Effect of Hidden Relatedness on Single-Step Genetic Evaluation in an Advanced Open-Pollinated Breeding Program

Jaroslav Klápště, Mari Suontama, Heidi S. Dungey, Emily J. Telfer, Natalie J. Graham, Charlie B. Low, and Grahame T. Stovold

From the Scion (New Zealand Forest Research Institute Ltd.), 49 Sala Street, Whakarewarewa, Rotorua 3010, New Zealand. Mari Suontama is now at Skogforsk, Umeå, Box 3, Sävar SE-918 21, Sweden.

Address correspondence to J. Klápště at the address above, or e-mail: jaroslav.klapste@scionresearch.com.

Received June 11, 2018; First decision August 20, 2018; Accepted September 27, 2018.

Corresponding Editor: John Stommel

Abstract

Open-pollinated (OP) mating is frequently used in forest tree breeding due to the relative temporal and financial efficiency of the approach. The trade-off is the lower precision of the estimated genetic parameters. Pedigree/sib-ship reconstruction has been proven as a tool to correct and complete pedigree information and to improve the precision of genetic parameter estimates. Our study analyzed an advanced generation Eucalyptus population from an OP breeding program using single-step genetic evaluation. The relationship matrix inferred from sib-ship reconstruction was used to rescale the marker-based relationship matrix ($G$ matrix). This was compared with a second scenario that used rescaling based on the documented pedigree. The proposed single-step model performed better with respect to both model fit and the theoretical accuracy of breeding values. We found that the prediction accuracy was superior when using the pedigree information only when compared with using a combination of the pedigree and genomic information. This pattern appeared to be mainly a result of accumulated unrecognized relatedness over several breeding cycles, resulting in breeding values being shrunk toward the population mean. Using biased, pedigree-based breeding values as the base with which to correlate predicted GEBVs, resulted in the underestimation of prediction accuracies. Using breeding values estimated on the basis of sib-ship reconstruction resulted in increased prediction accuracies of the genotyped individuals. Therefore, selection of the correct base for estimation of prediction accuracy is critical. The beneficial impact of sib-ship reconstruction using $G$ matrix rescaling was profound, especially in traits with inbreeding depression, such as stem diameter.

Keywords: Eucalyptus nitens, hidden relatedness, inbreeding depression, sib-ship reconstruction, single-step evaluation
(Burdon and Shelbourne 1971). OP strategies can not fully track pedigree and so do suffer from the presence of hidden relatedness, the proportion of which is affected by conditions under which reproduction was performed (i.e., wild stands vs. breeding arboretum vs. polymix breeding). Hidden relatedness can affect the accuracy of genetic parameter estimation and rankings of estimated breeding values (Squillace 1974; Askew and El-Kassabyy 1994; Namkoong et al. 1988; Vidal et al. 2013; Tambarussi et al. 2018). The development of highly polymorphic genetic markers, such as simple sequence repeats, has enabled pedigree reconstruction to be performed, eliminating the deleterious effect of hidden relatedness on accuracy of genetic parameters and breeding values in genetic evaluations (Lambeth et al. 2001; Grattapaglia et al. 2004; Doerksen and Herbinger 2010; Hansen and McKinney 2010; El-Kassabyy et al. 2011).

More recently, the development of next-generation sequencing technologies has facilitated the development of genomic resources, even for organisms with missing reference genomes such as forest trees (Elshire et al. 2011; Chen et al. 2013; Neves et al. 2013; Plomion et al. 2014; Silva-Junior et al. 2015). These technologies generate abundant genome-wide genetic markers, such as single nucleotide polymorphisms (SNPs), which allow the construction of a marker-based relationship matrix (Nejati-Javaremi et al. 1997; VanRaden 2008). Such matrices provide a tool to track Mendelian segregation (Visscher et al. 2006; Zapata-Valenzuela et al. 2013), historical relatedness before base population defined by pedigree (Powell et al. 2010) and linkage disequilibrium (LD) between markers and quantitative trait loci (QTLs) (Habier et al. 2013). In particular, tracking LD improves the ability to estimate genetic covariance and helps achieve the more accurate estimation of genetic variance (Lippert et al. 2013). The marker-based relationship matrix can then be used as a tool to predict phenotypes for individuals with genotypes through genomic selection (GS) models (Resende et al. 2012; Beaulieu et al. 2014; Muñoz et al. 2014; Gamal El-Dien et al. 2015; Ratcliffe et al. 2015; Bartholomé et al. 2016; Isik et al. 2016).

Forest tree species are a challenge as they are often characterized by high genetic diversity, large effective population size, and rapid LD decay, which requires genotyping of large training populations to fully utilize all the benefits of the genomics approach. The complete genotyping of a forest tree progeny test is currently cost-prohibitive due to their large dimensions (thousands of trees), and a reasonable alternative should be used (Beaulieu et al. 2014). El-Kassabyy and Lstibûrek (2009) proposed a partial pedigree reconstruction as an efficient alternative to full pedigree reconstruction to improve the comparative precision of genetic parameters (El-Kassabyy et al. 2011). Single-step evaluation (Legarra et al. 2009; Misztal et al. 2009) can be seen as a genomic-based equivalent of the above mentioned partial pedigree reconstruction to reasonably implement genomics into forest tree testing schemes. This strategy has already been successfully applied in animal breeding and also in some forest tree genetic evaluations (Christensen and Lund 2010; Meuwissen et al. 2011; Christensen et al. 2012; Cappa et al. 2017, 2018; Ratcliffe et al. 2017). The rescaling of the marker-based relationship matrix to that inferred from the documented pedigree is the greatest challenge in single-step genetic evaluation to avoid any inaccuracy of genetic parameter estimates. Usually, the marker-based matrix G is adjusted regarding differences of average diagonal and average off-diagonal elements to its pedigree-based counterpart. Nevertheless, the rescaling effects are highly variable and depend on the method used for G matrix construction (Forni et al. 2011). Several rescaling approaches have already been developed (Forni et al. 2011; Vitezica et al. 2011; Gao et al. 2012). However, there is lack of knowledge on the effect of incomplete pedigree information on accuracy of predicted breeding values in single-step evaluation. The rescaling of the G matrix based on incomplete pedigree-based relationship appears to be causing an issue. Individuals with shallow, single-generation pedigrees are causing the G matrix elements to be larger, on average, compared with the pedigree-based matrix A. In contrast, individuals with deep pedigrees have, on average, G matrix elements that are smaller (Misztal et al. 2013). The strategy to avoid this issue is through implementing patterns of population history. Misztal et al. (2013) developed a strategy based on implementation of unknown parental groups in a multibreed population. We found this strategy, however, unsuitable in our case due to the lack of isolation in mating events and rather we focused on reconstruction of hidden relatedness. A previous study performed on the material used in the current study was focused on sib-ship reconstruction and found a reasonable proportion of relatedness (including selfing), unrecognized by documented pedigree. The implementation of the relationship matrix based on sib-ship reconstruction improved the precision of genetic parameters and response to selection especially in traits suffering from inbreeding depression (Klápště et al. 2017). This study, therefore, investigates the efficiency of single-step genetic evaluation in an advanced generation of a Eucalyptus nitens breeding population, with an only partially tracked pedigree. It compares the effect of using relatedness inferred from sib-ship reconstruction versus the documented pedigree in the process of marker-based relationship matrix rescaling. In addition, the pedigree-based matrix was modified to take into account the probability of selfing in an attempt to further improve the accuracy of this strategy.

Methods

Material

The studied population is a third generation breeding population, derived from 2 seed orchards (Klápště et al. 2017). The experiment includes 3593 individuals structured into 116 half-sib families, of which 691 were randomly selected, representing 72 tested families analyzed through sib-ship reconstruction in previous study (Klápště et al. 2017). The individuals were measured for diameter at breast height (DBH) and scored for straightness (STR) using a 9° scale from 1—crooked to 9—straight and malformation (MAL) coded as a binary trait where 1 is perfectly formed and 0 otherwise.

Genetic markers were generated through EUChip60K SNP chip (Silva-Junior et al. 2015) and filtered for GenTrain score > 0.5, GenCall > 0.15, minor allele frequency (MAF) > 0.05 and SNP call rate > 0.6 which generated 13844 markers.

Statistical Analysis

Pedigree-Based Analysis

Genetic parameters such as additive genetic variance and heritability were estimated using a linear mixed model, implemented in the ASReml-R package (Butler et al. 2009) as follows:

\[ y = X \beta + Z a + Z r + Z r(s) + e \]

where y is the vector of observations, \( \beta \) is the vector of fixed effects such as intercept and seed orchard, a is the vector of random effects for breeding values following \( \text{var}(a) = N(0, \sigma^2_a) \), A is the average numerator relationship matrix (Wright 1922) which is substituted by the combined relationship matrix H using both
pedigree and marker information in the single-step evaluation (see below) and \( \sigma^2_g \) is the additive genetic variance, \( r \) is the vector of random replication effects following \( \text{var}(r) = N(0, I \sigma^2_r) \), where \( I \) is the identity matrix and \( \sigma^2_r \) is the replication variance, \( s(is) \) is the vector of random set nested within replication effects following \( \text{var}(s(is)) = N(0, I \sigma^2_is) \) (set represents incomplete block within replication having fixed number of families from each seed orchard), \( e \) is the vector of residuals following \( \text{var}(e) = N(0, I \sigma^2_e) \), where \( \sigma^2_e \) is the residual variance, \( X \) and \( Z \) are incidence matrices assigning fixed and random effects to observations in vector \( y \).

**Single-Step Genetic Evaluation**

Since the marker-based relationship matrix is reflecting both temporal and historical relatedness (Powell et al. 2010), the reference (base) population is different compared with the pedigree-based counterpart. Such discrepancies can result in biased estimations of genetic parameters and reduced accuracy of breeding values (Vitezica et al. 2011). Therefore, the adjustment of the marker-based relationship matrix is the most crucial step in the single-step evaluation. The marker-based relationship matrix \( G \) was constructed following (VanRaden 2008):

\[
G = \frac{ZZ'}{2 \sum p_i(1-p_i)}
\]

where \( Z = M - P, M \) is the matrix of genotypes coded 0, 1, and 2 as reference allele homozygote, heterozygote, and alternative allele homozygote, respectively, and \( P \) is the vector of doubled frequencies for alternative alleles, \( p_i \) is the frequency of the alternative allele at \( j \)th loci. The rescaling of the marker-based relationship matrix to adjust for a base population defined by the documented pedigree was performed following (Gao et al. 2012):

\[
\begin{align*}
\text{Avg.diag}(G)\beta + \alpha &= \text{Avg.diag}(A_{22}) \\
\text{Avg оф.diag}(G)\beta + \alpha &= \text{Avg оф.diag}(A_{22})
\end{align*}
\]

Since the investigated field experiment is derived from a 3rd generation breeding population in a program with incomplete tracking of relatedness, 2 \( A_{22} \) matrices were implemented to rescale the \( G \) matrix: 1) based on tracked pedigree (HBLUP1), and 2) based on sibship reconstruction performed in a previous study (Klášťová et al. 2017) (HBLUP2). We hypothesize that the implementation of a relationship matrix based on sibship reconstruction should result in a more precise adjustment of the marker-based relationship matrix to pedigree. The \( G \) matrix is usually not positive semi-definite, which is one of the mixed linear model assumptions, and weighting of the genomic and pedigree-based relationship matrices is required as follows:

\[
G_w = G(1-w) + A_{22}w
\]

Alternatively, the pedigree-based relationship matrix was modified to take into account partial selfing following (Dutkowski et al. 2001; Gilmour and Dutkowski 2004). This pedigree-based matrix was produced by using the “selfing” option in “asreml.Ainverse” function, implemented in the ASReml-R package (Butler et al. 2009).

The \( H \) matrix, implementing both marker and pedigree-based information, was constructed as follows:

\[
H = \begin{bmatrix}
A_{11} + A_{12}A_{22}^{-1}(G_w - A_{22})A_{21} & A_{12}A_{22}^{-1}G_w \\
G_wA_{22}^{-1}A_{21} & G_w
\end{bmatrix}
\]

where \( A_{11} \) is the relationship matrix for nongenotyped individuals, \( A_{12} \) and \( A_{22} \) are the relationship matrices between genotyped and nongenotyped individuals and \( A_{21} \) is the pedigree-based relationship matrix for genotyped individuals, \( G \) is the marker-based relationship matrix which is only available for genotyped individuals.

Narrow-sense heritability for continuous traits was estimated as follows:

\[
\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \phi \hat{\sigma}_e^2}
\]

and its alternative for binary trait was estimated as follows:

\[
\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \phi \hat{\sigma}_e^2}
\]

where \( \phi \) is the over/under dispersion coefficient. The theoretical accuracy of breeding values was estimated as follows:

\[
r = \sqrt{1 - \frac{\text{PEV}}{\text{var}(\hat{y})}}
\]

where PEV is prediction error variance (Mrode 2014), and \( F_i \) is the inbreeding coefficient of the \( i \)th individual. The leave-one-out cross-validation strategy was implemented as an independent evaluation of the tested models. Prediction accuracy for continuous traits was estimated as the correlation between breeding values estimated in the pedigree-based analysis and those predicted in the cross-validation procedure. Additionally, the predicted genomic breeding values for genotyped individuals were correlated with breeding values estimated in the independent analysis using the relationship matrix based on information from sibship reconstruction. Correlations were only estimated using the set of genotyped individuals. The area under a ROC curve (AUC) was used to estimate prediction accuracy for binary traits.

**Results**

The pedigree-based analysis resulted in heritability from 0.05 (MAL) to 0.28 (STR) for form traits and 0.22 (DBH) for the growth trait analyzed. The estimates for all traits were found to be statistically significant with regard to their standard errors (\( \alpha = 0.05 \)). The accuracy of the breeding values was moderate and reached 0.54 for the growth trait DBH and from 0.32 to 0.58 for form traits (MAL and STR) (Table 1). The LD in our population decayed to an \( r^2 \) of 0.2 within 3 kb, which is a common pattern in forest trees (Figure 1). The comparison of marker-based and sibship reconstruction-based relationship coefficients showed a clear deflation of marker-based estimates across the whole spectrum of relationship coefficients (Figure 2). The marker-based relationship matrix \( G \) was rescaled following Gao et al. (2012), using pedigree-based and sibship reconstruction-based relationship matrices. The parameters \( \alpha \) and \( \beta \) reached values of 0.005909189 and 1.322984116 in the pedigree-based scenario and 0.01343057 and 1.33787272 in the sibship reconstruction scenario.
The single-step evaluation resulted in heritability estimates ranging from 0.05 to 0.28 in the documented pedigree-based scenario and from 0.05 to 0.28 in the sib-ship reconstruction scenario. A slight increase in heritability from 0.22 to 0.24 was observed in the HBLUP1 scenario for DBH but was not accompanied by any concurrent increase in model fit. STR was the only trait to show improvement in the theoretical accuracy of breeding values when using information from sib-ship reconstruction to rescale the $G$ matrix compared with the pedigree-based scenario. The trend in the theoretical accuracy of the breeding values, however, is a reflection of the trend in heritability, which was not always a reflection of the model fit. The prediction accuracy was investigated through a leave-one-out strategy only in the default scenario (no selfing probability and 0.05 weight on pedigree information). Our study found

![Figure 1. LD decay in population under study.](image)

### Table 1. Variance components, heritability, their standard errors in parentheses, breeding values accuracy, their prediction accuracy (PA) in parentheses [2 prediction accuracies are reported for genotyped individuals regarding base to which are correlated (a) documented pedigree-based breeding value estimates; (b) sib-ship–based breeding value estimates—bold], and model fit for pedigree-based model (ABLUP), single-step evaluation where $G$ matrix is rescaled to documented pedigree (HBLUP1) and single-step evaluation where $G$ matrix is rescaled to information from sib-ship reconstruction (HBLUP2) under no selfing probability

| Model  | Parameter                | DBH       | STR       | MAL       |
|--------|--------------------------|-----------|-----------|-----------|
| ABLUP  | Additive genetic var.    | 132.4 (28.88) | 0.488 (0.096) | 0.218 (0.078) |
|        | Replicate var.           | 0.000 (0.000) | 0.074 (0.025) | 0.011 (0.013) |
|        | Rep(set) var.            | 0.000 (0.000) | 0.016 (0.011) | 0.000 (0.000) |
|        | Residual var.            | 480.5 (26.81) | 1.280 (0.052) | 1.000 (0.000) |
|        | Heritability             | 0.216 (0.045) | 0.278 (0.052) | 0.050 (0.017) |
|        | Acc (PA)—total          | 0.54 (0.69) | 0.58 (0.68) | 0.32 (0.56) |
|        | Acc (PA)—mother         | 0.54 (NA) | 0.56 (NA) | 0.35 (NA) |
|        | Acc (PA)—offspring      | 0.54 (0.69) | 0.58 (0.68) | 0.31 (0.56) |
|        | AIC                      | 25153.28 | 5290.2 | 8271.75 |
| HBLUP1 | Additive genetic var.    | 147.7 (31.47) | 0.484 (0.086) | 0.216 (0.077) |
|        | Replicate var.           | 0.000 (0.000) | 0.073 (0.025) | 0.011 (0.013) |
|        | Rep(set) var.            | 0.000 (0.000) | 0.017 (0.013) | 0.000 (0.000) |
|        | Residual var.            | 469.8 (27.75) | 1.277 (0.076) | 1.000 (0.000) |
|        | Heritability             | 0.239 (0.048) | 0.275 (0.047) | 0.050 (0.017) |
|        | Acc (PA)—total          | 0.57 (0.66) | 0.59 (0.64) | 0.34 (0.56) |
|        | Acc (PA)—mother         | 0.55 (NA) | 0.56 (NA) | 0.36 (NA) |
|        | Acc (PA)—offspring NonGen| 0.55 (0.66) | 0.58 (0.68) | 0.32 (0.56) |
|        | Acc (PA)—offspring Gen   | 0.63 (0.38, 0.37) | 0.65 (0.47, 0.58) | 0.40 (0.57) |
|        | AIC                      | 25148.78 | 5286.23 | 8272.943 |
| HBLUP2 | Additive genetic var.    | 131.9 (28.39) | 0.488 (0.088) | 0.231 (0.078) |
|        | Replicate var.           | 0.000 (0.000) | 0.074 (0.025) | 0.011 (0.013) |
|        | Rep(set) var.            | 0.000 (0.000) | 0.017 (0.011) | 0.000 (0.000) |
|        | Residual var.            | 480.2 (25.85) | 1.272 (0.077) | 1.000 (0.000) |
|        | Heritability             | 0.215 (0.044) | 0.277 (0.047) | 0.053 (0.017) |
|        | Acc (PA)—total          | 0.55 (0.67) | 0.59 (0.64) | 0.34 (0.56) |
|        | Acc (PA)—mother         | 0.54 (NA) | 0.56 (NA) | 0.36 (NA) |
|        | Acc (PA)—offspring NonGen| 0.53 (0.66) | 0.58 (0.68) | 0.32 (0.56) |
|        | Acc (PA)—offspring Gen   | 0.62 (0.58, 0.42) | 0.66 (0.47, 0.58) | 0.39 (0.57) |
|        | AIC                      | 25137.04 | 5283.96 | 8275.39 |
Discussion

Controlled pollination in forest tree breeding is expensive, time-consuming, and labor-intensive and its efficiency is affected by both biological and environmental limitations. Therefore, open pollination has been preferred in forest tree breeding programs, such as in the case of the *E. nitens* program in New Zealand (Burdon and Shelbourne 1971). However, this strategy comes at the cost of incomplete knowledge of genealogy, likely to cause the estimation of genetic parameters in quantitative genetic evaluations that are less reliable (Ratcliffe et al. 2017). The development of genetic markers has allowed recovery of missing relatedness and genealogy through pedigree/sib-ship reconstruction (Askew and El-Kassaby 1994; Lambeth et al. 2001; Vidal et al. 2015). Dense marker arrays have also allowed the construction of realized relationship matrices (Nejati-Javaremi et al. 1997; VanRaden 2008), which usually increase the accuracy of genetic parameter estimates and allow for more efficient selection of superior genotypes (Resende et al. 2012; Gamal El-Dien et al. 2015; Ratcliffe et al. 2015, 2017; Suontama et al. 2018).

El-Kassaby and Lštibýrek (2009) and El-Kassaby et al. (2011) found partial pedigree reconstruction as a feasible and cost-effective alternative to full pedigree reconstruction to improve the precision of genetic parameters. Our previous study (Klápště et al. 2017) focused on the effect of sib-ship reconstruction to improve genetic parameters. A significant benefit was demonstrated for those traits suffering from inbreeding depression, achieved by recognizing selves in the population, consequently leading to an increase in additive genetic variance, heritability, and improvement in estimated genetic gain. Improvement in breeding value accuracy was also observed for traits free of inbreeding depression due to the recovery of hidden relatedness and potential correction of pedigree errors. The analysis found 630 pair-wise relationships originally defined as half-sibs to be unrelated (See figure 2 in Klápště et al. 2017). However, defining all pedigree errors was not possible due to an inability to assign parents to each individual in the sib-ship reconstruction strategy. Similarly, the current study found a benefit when rescaling the marker-based G matrix according to the relationship matrix based on information from sib-ship reconstruction, rather than the documented pedigree. The benefit seen in the improved model fit (Table 1) and breeding values accuracy (Table 1—bold numbers) was more evident in production trait (DBH), which was more likely to suffer from inbreeding depression (Hardner and Tibbits 1998). Therefore, parentage/sib-ship reconstruction should be performed before G matrix rescaling in single-step evaluations, when applied in OP breeding programs. However, the low correlations between breeding values estimated on the basis of sib-ship reconstruction with those predicted in single-step evaluation is a result of the high influence of unrecognized relatedness and pedigree errors from nongenotyped individuals (contributing by 80% of the total population size) on breeding values predicted from single-step evaluation. On the other hand, there was no improvement in the accuracy of breeding values estimated in the nongenotyped part of the population after implementation of genomic information. This can be again caused by a high level of uncertainty in relatedness (coming from both the hidden relatedness and pedigree errors) across the population. In this case, we recommend the pedigree/sib-ship reconstruction of the whole population to reach a higher accuracy of predicted breeding values.

Results presented in this study showed that accumulation of unrecognized relatedness and pedigree errors across several
generations of breeding cycles resulted in virtually nonexistent between-family variation, with the main source of genetic variation generated by within-family variation (Figure 3). In contrary, the analysis using traits having similar level of heritability but complete pedigree information found large proportion of the genetic variance attributed to between-family variance (Thistlethwaite et al. 2017). Therefore, the missing pedigree information on the paternal side of the current progeny population, as well as for the parents in previous generations, appears to undermine the ability of the REML algorithm to differentiate families, and breeding values are shrunk toward the population mean (Figure 3) (Henderson 1975; Garrick et al. 2009). On the other hand, using genomic markers allowed the recovery of hidden relatedness and pedigree errors, resulting in a more disperse distribution of genomic breeding values compared with their pedigree-based equivalents (Figure 4). There are several strategies developed in animal breeding to overcome uncertain paternity using phenotypic data (Sapp et al. 2007) or construction of a sire probability matrix (Henderson 1988). However, the probability for many possible males (as would be the most likely scenario in forest trees) assigned to each nongenotyped offspring is not sufficient to increase the accuracy of genetic parameter estimates (Konigsberg and Cheverud 1992). The purpose of GS is primarily the approximation of pedigree-based breeding values through the implementation of genetic markers (Meuwissen et al. 2001). When the pedigree-based estimates of breeding values are imprecisely estimated, however, the resulting prediction accuracy (in terms of correlation between pedigree-based estimated and marker-based predicted breeding values) will undermine the efficiency of genomic predictions. Under such conditions, we would highly recommend implementation of genetic markers across the whole population and perform either pedigree or sib-ship reconstruction to obtain relatedness structure approaching the reality. The breeding values estimated on the basis of pedigree/sib-ship reconstruction will reach higher accuracy and provide a better base for the estimation of prediction accuracy.

The construction of a relationship matrix based on information from genetic markers allows tracking of not only temporal relatedness, as defined by the pedigree-based base population, but also Mendelian sampling (Visscher et al. 2006; Zapata-Valenzuela et al. 2013) and historical relatedness (Powell et al. 2010). This is highly beneficial in species in the initial phase of domestication, where pedigrees are shallow and simple, such as forest trees. Additional information from all genotyped individuals increases the precision of breeding values considerably (Table 1). Ratcliffe et al. (2017) investigated the effect of genotyping intensity in a single-step evaluation in white spruce and found continuous improvement in the accuracy of genetic parameters and model fit with increasing genotyping intensity. The study demonstrates the high value of genomic information, implemented in the initial phase of breeding programs, where pedigrees are simple and incomplete. Similarly, we found a large increase in the theoretical accuracy of breeding values for genotyped individuals compared with those without genotypes showing no improvement (Table 1). However, the prediction accuracy of genotyped individuals increased only when sib-ship reconstruction-based breeding values were used as a base. The fact that nongenotyped individuals reached higher prediction accuracy than genotyped individuals can be explained by the highly biased estimates of family means targeted in pedigree-based predictions (Zapata-Valenzuela et al. 2013). On the other hand, within-family variation targeted by genomic-based prediction is largely unreliable due to accumulated unrecognized relatedness and pedigree errors.
across the breeding cycles. Therefore, genomic approaches remain a very attractive option in forest species even where shortening of the breeding cycle is not possible due to late flowering. Gains can be made instead through a more complete understanding of underlying relationships and more accurate estimation of genetic parameters.

The precision of marker-based estimates of genetic parameters remain sensitive to selection, and a combination of marker and pedigree information is still recommended (Ducrocq and Patry 2010; Vitezica et al. 2011). In addition, the definition of the reference population in marker-based relationship matrices is rather arbitrary (Speed and Balding 2015) and should be rescaled with respect to the population in marker-based relationship matrices is rather arbitrary (Speed and Balding 2015) and should be rescaled with respect to the pedigree. de los Campos et al. (2015) argued that total heritability can be recovered only when all QTLs are included in the marker array, and is partially lost due to imperfect LD when only SNPs surrounding QTLs are available. Lippert et al. (2013) investigated the effect of using both QTL and non-QTL markers to construct a marker-based relationship matrix and found that only using QTL markers provides the most accurate estimate of additive genetic variance and heritability. Our analysis was performed using a multispecies Eucalyptus SNP chip (Silva-Junior et al. 2015), and ~10k markers were informative in this E. nitens population. The decay in LD was fast, as is common in forest tree species, and disappeared within ~3 kb (Figure 1). Therefore, capturing markers linked to QTLs is rather unlikely, and relatedness with co-segregation is probably the major source of capturing QTL effects. On the other hand, using an overwhelming amount of the genomic data does not increase the accuracy of prediction model after reaching saturation (Habier et al. 2013) and tract specific SNP prioritization should be applied (Lippert et al. 2013). However, the genetic complexity of the investigated traits can prohibit the reliable selection of causal variants, and therefore, prediction models rely rather on relatedness and co-segregation (Habier et al. 2007, 2013). Our previous analysis (Suontama et al. 2018) found that the sample coming from Tinkers (seed orchard undergoing more intensive selection) had a higher GEBV accuracy compared with the sample coming from the Waiouru seed orchard (seed orchard established as a clonal archive having broader genetic diversity), which was reached thanks to slower LD decay, capturing longer effective chromosomal fragments. Due to the fact that sample from Waiouru seed orchard had twice the sample size and produced lower accuracy of GEBVs, we found that the model didn’t reach a saturation point where any additional markers wouldn’t increase accuracy of GEBVs. Therefore, there appeared to still be space for improvement of genomic resources in Eucalyptus species to create a robust genomic prediction model.

The recovered relatedness through genetic markers in the set of genotyped individuals was underestimated compared with expectations (Figure 2). This can be caused by several cycles of selection and a lack of unrelated individuals to provide a reference for inferring actual relatedness among related individuals (Speed and Balding 2015), and had to be rescaled with respect to the pedigree-based counterpart before blending with the pedigree-based relationship matrix. The rescaling of the marker-based relationship matrix with respect to the pedigree-based equivalent is the most crucial step in single-step genetic evaluation. The difference in the scale of relationship coefficients between marker-based and pedigree-based counterparts causes a decrease in the accuracy of genomic breeding values (Ducrocq and Patry 2010; Forni et al. 2011; Vitezica et al. 2011). We tested 2 scenarios: 1) rescaling of the marker-based relationship matrix with regard to the documented pedigree and 2) rescaling the marker-based relationship matrix with regard to the relationship matrix derived from sib-ship reconstruction. The implementation of information from sib-ship reconstruction in the G matrix rescaling process resulted in a considerable improvement in model fit compared with the model that used the documented pedigree. This trend is especially observed in traits suffering from inbreeding depression, such as DBH (Hardner and Tibbits 1998). These improvements were achieved in spite of the fact that the sib-ship reconstruction could only recover higher classes of relatedness, such as full-sibs and half-sibs, but not first and second order cousins as found in the documented pedigree. This means that the greater degree of relatedness recovered by sib-ship reconstruction has a more significant impact on the improvement of genetic parameter estimates through the G matrix rescaling process than ignored or undiscovered lower degrees of relatedness. Therefore, pedigree/sib-ship reconstruction is highly recommended prior to G matrix rescaling in the single-step genetic evaluation, especially in species with an OP breeding program, where selfs are viable. However, we could not utilize the full potential of relatedness recovered by sib-ship reconstruction due to loss of connectivity with the remainder of the pedigree, as a simple blending of the sib-ship reconstruction-based relationship matrix (sib-ship-based \( A_s \)) into the pedigree-based relationship matrix would cause the resulting matrix not to be positive definite. Therefore, newly obtained relatedness information should be used only in the rescaling, but not in the weighting step. A more useful strategy would be to perform parentage analysis instead of sib-ship reconstruction when genomic information is also available for parental populations. In this case, consistency between original pedigree and the
reconstructed part would remain and positive definite nature of resulting relationship matrix warranted.

In some cases, marker information is not sufficient to capture all additive genetic variance, and residual polygenic effects have to be included in the prediction model (Aguilar et al. 2010; Christensen and Lund 2010). In addition, the implementation of a residual polygenic effect reduces the bias in SNP effects and increases their transferability over generations (Solberg et al. 2009). Similarly, in the single-step genetic evaluation, the weighting of marker and pedigree information is applied. We tested a broad range of weights from 0.05 to 0.5 for the pedigree information, but any increase resulted in a decrease in breeding value accuracies for genotyped individuals, while no effect was observed in nongenotyped individuals (Supplementary File 1).

Our previous analysis identified ~4% selfing in the genotyped sample (Klášť et al. 2017) and, therefore, we modified the pedigree-based matrix for selfing probability as proposed by Dutkowski et al. (2001) before blending with the marker-based G matrix. The modified selfing probability did not result in any additional improvement in the accuracy of breeding values, with decreases observed once probability exceeded 3% (Supplementary File 1). These results confirm our finding of 4% selfing in previous sib-ship reconstruction analysis (Klášť et al. 2017), it is, therefore, beneficial to implement selfing probability in any single-step genetic evaluation in species where there is strong evidence of viable selfing.

In this study, we have shown how the increase in connectivity between genotyped individuals through genomic similarity has a big impact on the resulting accuracy of breeding values compared with information from a sparse pedigree. In addition, implementation of genomic information in a quantitative genetic evaluation can dissect genetic and environment effects more precisely (Gamal El-Dien et al. 2016). Modification of the relationship matrix for selfing before blending and/or rescaling was found to be important in our population and would be recommended for other OP tree breeding programs.

**Supplementary Material**
Supplementary data are available at *Journal of Heredity* online.

**Funding**
This work was supported by the Specialty Wood Partnership Program (SWP) contract number C04X1104 and Scion Strategic Investment Funding.

**Conflict of Interest**
The authors declare no conflict of interest.

**Data Availability**
All genomic and phenotypic data is publically available from DRYAD data depository, doi:10.5061/dryad.cb4m96b.

**References**
Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci.* 93:743–752.

Askew GR, El-Kassaby YA. 1994. Estimation of relationship coefficients among progeny derived from wind-pollinated orchard seeds. *Theor Appl Genet.* 88:267–272.

Bartholomé J, Van Heerwaarden J, Isik F, Boury C, Vidal M, Plomson C, Bouffier L. 2016. Performance of genomic prediction within and across generations in maritime pine. *BMC Genomics.* 17:604.

Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J. 2014. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity (Edinb).* 113:343–352.

Burdon R, Shelbourne C. 1971. Breeding populations for recurrent selection: conflicts and possible solutions. *N Z J For Sci.* 1:174–193.

Butler D, Cullis BR, Gilmour A, Gogel B. 2009. *Asreml-reference manual.* Brisbane: The State Queensland, Department of Primary Industries and Fishery.

Cappa EP, El-Kassaby YA, Muñoz F, García MN, Villalba PV, Klášť J, Marcucci Poltri SN. 2017. Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models. *Mol Breed.* 37:125.

Cappa EP, El-Kassaby YA, Muñoz F, García MN, Villalba PV, Klášť J, Marcucci Poltri SN. 2018. Genomic-based multiple-trait evaluation in *Eucalyptus grandis* using dominant DAfT markers. *Plant Sci.* 271:27–33.

Chen C, Mitchell SE, Elshire RJ, Buckler ES, El-Kassaby YA. 2013. Mining conifers’ mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genet Genomes.* 9:1537–1544.

Christensen OF, Lund MS. 2010. Genomic prediction when some animals are not genotyped. *Genet Sel Evol.* 42:1.

Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. 2012. Single-step methods for genomic evaluation in pigs. *Animal.* 6:1565–1571.

de Los Campos G, Sorensen D, Gianola D. 2015. Genomic heritability: what is it? *Plos Genet.* 11:e1005048.

Doerksen TK, Herbering CM. 2010. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. *Tree Genet Genomes.* 6:591–600.

Ducrocq V, Patry C. 2010. Combining genomic and classical information in national BLUP evaluation to reduce bias due to genomic pre-selection. *Interbull Bull.* 41:33–36.

Dutkowski G, Gilmour A, Borralho N. 2001. Modification of the additive relationship matrix for open pollinated trial. In: S Barros, R Ipinzà, editors. Developing the eucalypt of the future. Proceedings of IUFRO Working Group 2.08.03 Conference, 10-15 September, Valdivia, Chile.

El-Kassaby YA, Cappa EP, Liewalskaneyanawin C, Klášť J, Lštibrek M. 2011. Breeding without breeding: is a complete pedigree necessary for efficient breeding? *PLoS One.* 6:e25737.

El-Kassaby YA, Lštibrek M. 2009. Breeding without breeding. *Genet Res (Camb).* 91:111–120.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 6:e19379.

Forni S, Aguilar I, Misztal I. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol.* 43:1.

Gamal El-Dien O, Ratcliffe B, Klášť J, Chen C, Porth I, El-Kassaby YA. 2015. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genomics.* 16:370.

Gamal El-Dien O, Ratcliffe B, Klášť J, Porth I, Chen C, El-Kassaby YA. 2016. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. *G3 (Bethesda).* 6:743–753.

Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G. 2012. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet Sel Evol.* 44:8.

Garrick DJ, Taylor JE, Fernando RL. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol.* 41:55.
