	298.65 [118.70]	183.55 [76.14]	927.43 [67.95]	0.35 [0.12]	0.37
	(V49, Holo)	P-MT2	102.60	298.61 [118.70]	182.90 [75.93]	927.75 [67.98]	0.35 [0.12]	0.37
	(V49, TL)	P-MT2	102.59	298.17 [118.54]	183.46 [76.11]	927.55 [67.96]	0.35 [0.12]	0.37
	(V49, Holo, TL)	P-MT3	102.51	300.08 [119.18]	184.20 [76.35]	927.42 [67.94]	0.35 [0.12]	0.37
TL	TL	P-ST	36.44	0.15 [0.12]	0.09 [0.10]	3.49 [0.29]	0.24 [0.21]	0.25
	(TL, V49)	P-MT2	36.45	0.15 [0.12]	0.09 [0.10]	3.49 [0.29]	0.23 [0.20]	0.25
	(TL, Holo)	P-MT2	36.45	0.13 [0.11]	0.08 [0.10]	3.51 [0.29]	0.23 [0.21]	0.23
	(TL, V49, Holo)	P-MT3	36.46	0.13 [0.11]	0.08 [0.09]	3.51 [0.29]	0.23 [0.21]	0.23
Holo	Holo	P-ST	60.79	0.70 [0.37]	0.35 [0.26]	7.00 [0.58]	0.26 [0.16]	0.29
	(Holo, V49)	P-MT2	60.75	0.71 [0.38]	0.33 [0.25]	7.00 [0.58]	0.25 [0.15]	0.28
	(Holo, TL)	P-MT2	60.77	0.67 [0.36]	0.33 [0.25]	7.03 [0.58]	0.25 [0.16]	0.28
	(Holo, TL, V49)	P-MT3	60.73	0.67 [0.36]	0.31 [0.24]	7.03 [0.58]	0.24 [0.16]	0.27

SD: standard deviation

Table 4 Variance components, heritability, and goodness-of-fit using the GBLUP method with the single-trait individual model (I-ST), bivariate individual model (I-MT2), and trivariate individual model (I-MT3)

Targeted trait	Associated traits	Model	$\hat{\mu}$	$\hat{\sigma}^2_b$ [SD]	$\hat{\sigma}^2_a$ [SD]	$\hat{\sigma}^2_r$ [SD]	h^2_i [SD]	r_{I_full}
V49	V49	I-ST	103.34	294.42 [117.67]	278.96 [77.65]	648.69 [97.79]	0.23 [11.9]	0.79
	(V49, Holo)	I-MT2	103.34	293.98 [117.54]	279.93 [77.87]	647.80 [97.89]	0.23 [0.07]	0.79
	(V49, TL)	I-MT2	103.35	287.32 [115.22]	279.56 [77.84]	648.37 [97.89]	0.23 [0.07]	0.79
	(V49, Holo, TL)	I-MT3	103.19	290.74 [116.15]	Unconstrained	616.3 [0 97.12]	NA	NA
TL	TL	I-ST	36.43	0.15 [0.12]	0.18 [0.19]	3.31 [0.38]	0.05 [0.07]	0.66
	(TL, V49)	I-MT2	36.42	0.15 [0.12]	0.16 [0.19]	3.33 [0.38]	0.04 [0.05]	0.54
	(TL, Holo)	I-MT2	36.44	0.14 [0.12]	0.22 [0.2]	3.25 [0.38]	0.06 [0.06]	0.63
	(TL, V49, Holo)	I-MT3	36.44	0.14 [0.11]	Unconstrained	3.21 [0.39]	NA	NA
Holo	Holo	I-ST	60.80	0.72 [0.38]	0.70 [0.5]	6.27 [0.84]	0.09 [0.12]	0.74
	(Holo, V49)	I-MT2	60.79	0.75 [0.39]	0.69 [0.5]	6.25 [0.84]	0.09 [0.07]	0.73
	(TL, Holo)	I-MT2	60.77	0.70 [0.38]	0.78 [0.5]	6.15 [0.82]	0.10 [0.07]	0.70
	(TL, V49, Holo)	I-MT3	60.76	0.72 [0.38]	Unconstrained	6.16 [0.83]	NA	NA

Unconstrained: this parameter cannot be estimated due to convergence difficulty of the algorithm

NA: this parameter cannot be estimated due to non-convergence of the algorithm

SD: standard deviation

and tri-variate approaches did not improve and even seemed to worsen accuracy, but variations did not exceed 0.02.

With the *I* model, the goodness-of-fit was higher than for the *P* model with $r_{I\_Full} = 0.79$ for V49, varying from $r_{I\_Full} = 0.54$ to 0.66 for TL and from $r_{I\_Full} = 0.70$ to 0.74 for Holo (Table 4). With this model, the bi- and tri-variate approaches did not improve the goodness-of-fit, the correlations sometimes being lower than with the single-trait approach.

Prediction accuracy of the models

The effect of models and multi-trait options on prediction accuracy, using the cross-validation procedure, is illustrated in Table 6. Only one model with individual genetic estimation was considered. For V49, the prediction accuracy was, on average, $r_I = 0.30$ for the *I* model (Table 6), while goodness-of-fit was $r_{I\_Full} = 0.79$ (Table 5). For TL, mean accuracy was $r_I = 0.05$ (Table 7), while with the whole dataset goodness-of-fit was $r_{I\_Full} = 0.60$ (Table 4). Finally, for Holo, mean accuracy was $r_I = 0.09$ (Table 6), while with the whole dataset goodness-of-fit was $r_{I\_Full} = 0.72$ (Table 4).

Evolution of genetic gains by index selection

The evolution of genetic gains for each trait is presented according to the three pairs of economic coefficients (0.75/0.25, 0.50/0.50, and 0.25/0.75) and the two models (Figs. 1 and 2). The gain obtained by index selection was compared to the gain obtained by single trait selection.

For V49, the trends in genetic gain were presented for two index types: V49 associated with TL and V49 associated with Holo. As expected, the gain decreased when the selection intensity decreased, i.e., when the percentage of trees selected varied from 5 to 15%. This decrease is generally moderate, around 5% (Fig. 1a and b). We also noted a decrease in the

Table 5 Phenotypic correlations (ρ_P), environmental block effect correlations (ρ_{Bloc}), provenance correlation (ρ_{prov}), additive correlations (ρ_a), and residual correlation (ρ_r) according to the three linear mixed models

Bi-variate model	Associated traits	ρ_P	ρ_{Bloc}	ρ_{prov} [SD]	ρ_a [SD]	ρ_r [SD]
P model: provenance model (P-MT2)	(V49, Holo)	0.17	0.29 [0.30]	0.41 [0.37]	NA	0.13 [0.06]
	(V49, TL)	− 0.07	− 0.05 [0.40]	− 0.09 [0.48]	NA	− 0.07 [0.06]
	(TL, Holo)	− 0.91	− 0.91 [0.10]	− 0.92 [0.11]	NA	− 0.91 [0.01]
I model: Individual model (I-MT2)	(V49, Holo)	0.17	0.36 [0.29]	NA	0.21 [0.31]	0.15 [0.10]
	(V49, TL)	− 0.07	− 0.15 [0.40]	NA	0.38 [0.48]	− 0.18 [0.10]
	(TL, Holo)	− 0.91	− 0.91 [0.09]	NA	− 0.91 [0.10]	− 0.92 [0.01]

SD: standard deviation

NA: this parameter cannot be estimated

V49 genetic gain when the economic coefficient changed from 0.75 to 0.25, whatever the model, which was an expected result too. The *I* model led to a greater genetic gain varying from 26 to 37% for the best situation. The higher gains for V49 were obtained using a coefficient equal to 0.75 for both (V49, TL) and (V49, Holo) indexes. Whatever the model, the genetic gain using the ST approach was the best or similar to the index with coefficient pair equal to 0.75/0.25 (Fig. 1).

For TL and Holo, the major difference with V49 was the gain in magnitude, which did not exceed 2% under the most favorable conditions (Fig. 2a and b). Except that, the same trends were observed for wood chemical traits, i.e., a decrease of genetic gain with percentage of trees selected (from 5 to 15%); however, these differences are smaller than 1% between the two extreme indices. As expected, the most important gain for TL and Holo was obtained with economic coefficient equal to 0.75.

Provenance representativeness in the selected population

We studied the evolution of diversity within the selected population by counting the number of provenances represented in the selected population. The comparisons were done according to the different selection strategies combining model-, single-, or multi-trait and selection intensity. Table 7 shows that, with the *P* model, one to three provenances were represented in the selected population with a selection intensity varying from 5 to 15%, whereas with the *I* model, the number of provenances varied between 6 and 16 depending on the indexes. The single-trait approach was close to the bivariate approach, as the number of provenances represented in the selected population was not very different and no clear trend was observed. Moreover,

Table 6 Prediction accuracy[a] (r_I) for all trait using single-trait (ST) and bivariate (MT2) of the *P* and *I* models

Trait	Model	Associated trait	$\mathbf{r_I}$ (*P* model)	$\mathbf{r_I}$ (*I* model)
V49	Single-trait (ST)	V49	0,29 [0,07]	0,30 [0,05]
	Bivariate (MT2)	(V49, Holo)	0,28 [0,06]	0,29 [0,05]
	Bivariate (MT2)	(V49, TL)	0,28 [0,06]	0,30 [0,05]
TL	Single-trait (ST)	TL	0,05 [0,06]	0,05 [0,06]
	Bivariate (MT2)	(TL, V49)	0,02 [0,07]	0,04 [0,07]
	Bivariate (MT2)	(TL, Holo)	0,03 [0,07]	0,05 [0,07]
Holo	Single-trait (ST)	Holo	0,11 [0,07]	0,09 [0,06]
	Bivariate (MT2)	(Holo, V49)	0,09 [0,08]	0,07 [0,07]
	Bivariate (MT2)	(Holo, TL)	0,10 [0,08]	0,10 [0,07]

[a] Accuracies resulted from 100 cross-validations and were based on a training population including two third of the total population and a validation population including one third of the total population

Table 7 Number of provenances represented, which corresponds to the genetic diversity, in each selected set with a selection intensity "i" varying from 5 to 15% according to the different models and selection indices

	Single trait (ST) selection				Multi-trait (MT) selection					
i	Model	V49	TL	Holo	MT2 (V49,TL)			MT2 (V49,Holo)		
					0.75 V49 + 0.25TL	0.50 V49 + 0.50TL	0.25 V49 + 0.75TL	0.75 V49 + 0.25Holo	0.50 V49 + 0.50Holo	0.25 V49 + 0.75Holo
5%	*P* model	1	1	1	1	1	1	1	1	1
	I model	8	6	9	8	7	7	10	8	7
10%	*P* model	2	2	2	2	2	2	2	2	2
	I model	12	9	12	11	11	11	12	11	10
15%	*P* model	3	3	3	3	3	3	3	3	3
	I model	12	13	16	13	15	13	13	14	13

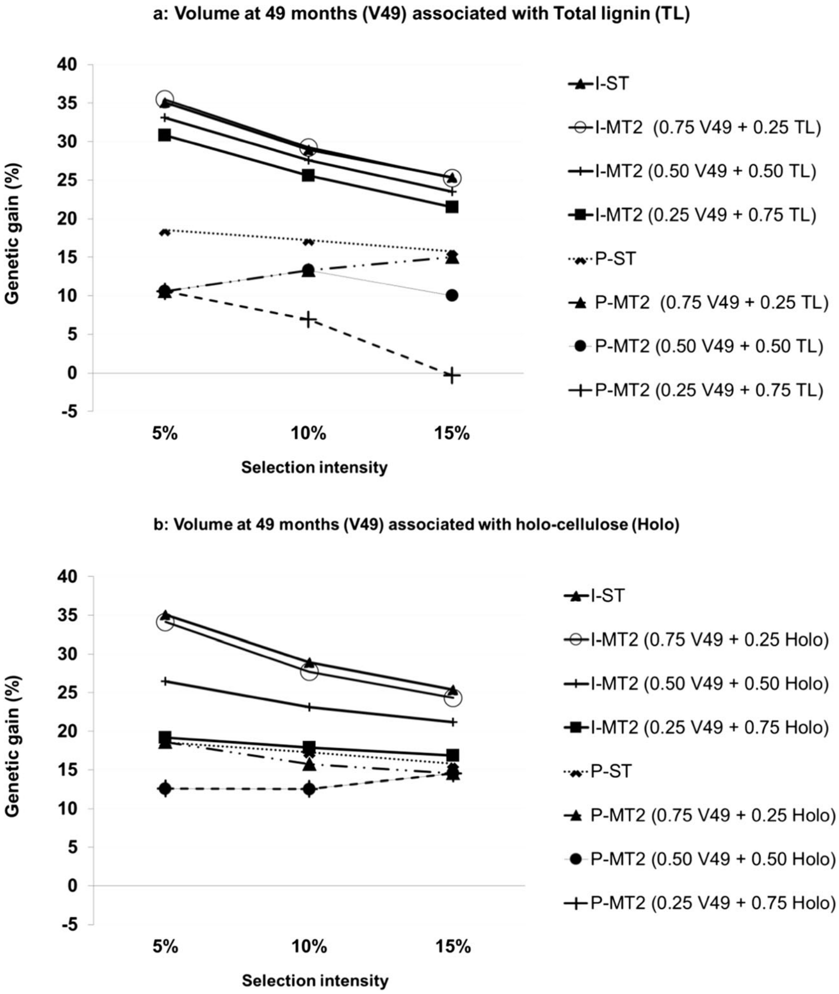


Fig. 1 Trends in genetic gain for volume at 49 months (V49) according to the *I* and *P* models with single-trait (I-ST and P-ST) and bivariate (I-MT2 and P-MT2) models with varying index selection coefficients: (0.75, 0.25), (0.50, 0.50), and (0.25, 0.75)

the representativeness of provenances was more or less the same using of the different indexes.

Discussion

Heritability and correlations between traits

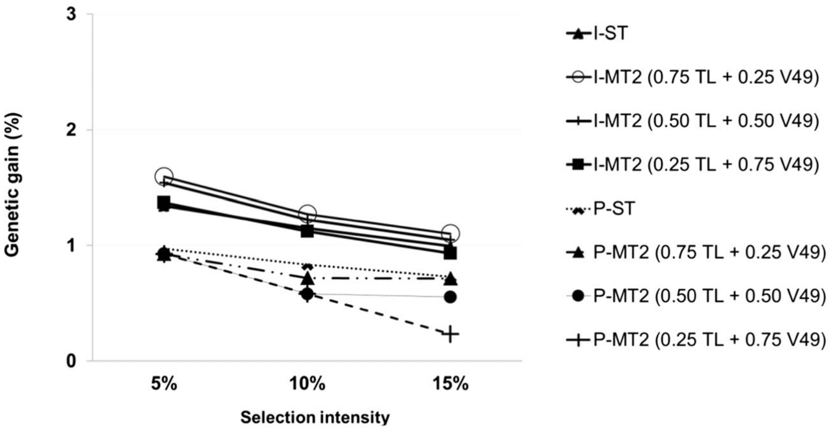
Although the provenance trial underwent successive thinnings, the genetic diversity was not markedly affected. The phenotypic coefficients of variation for V49 and wood traits were very close to those observed in other experiments on *Eucalyptus* (Tripiana et al. 2007; Denis et al. 2013; Razafimahatratra et al. 2016).

In our study, whatever the model, we found a lower heritability for TL and Holo than for V49 (Tables 3 and 4). Generally, the wood chemical properties appeared to be more heritable than growth traits. This is the case for lignin in *Eucalyptus* (Poke et al. 2006; Stackpole et al. 2011; Hein et al. 2012; Mandrou et al. 2012; Makouanzi et al. 2017) and for the syringyl-to-guaiacyl ratio (Hamilton et al. 2009; Poke et al. 2006; Hamilton et al. 2009; Stackpole et al. 2011; Hein et al. 2012; Makouanzi et al. 2017). In our experiment, this unexpected result can be explained by a preponderant residual variance for the wood properties. For example, in the *I* model, the additive coefficient of variation is three to four times smaller than the residual coefficient of variation for TL (CVa = 1.2% and CVr = 5.0%) and Holo (CVa = 1.4% and CVr = 4.0%), whereas for V49 this ratio is 1.5 (CVa = 16.2% and CVr = 24.6%). The preponderance of residual variance for wood chemical traits is probably related to the strong environmental impact that trees have undergone after 50 months. Strong winds affected the plantation; some of the trees surrounding the sampled trees were broken and produced shoot stumps, thus exacerbating the micro-environmental effects.

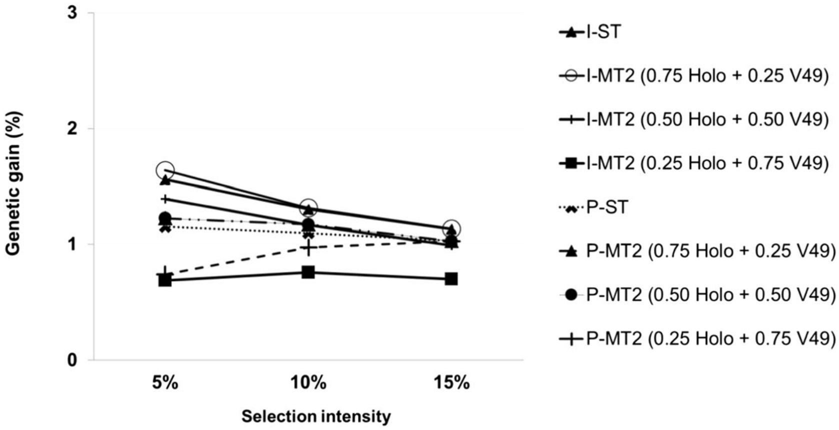
On average, whatever the models, the correlations between biomass and wood chemical properties showed a weak but positive estimate between V49 and Holo and a weak but negative estimate between V49 and TL. Although clear conclusions cannot be drawn given the large confidence intervals (Table 5), these results were consistent with other experiments with *Eucalyptus*, which tended to show a negative correlation between biomass and lignin and a positive correlation between growth and cellulose (Novaes et al. 2010; Denis et al. 2013;

Fig. 2 Trends in genetic gain for (**a**) total lignin (TL) and (**b**) holo-cellulose (Holo) according to the *I* and *P* models with single-trait (I-ST and P-ST) and bivariate (I-MT2 and P-MT2) models with varying index selection coefficients: (0.75, 0.25), (0.50, 0.50), and (0.25, 0.75)

a: Total lignin (TL) associated with volume at 49 months (V49)



b: Holo-cellulose (Holo) associated with volume at 49 months (V49)



Makouanzi et al. 2017). Note that these results have also been observed in other species, such as poplars (Novaes et al. 2010).

The performance of multi-trait genomic selection

One of the assumptions of our study was that the selection accuracy and consequently the genetic gain could be improved by the multi-trait approach. In the context of *Eucalyptus* breeding, we combined growth and wood chemical traits, both of which are of economic interest. With models that take into account only one or two traits, the algorithm converged easily (ST and MT2 approach). But convergence was difficult to achieve with three traits (MT3 approach). Convergence of the ASReml algorithm depends very much on the size and structure of the variance/covariance matrix to be estimated. This difficulty increases when the data are scarce, which is probably the case with our data set.

There was neither better goodness-of-fit with all the data (Tables 3 and 4) nor better prediction accuracy in the framework of GS (Table 6) by using the bivariate approach compared with a single-trait approach. Using simulations, Jia and Jannink (2012) showed that multi-trait GS increased accuracy, while low-heritability traits benefitted from correlated high-heritability traits when the genetic correlation was higher than 0.5. In our experiment, TL and Holo presented a lower heritability than V49, but the low correlations between V49 and both TL and Holo may explain the absence of impact on the accuracy of wood traits.

Consequence for genetic improvement of *Eucalyptus robusta*

Our study has shown that efficient individual selection can be achieved in a provenance trial using dense genotyping. We have adopted GBLUP, which is well adapted for implementing multi-trait selection rather than other prediction methods. Estimation of the **A** relationship matrix between individuals, by means of a consistent set of SNPs, makes it possible to estimate individual genetic value and to increase significantly the genetic gains compared with selection based on the provenance means.

Our analyses have produced various pieces of information to orient the constitution of the *E. robusta* breeding population. They have shown that biomass production, estimated through the volume of the tree, presented a higher variability than TL and Holo, leading to a higher genetic gain for growth than for wood chemical properties.

The association of volume with wood chemical property traits did not improve the selection accuracy compared with single-trait GS. This could be due to the weak correlations between growth and these wood chemical components. This result corroborates those found by Makouanzi et al. (2017) in *E. urophylla*×*E. grandis* hybrid, which suggested that selection for growth will not result in a significant reduction in wood quality traits.

Different selection strategies can impact genetic gain. For example, with equivalent coefficients for volume and wood chemical traits (0.50/0.50), index selection will still lead to a substantial gain in volume (around 30 and 25% associated with TL and Holo with selection intensity of 10% and *I* model), without affecting the gain for holo-cellulose and total lignin. With a higher coefficient for wood chemical traits (0.25/0.75), the gain in volume is reduced (around 20 and 15% associated with TL and Holo with selection intensity of 10% and *I* model), but the gain for wood traits changed only by 0.5%. This analysis showed that a strong emphasis on wood chemical traits is needed to achieve genetic gain in these traits during a breeding cycle.

Individual selection increased the genetic diversity in terms of the number of provenances represented in the selected population compared with selection based on provenance means. This result is encouraging if we want to build the breeding population while maximizing diversity. However, this conclusion should not obscure the fact that different provenances may have asynchronous flowering phenology, as has been observed in other species introduced in Madagascar, which may reduce panmixia within seed orchards and increase inbreeding (Chaix et al. 2003). It is therefore very important to incorporate flowering phenology as a selection trait for building the breeding population. On the basis of this latter observation, a two-stage selection may also be envisaged, first by selecting the best provenances on the basis of their growth and flowering phenology and then by selecting the individuals within these provenances using the indexes associating volume and wood properties.

Acknowledgements This study was conducted within the framework of the PhD fellowship funded by CIRAD for the research platform in the "Forest and Biodiversity in Madagascar" partnership. Field experiments were conducted at FOFIFA, the forest experimental station of Mahela in Madagascar in the framework of the *E. robusta* breeding program. Near infrared spectroscopy analyses were performed under the technical supervision of Gilles Chaix in the CIRAD laboratories in Montpellier, France. We thank our colleagues in Madagascar for their valuable help in sampling. We also thank the two reviewers for their valuable comments and suggestions to improve the first version of the manuscript.

Data archiving statement Data will be available from the Dryad Digital Repository if the manuscript is accepted. The data will consist in two files: a file with the phenotype of the 415 trees and a file with the 2919 SNP for each of the tree.

Author contributions Implementation and sampling: TR, JML, DV, and JMB; study and analysis design: TR, TC, JL, and JMB. TR and JMB supervised the writing and LR, TV, GM, and JML contributed ideas, comments, analyses, and revised manuscript versions.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Akaike H (1974) A new look at the statistical model identification. Trans Autom Control 19:716–723

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bouvet J-M, Makouanzi G, Cros D, Vigneron P (2015) Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. Heredity 116: 146–157. https://doi.org/10.1038/hdy.2015.78

Calus M, Veerkamp R (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43:26

Chaix G, Gerber SA, Razafimaharo V, Vigneron P, Verhaegen D, Hamon S (2003) Gene flow estimation with microsatellites in a Malagasy seed orchard of *Eucalyptus grandis*. Theor Appl Genet 107:705–712

Cros D, Denis M, Bouvet J-M, Sanchez L (2015) Long-term genomic selection for heterosis without dominance in multiplicative traits: case study of bunch production in oil palm. BMC Genomics 16(651):16. https://doi.org/10.1186/s12864-015-1866-9

Cros D, Denis M, Sanchez L, Cochard B, Flori A, Durand-Gasselin T, Nouy B, Omoré A, Pomies V, Riou V, Suryana E, Bouvet J-M (2015) Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet 128(3):397–410. https://doi.org/10.1007/s00122-014-2439-z

Denis M, Bouvet J-M (2011) Genomic selection in tree breeding: testing accuracy of prediction models including dominance effect. BMC Proc 5(Suppl 7):O13 http://www.biomedcentral.com/1753-6561/5/S7/O13

Denis M, Bouvet J-M (2013) Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. Tree Genet Genomes 9:37–51. https://doi.org/10.1007/s11295-012-0528-1

Denis M, Favreau B, Ueno S, Camus-Kulandaivelu L, Chaix G, Gion J-M, Nourrisier-Mountou S, Polidori J, Bouvet J-M (2013) Genetic variation of wood chemical traits and association with underlying genes in *Eucalyptus urophylla*. Tree Genet Genomes 9(4):927–942. https://doi.org/10.1007/s11295-013-0606-z

Donald CM (1968) The breeding of crop ideotypes. Euphytica 17:385–403

Gallais A, Poly J (1990) Théorie de la sélection en amélioration des plantes. Collection Sciences agronomiques, Masson, p 588

Gilmour AR, Gogel BJ, Cullis BR, and Thompson R (2009) ASReml User. Guide Release 3.0VSN International Ltd, Hemel Hempstead

Grattapaglia D (2014) Breeding forest trees by genomic selection: current progress and the way forward. In: Tuberosa R, Graner A, Frison E (eds) Genomics of Plant Genetic Resources. Springer, Netherlands, pp 651–682

Grattapaglia D, Resende MDV (2010) Genomic selection in forest tree breeding. Tree Genet Genomes 7:241–255

Guo G, Zhao F, Wang Y, Zhang Y, du L, Su G (2014) Comparison of single-trait and multiple-trait genomic prediction models. BMC Genet 15:30

Hamilton MG, Raymond CA, Harwood CE, Potts BM (2009) Genetic variation in *Eucalyptus nitens* pulpwood and wood shrinkage traits. Tree Genet Genomes 5:307–316

Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. Genome 53(11):876–883

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Sci 50: 1681–1690

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49(1):12

Hein PRG, Bouvet J-M, Mandrou E, Vigneron P, Clair B, Chaix G (2012) Age trends of microfibril angle inheritance and their genetic and environmental correlations with growth, density and chemical properties in *Eucalyptus urophylla* S.T. Blake wood. Ann For Sci 69(6): 681–691. https://doi.org/10.1007/s13595-012-0186-3

Higuchi T (1997) Biochemistry and molecular biology of wood. Springer-Verlag, Heidelberg

Hung TD, Brawner JT, Meder R, Lee DJ, Southerton S, Thinh HH, Dieters MJ (2015) Estimates of genetic parameters for growth and wood properties in *Eucalyptus pellita* F Muell. to support tree breeding in Vietnam. Ann For Sci 72:205–217

Isik F, Bartholomé J, Farjat A, Chancerel E, Raffin A, Sanchez L, Plomion C, Bouffier L (2016) Genomic selection in maritime pine. Plant Sci 242:108–119. https://doi.org/10.1016/j.plantsci.2015.08.006. http://prodinra.inra.fr/record/327160

Jacquin L, Cao TV, Ahmadi N (2016) A unified and comprehensible view of parametric and kernel methods for genomic prediction with application to rice. Front Genet 7:145. https://doi.org/10.3389/fgene.2016.00145

Jannink JL (2010) Dynamics of long-term genomic selection. Genet Sel Evol 42:35. https://doi.org/10.1186/1297-9686-42-35

Jia Y, Jannink J-L (2012) Multiple trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522. https://doi.org/10.1534/genetics.112.144246

Luan DT (2009) Genetic studies of Nile Tilapia (Oreochromis Niloticus) for farming in Northern Vietnam: growth, survival and cold tolerance in different farm environments. PhD theses. Norwegian University of Life Sciences

Makouanzi G, Chaix G, Nourissier S, Vigneron P (2017) Genetic variability of growth and wood chemical properties in a clonal population of *Eucalyptus urophylla* × *Eucalyptus grandis* in the Congo. South Forests 80:151–158. https://doi.org/10.2989/20702620.2017.1298015

Mandrou E, Hein PRG, Villar E, Vigneron P, Plomion C, Gion J-M (2012) A candidate gene for lignin composition in *Eucalyptus*: cinnamoyl-CoA reductase (CCR). Tree Genet Genomes 8:353–364. https://doi.org/10.1007/s11295-011-0446-7

Marchal A, Legarra A, Tisne S, Carasco-Lacombe C, Manez A, Suryana E, Omoré A, Nouy B, Durand-Gasselin T, Sanchez L, Bouvet JM, Cros D (2016) Multivariate genomic model improves analysis of oil palm (*Elaeis guineensis Jacq.*) progeny tests. Mol Breed 36(1):1–13

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829

Meuwissen TH, Mike Goddard ME (2010) Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. Genetics 185:623–631

Mrode RA, Thompson R (2005) Linear models for the prediction of animal breeding values. CABI, UK

Novaes E, Kirst M, Chiang V, Winter-Sederoff H, Sederoff R (2010) Lignin and biomass: a negative correlation for wood formation and lignin content in trees. Plant Physiol 154(2):555–561. https://doi.org/10.1104/pp.110.161281

Poke FS, Potts BM, Vaillancourt RE, Raymond CA (2006) Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus*. Ann For Sci 63:813–821. https://doi.org/10.1051/forest:2006080

Razafimahatratra AR, Ramananantoandro T, Razafimaharo V, Chaix G (2016) Provenance and progeny performances and genotype environment interactions of *Eucalyptus robusta* grown in Madagascar. Tree Genet Genomes 12:38. https://doi.org/10.1007/s11295-016-0999-6

Robbins, MD, Staub, JE, Fazio, G (2002) Deployment of molecular markers for multi-trait selection in cucumber. In: Proceeding Cucurbitaceae American Society for Horticultural Science, p 41–47

Stackpole DJ, Vaillancourt RE, Alves A, Rodrigues J, Potts BM (2011) Genetic variation in the chemical components of *Eucalyptus globulus* wood. G3 (Bethesda) 1:151–159. https://doi.org/10.1534/g3.111.000372

Tripiana V, Bourgeois M, Verhaegen D, Vigneron P, Bouvet J-M (2007) Combining microsatellites, growth, and adaptive traits for managing in situ genetic resources of *Eucalyptus urophylla*. Can J For Res 37(4):773–785

Van Raden PM (2007) Genomic measures of relationship and inbreeding genomic measures of relationship and inbreeding. Interbull Bull 37: 33–36

Van Raden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414–23

Verhaegen D, Randrianjafy H, Montagne P, Danthu P, Rabevohitra R, Tassin J, Bouvet JM (2011) Historique de l'introduction du genre *Eucalyptus* à Madagascar. Bois For Trop 309(3):17–25

Vigneron P, Bouvet J-M (1997) Les eucalyptus. In : L'amélioration des plantes tropicales. Charrier André (ed.), Jacquot Michel (ed.), Hamon Serge (ed.), Nicolas Dominique (ed.). CIRAD, Montpellier, pp 267–290

Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004) Diversity arrays technology (DArT) for whole-genome profiling of barley. PNAS 101(26):9915–9920. https://doi.org/10.1073/pnas.0401076101