BMC Genomics

RESEARCH ARTICLE **Open Access**

Genomic selection for non-key traits in radiata pine when the documented pedigree is corrected using DNA marker information

Yongjun Li[1,2*], Jaroslav Klápště[1], Emily Telfer[1], Phillip Wilcox[3], Natalie Graham[1], Lucy Macdonald[1^] and Heidi S. Dungey[1]

Abstract

Background: Non-key traits (NKTs) in radiata pine (*Pinus radiata* D. Don) refer to traits other than growth, wood density and stiffness, but still of interest to breeders. Branch-cluster frequency, stem straightness, external resin bleeding and internal checking are examples of such traits and are targeted for improvement in radiata pine research programmes. Genomic selection can be conducted before the performance of selection candidates is available so that generation intervals can be reduced. Radiata pine is a species with a long generation interval, which if reduced could significantly increase genetic gain per unit of time. The aim of this study was to evaluate the accuracy and predictive ability of genomic selection and its efficiency over traditional forward selection in radiata pine for the following NKTs: branch-cluster frequency, stem straightness, internal checking, and external resin bleeding.

Results: Nine hundred and eighty-eight individuals were genotyped using exome capture genotyping by sequencing (GBS) and 67,168 single nucleotide polymorphisms (SNPs) used to develop genomic estimated breeding values (GEBVs) with genomic best linear unbiased prediction (GBLUP). The documented pedigree was corrected using a subset of 704 SNPs. The percentage of trio parentage confirmed was about 49% and about 50% of parents were re-assigned. The accuracy of GEBVs was 0.55–0.75 when using the documented pedigree and 0.61–0.80 when using the SNP-corrected pedigree. A higher percentage of additive genetic variance was explained and a higher predictive ability was observed when using the SNP-corrected pedigree than using the documented pedigree. With the documented pedigree, genomic selection was similar to traditional forward selection when assuming a generation interval of 17 years, but worse than traditional forward selection when assuming a generation interval of 14 years. After the pedigree was corrected, genomic selection led to 37–115% and 13–77% additional genetic gain over traditional forward selection when generation intervals of 17 years and 14 years were assumed, respectively.

Conclusion: It was concluded that genomic selection with a pedigree corrected by SNP information was an efficient way of improving non-key traits in radiata pine breeding.

Keywords: Genomic selection, Non-key traits, Radiata pine, Pedigree correction, Accuracy, Predictive ability

* Correspondence: Yongjun.Li@agriculture.vic.gov.au
[1]Scion (New Zealand Forest Research Institute), Private Bag 3020, Rotorua 3046, New Zealand
[2]Agriclture Victoria, AgriBio Centre, 5 Ring Road, Bundoora, VIC 3083, Australia
Full list of author information is available at the end of the article

Introduction

Genomic selection (GS) is an approach for improving quantitative traits in forest tree breeding populations that uses high density markers dispersed across the whole genome [1–7]. Genomic predictions are estimated based on information from markers, phenotypes and pedigrees to increase the accuracy of breeding values. There are two groups of individuals that are used in genomic selection: the training individuals and the selection candidates. Marker and pedigree information is available for both groups of individuals, but phenotypes are only available for the training individuals. The breeding values of selection candidates can be estimated without the need to ascertain their own individual phenotypes. In traditional tree breeding, selection candidates must be tested in field trials over a number of years to obtain their performance measurements. With genomic selection, breeding cycles can skip the field performance testing phase thereby significantly reducing the generation interval. This benefit of genomic selection is particularly important to species with long generation intervals and requiring large field testing experiments such as forest trees [1, 5, 8] and is particularly useful for those traits that express late in life (e.g. wood density) or have low to medium heritability (e.g. growth and disease resistance) [9, 10]. Until recently, obtaining sufficient single nucleotide polymorphisms (SNPs) to cover the entire genome and hence capture enough genomic variation was prohibitively expensive. The development of next-generation sequencing techniques has enabled researchers to obtain tens of thousands of SNPs at a reasonable cost through genotyping-by-sequencing (GBS) [11]. GBS uses methylation-sensitive restriction enzymes to reduce genome complexity and avoid the repetitive fraction of the genomes. It is becoming increasingly important to acquire genomic information in plant species with complex genomes that lack reference genomes. Where expressed sequence data are available, exome-capture GBS offers an alternative that allows researchers to focus on gene regions, generating a smaller, more manageable dataset and a cost-effective sequencing solution for studying genomes in species with large genomes, such as loblolly pine (*Pinus taeda*) [12].

SNPs have been found to be associated with phenotypic performance [13–15]. Traits under selection can follow specific genetic architectures so several models assuming different distributions of marker effects should be investigated. There are essentially two types of genetic architecture: (1) genetic effects follow a mixed inheritance process where there are few genetic variants of large effects and many variants of very small effects, or (2) genetic effects follow Fisher's infinitesimal model and each effect contributes only a very small fraction of the total genetic variance. Variable selection procedures such as Bayesian methodologies (including BayesB, BayesC, BayesCπ, etc.) are successively used in traits with the first type of genetic architecture,

where the marker effects are modeled to follow a priori distributions [16–18]. These Bayesian methodologies for genomic selection are implemented in two steps: 1) breeding values are estimated using phenotypes and pedigree information, and 2) prediction equations using SNP markers are estimated using de-regressed estimated breeding values (EBVs) as inputs, and then used to derive genomic EBVs (GEBVs) [19–21]. The genomic breeding values of selection candidates are calculated based on the prediction equations and their marker genotypes. However, this two-step procedure has been found to inflate the accuracy of genetic evaluation when individuals with only small numbers of offspring were used [22]. The second type of genetic architecture can be successively fitted by genomic best linear unbiased prediction (GBLUP) which estimates genomic breeding values by incorporating genomic relationships derived from markers in a mixed model framework. No prediction equations are estimated for individual markers. The GBLUP method is preferred for forest tree breeding programmes since only shallow and simple pedigrees are usually available, so reliable de-regression of EBVs cannot be undertaken [23]. Moreover, experimental design features can be included in the model. The genotype by environment interaction can also be formulated and variance-covariance structures incorporated into GBLUP models to account for genetic/residual heterogeneity [3].

Growth, wood density and stiffness are the most economically important traits for radiata pine (*Pinus radiata* D. Don) growers and the improvement of these traits has been the main focus of radiata pine breeding programmes. They are called the key traits (KTs) in radiata pine breeding, while other traits of interest to radiata pine breeders are called non-key traits (NKTs) [24–26]. Branch-cluster frequency, stem straightness, external resin bleeding, and internal checking are examples of such traits. These non-key traits have been targeted for improvement in previous radiata pine research programmes [27, 28]. Selection indices have been proposed to incorporate non-key traits together with the key traits into breeding programmes in New Zealand [25, 26].

New Zealand's Radiata Pine Breeding Company (RPBC) has established a genomic selection project as part of its overall goal of genetically improving the growth, form, wood quality, and resistance to pests and diseases of radiata pine. Phenotypes for two form traits (branch-cluster and stem straightness) and two wood quality traits (internal checking and external resin bleeding) were available for the training population of this genomic selection project. Branch-cluster frequency refers to the frequency of branch-clusters between one and six metres above the ground on the main stem. It affects both branch size and mean internode length, particularly in the first 3–11 m of the tree bole above the ground. Stem straightness affects log grade, log length and sawn-timber recovery [25, 29].

Li *et al. BMC Genomics*   (2019) 20:1026

Page 3 of 10

External resin bleeding and internal checking are two wood defects in radiata pine timber and lower the value of appearance-grade timber, leading to large economic losses for the forest industry [28]. Stem straightness and branch-cluster frequency both have medium to high heritabilities [30] while external resin bleeding and internal checking have low to high heritabilities [26, 31, 32].

This study was the application of genomic selection in radiata pine breeding with a limited number of genotypes in the training population. The objective of this study was to demonstrate the efficacy of applying genomic selection in radiata pine breeding for the non-key traits described. The accuracy of genomic breeding values, the predictive ability of genomic selection, and the expected genetic gains for these non-key traits in radiata pine were investigated in this study.

Results

Pedigree correction

The training population in this study comprised two clonally propagated radiata pine breeding trial series planted in New Zealand: POP2 and POP3. Trio parentage assignments for POP2 and POP3 was conducted with Cervus [33, 34]. The percentage of trio parentage confirmed was 48.91 and 49.33% for POP2 and POP3, respectively. There were 83 parents in total in the documented pedigree of POP2 and POP3. About 50% of parents were re-assigned in the SNP-corrected pedigree. The total number of parents in the SNP-corrected pedigree was 107.

Heritability and accuracy of breeding values

The heritability estimates (h_a^2) in ABLUP (best linear unbiased prediction using the average numerator relationship matrix) and the combined heritability estimates (h_{am}^2) in GBLUP were lower when the SNP-corrected pedigree was used compared with the documented pedigree for branch-cluster frequency and stem straightness (Table 1). However, the heritability estimates were similar when comparing the SNP-corrected pedigree and the documented pedigree for internal checking and external resin bleeding. In GBLUP, the marker-based heritability (h_m^2) was higher when using the SNP-corrected pedigree than using the documented pedigree. The combined heritability estimated in GBLUP was higher than the heritability estimated in ABLUP for branch-cluster frequency, stem straightness and internal checking, whereas heritability estimates for external resin bleeding were similar in GBLUP and ABLUP.

The accuracy of GEBVs was lower (0.55–0.75) for branch-cluster frequency, stem straightness, internal checking and external resin bleeding) than EBVs from ABLUP (0.73–0.84) when using the documented pedigree. The accuracy of GEBVs was higher for branch-cluster frequency (0.80) than that of EBVs from ABLUP (0.73) while that of GEBVs for stem straightness, internal checking and external resin bleeding (0.61–0.70) was lower than that of EBVs from ABLUP (0.73–0.87) when using the SNP-corrected pedigree. Higher accuracy was observed for external resin bleeding in ABLUP when using the documented pedigree than using the SNP-corrected pedigree. Similar accuracy was observed in ABLUP when using the documented pedigree and SNP-corrected

Table 1 Heritabilities, accuracy of EBVs and GEBVs, and the percentage of genetic variation explained by SNP markers (%VA) for branch-cluster frequency, stem straightness, internal checking and external resin bleeding when using documented or SNP-corrected pedigrees

Statistical model	Pedigree	Genetic parameter	Branch-cluster frequency	Stem straightness	Internal checking	External resin bleeding
ABLUP	Documented	h_a^2	0.17	0.13	0.23	0.33
		r_{EBV}	0.77	0.73	0.78	0.84
	SNP-corrected	h_a^2	0.09	0.10	0.23	0.34
		r_{EBV}	0.73	0.73	0.77	0.87
GBLUP	Documented	h_m^2	0.18	0.09	0.11	0.12
		h_{am}^2	0.28	0.18	0.29	0.35
		r_{GEBV}	0.75	0.68	0.55	0.56
		%VA	64%	54%	39%	36%
	SNP-corrected	h_m^2	0.21	0.1	0.16	0.18
		h_{am}^2	0.22	0.13	0.28	0.34
		r_{GEBV}	0.80	0.70	0.65	0.61
		%VA	96%	74%	59%	46%

h_a^2: heritability from ABLUP, h_m^2: marker-based heritability from GBLUP, h_{am}^2: the combined heritability based on variance explained by SNP markers and residual additive genetic variance from GBLUP. Heritabilities and residual variances reported here are the average across seven sites for branch-cluster frequency and stem straightness, and across four sites for internal checking and external resin bleeding

Li *et al. BMC Genomics*   (2019) 20:1026

Page 4 of 10

pedigree for stem straightness and internal checking. Lower accuracy in ABLUP was observed for branch-cluster frequency when using the documented pedigree compared with the SNP-corrected pedigree. Lower accuracy was observed in GBLUP when using the documented pedigree than when using the SNP-corrected pedigree for all traits.

For branch-cluster frequency and stem straightness, the percentage of additive genetic variance explained by SNP markers was 54–64% in GBLUP using the documented pedigree, and 74–96% in GBLUP using the SNP-corrected pedigree. For internal checking and external resin bleeding, the percentage of additive genetic variance explained by SNP markers was 36–39% in GBLUP using the documented pedigree and 46–59% in GBLUP using the SNP-corrected pedigree.

Predictive ability of genomic selection

The predictive ability, defined as the average correlation between GEBVs from GBLUP in the cross-validation and EBVs from ABLUP using all phenotypes, increased for branch-cluster frequency, stem straightness, internal checking and external resin bleeding when using the SNP-corrected pedigree over the documented pedigree (Table 2). The predictive ability of genomic selection ranged from 0.47 to 0.54 for the four traits examined when using the documented pedigree and ranged from 0.55 to 0.70 when using the SNP-corrected pedigree. The predictive ability of traditional BLUP was higher than that from genomic selection, ranging from 0.65 to 0.77 for the four traits examined when using the documented pedigree, and ranged from 0.64 to 0.78 when using the SNP-corrected pedigree.

When using the documented pedigree, genomic selection was only superior to the traditional BLUP selection for branch-cluster frequency, and only reached 91–98% of the efficiency of traditional forward selection, with no clonal archive establishment, for the other traits. When using forward selection with the establishment of a clonal archive, the generation interval reduced from 17 years to 14 years and the efficiency of genomic selection was reduced. Genomic selection reached 75–81% of the efficiency of forward selection for stem straightness, internal checking and

external resin bleeding, and had similar efficiency to forward selection for branch-cluster frequency.

However, when the pedigree was corrected using SNP information, the efficiency of genomic selection over forward selection increased. When forward selection with a generation interval of 17 years was used, genomic selection was equivalent to forward selection for external resin bleeding but led to 37–115% additional benefit over forward selection for branch-cluster frequency, stem straightness and internal checking. When forward selection with a generation interval of 14 years was used, genomic selection only reached 84% of the efficiency of forward selection for external resin bleeding, but still obtained 13–77% extra genetic gain for branch-cluster frequency, stem straightness and internal checking.

Discussion

Genomic selection has been conducted for growth and wood properties in *Eucalyptus* [5], white spruce (*Picea glauca* (Moench) Voss) [6, 7], interior spruce (*Picea engelmannii* x *glauca*) [1, 23], and loblolly pine (*Pinus taeda* L.) [2]. Isik et al. [3] conducted genomic selection on growth and stem sweep in maritime pine (*Pinus pinaster* Ait.). For *Eucalyptus*, the accuracy of GEBVs across sites was 0.66–0.79 for growth traits and 0.65–0.88 for wood specific gravity within site [5]. For loblolly pine, the accuracy of genomic breeding values across four sites was 0.65–0.75 for diameter at breast height (DBH) and 0.63–0.74 for height [2]. The current study added two form traits and two wood defect traits that were evaluated for genomic selection in radiata pine. The predictive ability of genomic selection in cross-validation was 0.47–0.50 for branch-cluster frequency and stem straightness in the current study. A similar predictive ability (0.49) of genomic selection was reported for stem sweep in maritime pine [3]. Predictive abilities reported for growth traits were 0.46–0.55 in *Eucalyptus* [2] and 0.43–0.47 in maritime pine [3]. In white spruce, the predictive ability was 0.32–0.44 for wood and growth traits when both training and validation datasets shared individuals of the same families but decreased to 0.13–0.28 when training and validation datasets were made up of individuals from different families [6].

Table 2 The predictive abilities of genomic selection (r_{IH_g}) and traditional selection (r_{IH_a}) and the relative efficiency (E_{17} or E_{14}) of genomic selection over traditional BLUP selection for branch-cluster frequency, stem straightness, internal checking and external resin bleeding

Trait	Documented pedigree				SNP-corrected pedigree			
	r_{IH_g}	r_{IH_a}	E_{17} [§]	E_{14} [†]	r_{IH_g}	r_{IH_a}	E_{17}	E_{14}
Branch-cluster frequency	0.50	0.77	1.28	1.05	0.70	0.78	2.15	1.77
Stem straightness	0.47	0.73	0.98	0.81	0.55	0.72	1.51	1.25
Internal checking	0.51	0.71	0.92	0.75	0.59	0.70	1.37	1.13
External resin bleeding	0.54	0.65	0.91	0.75	0.57	0.64	1.02	0.84

[§] L_a = 17 years, [†] L_a = 14 years

When a SNP-corrected pedigree was used, the predictive ability was quite high (0.55–0.70) and seemed overestimated for branch-cluster frequency and stem straightness, given low heritabilities (0.13–0.22) reported in this paper. The low heritability might be because a narrow-sense heritability rather than a broad-sense heritability was reported. Another reason for the high predictive ability for these two traits could be because the EBV was estimated using a pedigree that was corrected by 704 markers. These 704 markers were a well-selected subset of whole genomic markers that were used in the GBLUP method to estimate GEBV.

Genomic selection can increase the amount of genetic gain per year that is delivered to the forest by shortening the breeding cycle. In the current study, the selection efficiency of genomic selection was 37–115% higher than traditional forward selection when the breeding cycle was reduced from 17 years to 9 years for branch-cluster frequency, stem straightness and internal checking. This is very similar to the efficiency of genomic selection reported in loblolly pine, where the selection efficiency per unit time in genomic selection was 53–112% higher than selection through phenotypes, assuming a reduction of 50% in the breeding cycle [2]. A higher selection efficiency of genomic selection was reported in interior spruce with an increase of 106–133%, assuming a 25% reduction in the breeding cycle [23]. In *Eucalyptus*, the efficiency of genomic selection over traditional selection was 50–100% for a reduction of 50% in the breeding cycle and 200–300% for a reduction of 75% in the breeding cycle [2]. However, simulations of a conifer breeding programme, with a training population size of 2000 and assuming a reduction in the breeding cycle from 17 years to 9 years, demonstrated additional genetic gain from genomic selection was 40% for a trait with low heritability and 95% for a trait with high heritability [18].

The best linear unbiased prediction (BLUP) methodology has been widely applied in livestock and plant breeding programmes to rank selection candidates [35]. It employs an average numerator relationship matrix, derived from the pedigree and based on expected relatedness between individuals, and incorporated in the mixed linear model equations [36]. Correct pedigree information is essential for accurately selecting the right individuals as parents of the next generation. However, pedigree errors are common in breeding programmes for both livestock and plant species, with an average of 10% error reported [37–41]. In the current study, the SNP-corrected pedigree re-assigned half of the documented parents, suggesting parentage error was around 50% in the training population. This pedigree error seemed high compared with that reported in the livestock and crop programmes mentioned above. Both error and missing genomic data could be contributing to the high parentage re-assignment we observed. The genotyping error rate in the exome capture GBS data was estimated to

be approximately 5%, based on replicated samples. The rate of missing genotypes in the data used for this parentage reconstruction was about 8%. Additional errors may also have been introduced by human operation throughout the whole process, from pollination to planting in the forest, and sample to collection to DNA extraction and genotyping.

Pedigree errors resulted in incorrect estimates of variance components and heritabilities and decreased breeding value accuracies. The genetic gains of breeding populations could be reduced by 4.3–17% when using incorrect pedigree information [37, 42, 43]. In the current study, the SNP-corrected pedigree considerably increased the accuracy of genomic selection, similar to that reported by Muñoz et al. [44]. The SNP-corrected pedigree also increased the percentage of variation explained by SNP markers from 36 to 64% to 46–96%, which suggests that it is the pedigree correction that increases the benefit of genomic selection over traditional BLUP selection.

Three types of narrow-sense heritabilities of branch-cluster frequency, stem straightness, internal checking and external resin bleeding were estimated using a model assuming homogeneous genetic variance and heterogeneous residual variances across sites. $\hat{h}_a^2$ was a pedigree-based heritability estimated through a BLUP model (ABLUP) that used the average numerator relationship calculated from pedigree. $\hat{h}_m^2$ was a marker-based heritability estimated thought a genomic BLUP (GBLUP) model that used genomic relationship matrix calculated from genomic data, which indicated a ratio of the additive genetic variation explained by genomic markers. $\hat{h}_{am}^2$ was a heritability estimated fitted both genomic relationship matrix and the average numerator relationship matrix simultaneously, which indicated a ratio of the additive genetic variation explained by genomic markers and the residual additive genetic variation that was not explained by genomic markers. Genomic selection was not quite efficient for capturing the additive genetic variations for all these NKT, only explaining 36–64% of total additive genetic variations. After correcting pedigree with the 704 parentage reconstruction markers, genomic selection captured most of the total additive genetic variation for branch-cluster frequency and stem straightness, however, it was not quite efficient for internal checking and external resin bleeding. Therefore, it is important for some traits to fit the residual polygenic genetic effects to capture the residual additive genetic variance when conducting genomic selection.

Narrow-sense heritabilities in the training population ranged from 0.09 to 0.28 for branch-cluster frequency and from 0.10 to 0.18 for straightness in the current study, where data from two trial series were combined in one analysis. Similar narrow-sense heritabilities within each trial series of the training population were reported

Li *et al. BMC Genomics* (2019) 20:1026

Page 6 of 10

by Li et al. [45], ranging from 0.13 to 0.28 for branch-cluster frequency and from 0.04 to 0.18 for stem straightness. The heritabilities for these two traits were also within the range reported in the literature. Heritability of branch-cluster frequency in radiata pine has been estimated as 0.19 in control-pollinated populations [46] and 0.37 in juvenile clones [47]. The heritability of stem straightness in radiata pine has been estimated as 0.11 to 0.17 in control-pollinated populations [30, 48, 49] and 0.28 for juvenile clones [47].

For radiata pine, low to moderate heritabilities were reported for external resin bleeding, and low to high heritabilities reported for internal checking in the literature. The narrow-sense heritability was 0.33 for external resin bleeding at a single site, and 0.40 for internal checking across two sites, in an open-pollinated progeny test of 224 first-generation families [31]. In a control-pollinated trial series with 150–165 pollen parents crossed to five Female Testers, the narrow-sense heritability was 0.16 for internal checking across two sites [31]. In another study, heritability for internal checking was 0.04–0.61 with an average of 0.35 at nine sites in six trial series [32].

The training population used in this study was limited both in terms of population size and available phenotypes, with only 988 clonally replicated genotypes available for branch-cluster frequency and stem straightness, and 465 for internal checking and external resin bleeding. Nevertheless, we found that the combination of a SNP-corrected pedigree and GBLUP resulted in accuracies that were acceptable (0.61–0.80). This is a very encouraging result for a population of this size. Accuracies of genomic selection are related to the size of the training population available and simulations suggest that higher accuracies can be achieved with larger training populations [10]. The accuracies of GEBVs for internal checking and external resin bleeding, for which less than half the individuals were phenotyped, were lower than that for branch-cluster frequency and stem straightness (Table 2). Increasing the number of genotypes tested for internal checking and external resin bleeding should increase the accuracy of their GEBVs. The accuracy of GEBVs will likely increase in the future as additional genotypes and phenotypes become available for an expanded training population.

The genotypes used in this study were tested in multiple environments (sites). The genetic model used assumed homogeneous genetic variance and heterogeneous residual variances across different environments. No genotype by environment interaction was considered in this study. A low level of genotype by environment interaction has been previously reported for internal checking, branch-cluster and stem straightness [31, 32, 50]. Li et al. [45] found there were considerable genotype by environment interactions

for branch-cluster frequency and stem straightness in both the POP2 and POP3 populations. Models that assumed heterogeneous genetic and residual variances, including factor analytic models [51], were also attempted but were unstable and would not converge.

Nevertheless, the accuracy and predictive ability of GEBVs for the traits investigated in this study are promising for the RPBC stakeholders, with potential applications for accelerating breeding of radiata pine. In the future, genomic selection will also be available for testing on additional traits, including key traits and resistance to diseases.

Conclusion

This study presents the first GEBVs for four non-key traits in the New Zealand radiata pine breeding programme, with a theoretical accuracy of 0.61–0.80 and a predictive ability of 0.55–0.70 for the traits examined, when using a pedigree corrected by SNP marker information. The predictive ability reported for the non-key traits in this study indicates that GEBVs are able to achieve an accuracy of 0.55–0.70 when used to predict individuals that are not included in the training population but have relatedness in common with the training population. These results are encouraging and indicate the method will be effective for operational implementation for these traits in radiata pine improvement. The results from this study appeared to favour the forward selection genomics approach, which will significantly reduce the generation interval of radiata pine. This has the potential to deliver benefits over forward selection of 13–77% or 37–115% for branch-cluster frequency, stem straightness and internal checking, with or without clonal archive establishment, respectively.

Materials and methods
Genetic material
Genetic material used in this study was provided by RPBC and data were collected from two RPBC clonally propagated radiata pine breeding trial series planted in New Zealand. Planting of the genetic material and collection of the data complied with the RPBC genetic material planting and data collection guidelines. Details of these two trial series are described by Li et al. [45], where the former was called POP2 and the latter POP3. The first trial series POP2 comprised 457 progeny from 63 parents and were planted in 1997 at two sites (Tarawera and Woodhill forests), with a single-paired mating design. The second trial series POP3 comprised 524 progeny from 24 parents and was planted in 1999 at three sites (Kinleith, Tarawera and Woodhill forests) with a factorial mating design. Tarawera and Kinleith forests located in the central North Island. Woodhill forest locates in the northwest of the North Island. The effective population size was 30.07

based on the status number for the training population [52, 53].

Phenotypic data

Branch-cluster frequency and stem straightness were assessed at age seven in POP3 and at age 8 in POP2. Branch-cluster frequency was assessed using a 9-point system where 1 = uninodal and 9 = extremely multinodal [49]. Stem straightness was also assessed using a 9-point subjective scale where 1 = crooked and 9 = very straight [48]. Internal checking was assessed as a visual score on a scale of 0–3 in POP3, where 0 = none, 1 = low, 2 = moderate, and 3 = severe. Equivalent visual scores for internal checking in POP2 were obtained by converting the percentage of collapse in increment-cores at breast height, assessed at age 9; 0 = below 3.5%, 1 = 3.5–4.5%; 2 = 4.5–6.5%; and 3 = greater than 6.5% [32]. The severity of external resin bleeding from bark split was assessed at age 9 in the POP2 trial series on a scale of 0–3, where 0 = none, 1 = low, 2 = moderate, and 3 = severe. Although these phenotypes were assessed as categorical traits, the distribution of their scores was close to a normal distribution. A summary of branch-cluster frequency, stem straightness, internal checking, and external resin bleeding data is presented in Table 3.

Genomic data

Four-hundred and sixty-five progeny from POP2, 523 progeny from POP3, and 117 unrelated individuals from the wider radiata pine breeding population (including 53 parents of POP2 and 24 parents of POP3) were genotyped using the exome capture genotyping by sequencing (GBS) method [12]. Details of SNP discovery and capture probe design and testing are described in [54]. The total number of SNPs markers genotyped was 1,371,123. The allele frequencies of these SNPs were calculated using the 117 unrelated individuals. Those SNP markers with a minor allele frequency of less than 0.03 were excluded from the analysis, leaving 67,168 SNP markers to be used in this study. The call rate of SNP markers for individual genotypes ranged from 0.60 to 0.93, with an average of 0.89. Where individual SNP genotypes were missing, substitution with the population mean for that SNP was used. Heterozygosity ranged from 0.11 to 0.35, with a mean of

0.28 and a standard deviation of 0.03 in POP2. Heterozygosity ranged from 0.11 to 0.41, with a mean of 0.27 and a standard deviation of 0.04 in POP3.

Statistical models

In this study, the predictive ability of GEBVs estimated using a GBLUP model that was based on the genomic relationship matrix was compared with those estimated using an ABLUP model that was based on the average numerator relationship matrix. The genomic relationship matrix was calculated based on genomic information whereas the average numerator relationship matrix on pedigree information. This study aimed to demonstrate the efficacy of genomic selection for non-key traits in radiata pine using existing clonally replicated trial datasets. The genetic parameters and EBVs from ABLUP were estimated through the linear mixed model described in eq. (1), with the assumptions of homogeneous additive and non-additive genetic variances, heterogeneous residual variances, heterogeneous variances for replication, set within replication and incomplete block across sites. Attempts were made assuming heterogeneous genetic variance across all sites, but a full genetic variance-covariance matrix was unable to estimate due to small numbers of genotypes at some sites.

$$y = X\beta + Z_a a + Z_d d + Z_r r + Z_w w + Z_b b + e \qquad (1)$$

where y is a vector of measurements, β is a vector of fixed effects (intercept and site), a is a vector of polygenic additive genetic effects following $Var(a) \sim N(0, \sigma_a^2 A)$ where σ_a^2 is the additive genetic variance and A is the pedigree-based average numerator relationship matrix [55], d is a vector of non-additive genetic effects following $Var(d) \sim N(0, \sigma_d^2 I)$ where σ_d^2 is the non-additive genetic variance fitting both dominance and epistatic effects and I is the identity matrix, r is a vector of replication effects following $Var(r) \sim N(0, P_0 \otimes I)$, where P_0 is a replication variance-covariance structure matrix with

$$P_0 = \begin{bmatrix} \sigma_{r_1}^2 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \sigma_{r_n}^2 \end{bmatrix}, \sigma_{r_i}^2$$ is the replication variance for

site i, w is a vector of set nested within replication following $Var(w) \sim N(0, W_0 \otimes I)$, where W_0 is a set nested within replication variance-covariance structure matrix with $W_0 = \begin{bmatrix} \sigma_{w_1}^2 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \sigma_{w_n}^2 \end{bmatrix}$, $\sigma_{w_i}^2$ is the set nested within replication variance for site i, b is a vector of incomplete block effects following $Var(b) \sim N(0, B_0 \otimes I)$ where B_0 is a block variance-covariance structure matrix with $B_0 = \begin{bmatrix} \sigma_{b_1}^2 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \sigma_{b_n}^2 \end{bmatrix}$, $\sigma_{b_i}^2$ is the incomplete block variance for site i, e is a vector of residual effects following

Table 3 Summary of statistics for NKTs in POP2 and POP3

Trial	Trait	N	Mean	SD	Min	Max
POP2	Branch-cluster frequency	5290	6.52	1.49	1	9
	Stem straightness	5289	6.68	1.40	1	9
	Internal checking	2732	1.26	0.99	0	3
	External resin bleeding	2275	0.89	0.87	0	3
POP3	Branch-cluster frequency	6851	4.64	1.87	1	9
	Stem straightness	6851	6.51	1.68	1	9

$Var(e) \sim N(0, R_0 \otimes I)$ with $R_0 = \begin{bmatrix} \sigma^2_{e_1} & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \sigma^2_{e_n} \end{bmatrix}$, $\sigma^2_{e_i}$ is the residual variance for site i, X, Z_a, Z_d, Z_r, Z_w and Z_b are incidence matrices assigning fixed and random effects to measurements in vector y, and n is the number of sites.

The genetic parameters and GEBVs from GBLUP were estimated through the linear mixed model described in eq. (2), with the assumptions of homogeneous additive and non-additive genetic variances, heterogeneous residual variances, heterogeneous variances for replication, set within replication and incomplete block across sites:

$$y = X\beta + Z_m m + Z_a a^* + Z_d d + Z_r r + Z_w w + Z_b b + e \qquad (2)$$

where terms are the same as those in eq. (1) except m is the vector of genomic breeding values following $Var(m) \sim N(0, \sigma^2_m G)$, σ^2_m is the additive genetic variance explained by markers and G is the normalized marker-based relationship matrix [56], a^* is the vector of residual polygenic additive genetic effects following $Var(a^*) \sim N(0, \sigma^2_{a^*} A)$, where $\sigma^2_{a^*}$ is the additive genetic variance not explained by genetic markers, Z_m and Z_{a^*} are the incidence matrices assigning random genomic effects and random residual additive genetic effects to measurements in vector y.

Three types of heritabilities were estimated and reported. Firstly, the pedigree-based heritability ($\hat{h}^2_{a_i}$) was estimated on the basis of variance components obtained from the pedigree-based analysis (ABLUP) as $\hat{h}^2_{a_i} = \frac{\hat{\sigma}^2_a}{\hat{\sigma}^2_a + \hat{\sigma}^2_d + \hat{\sigma}^2_{e_i}}$, where $\hat{\sigma}^2_a$ is the additive genetic variance, $\hat{\sigma}^2_d$ is the non-additive genetic variance and $\hat{\sigma}^2_{e_i}$ is the environmental variance at site i. Secondly, marker-based heritability ($\hat{h}^2_{m_i}$) representing the proportion of genetic variance explained by SNP markers in GBLUP, was estimated as $\hat{h}^2_{m_i} = \frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_m + \hat{\sigma}^2_{a^*} + \hat{\sigma}^2_d + \hat{\sigma}^2_{e_i}}$, where $\hat{\sigma}^2_m$ is the additive genetic variance explained by SNP markers, $\hat{\sigma}^2_{a^*}$ is the residual additive genetic variance not explained by SNP markers but captured by pedigree, $\hat{\sigma}^2_d$ is the non-additive genetic variance and $\hat{\sigma}^2_{e_i}$ is the residual variance at site i. Finally, a combined marker-based and pedigree-based heritability ($\hat{h}^2_{ma_i}$), representing the additive genetic variance fitted jointly by SNP markers and pedigree in GBLUP, was estimated as $\hat{h}^2_{ma_i} = \frac{\hat{\sigma}^2_m + \hat{\sigma}^2_{a^*}}{\hat{\sigma}^2_m + \hat{\sigma}^2_{a^*} + \hat{\sigma}^2_d + \hat{\sigma}^2_{e_i}}$.

The accuracy of both pedigree- and marker-based breeding values was estimated as $r = \sqrt{1 - \frac{PEV}{(1+F_i)\hat{\sigma}^2_a}}$ [36], where r is the accuracy of breeding values, PEV is the prediction error variance, F_i is the inbreeding coefficient

of individual genotype i, and $\hat{\sigma}^2_a$ is the estimated additive genetic variance linked to the breeding values, which is equal to σ^2_a in ABLUP and to σ^2_m in GBLUP. The percentage of additive genetic variance explained by SNP markers in GBLUP (%VA) was calculated as $\%VA = \frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_m + \hat{\sigma}^2_{a^*}} \times 100\%$.

Predictive ability and relative efficiency of genomic selection

The predictive ability of genomic selection refers to the accuracy achieved when GEBVs are used to predict the performance of individuals that are not included in the training population. Firstly, EBVs of all genotypes in the whole population were estimated with the ABLUP model using pedigree information and all phenotypes available. Secondly, GEBVs of all genotypes were estimated using the GBLUP model through a ten-fold random cross-validation procedure with ten replications [23]. The whole population was randomly divided into ten groups, each group having one-tenth of the genotypes. In each replication, GEBVs were estimated with the GBLUP model with nine groups of genotypes as a training population and the remaining group of genotypes as a validation population. Phenotypes of genotypes in the validation population were set as missing values. A Pearson product-moment correlation was calculated between GEBVs and EBVs of the genotypes in the validation population for each replication. The average of these correlations from ten replications was reported as the predictive ability of genomic selection.

The relative efficiency (E) of genomic selection over traditional BLUP selection that is based on phenotypes and pedigree indicates the benefit of genomic selection over traditional BLUP selection when generation interval is considered. Genomic selection is more efficient than traditional BLUP selection when E was over 100%, less efficient when E is below 100%, and equally efficient when E is equal to 100%. It was evaluated as a ratio of the expected genetic gain from GBLUP over the expected genetic gain from ABLUP per unit time. Therefore, we have:

$$E = \frac{\frac{i r_{IH_g} \sigma_{A_g}}{L_g}}{\frac{i r_{IH_a} \sigma_{A_a}}{L_a}} \cdot 100\% = \frac{r_{IH_g} \sigma_{A_g}}{r_{IH_a} \sigma_{A_a}} \cdot \frac{L_a}{L_g} \cdot 100\%$$

where i is the selection intensity, r_{IH_g} is the predictive ability of genomic selection that was calculated from 10-fold random cross-validation with GBLUP, r_{IH_a} is the predictive ability of traditional BLUP selection that was calculated from 10-fold random cross-validation with ABLUP, σ_{A_g} is the square root of additive genetic vari-

Li et al. BMC Genomics (2019) 20:1026

Page 9 of 10

ance explained by SNP markers from GBLUP using all phenotypes, σ_{A_a} is the square root additive genetic variance in ABLUP using all phenotypes, L_g is the generation interval of radiata pine for a selection based on GEBVs, and L_a is the generation interval of radiata pine for a selection based on EBVs. Based on a scenario analysis conducted by Li and Dungey [18] in conifers, L_g was 9 years in a forward selection strategy and L_a was 14 or 17 years for a forward selection with or without the establishment of a separate clonal archive of field-tested material, respectively.

Comparison between two pedigrees

The documented pedigree of the training population was found with errors. The parentage of the training population was reconstructed using a panel of 704 SNPs with a call rate of greater than 0.75 and a minor allele frequency of 0.35 and 0.5. These parentage reconstruction SNPs did not deviate from Hardy-Weinberg equilibrium and did not show evidence of linkage disequilibrium between them. Details of selection criteria and parentage reconstruction performance of these SNPs were described in [54]. An exclusion analysis approach that was described by Telfer et al. [57] was used to determine a trio relationship (a progeny and two candidate parents). For a given trio combination of a progeny and two candidate parents, an exclusion for the trio was considered if a SNP genotype in the progeny was not one of four genotypic combinations that were possibly formed from genotypes of the SNP in the candidate parents. The number of exclusions was calculated for all 704 SNPs for all possible trio relationship combinations of the documented candidate parents. Among all possible trio relationships related to a progeny, the trio relationship with the lowest number of exclusions was assigned as pedigree for the progeny.

Abbreviations
ABLUP: Best linear unbiased prediction based on the average numerator relationship matrix; BLUP: Best linear unbiased prediction; EBV: Estimated breeding value; GBLUP: Best linear unbiased prediction based on genomic relationship matrix; GBS: Genotyping by sequencing; GEBV: Genomic estimated breeding value; GS: Genomic selection; KT: Key trait; NKT: Non-key trait; POP2 and POP3: Two training populations used in this study; RPBC: Radiata Pine Breeding Company; SNP: Single nucleotide polymorphism

Authors' contributions
YL: conception, design of the work, analysis and interpretation of data, and writing; JK: conception and design of the work; ET: generation of data; PW: conception and generation of data; NG: generation of data. LM: generation of data; HSD: acquisition and conception. All authors read and approved the final version of the manuscript.

Availability of data and materials
The data used in the manuscript are available without any restriction from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
The manuscript has been approved for publication by its funding bodies: the Radiata Pine Breeding Company, the Ministry of Business, Innovation and Employment (MBIE) of New Zealand and the Scion Strategic Science Investment Funding.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Scion (New Zealand Forest Research Institute), Private Bag 3020, Rotorua 3046, New Zealand. [2]Agriclture Victoria, AgriBio Centre, 5 Ring Road, Bundoora, VIC 3083, Australia. [3]University of Otago, 362 Leith Steet, North Dunedin, Dunedin 9016, New Zealand.

References
1. El-Dien OG, Ratcliffe B, Klápště J, Chen C, Porth I, El-Kassaby YA. Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. BMC Genomics. 2015;16(1):1–16.
2. Resende MF Jr, Munoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MD, Kirst M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol. 2012;193(3):617–24.
3. Isik F, Bartholomé J, Farjat A, Chancerel E, Raffin A, Sanchez L, Plomion C, Bouffier L. Genomic selection in maritime pine. Plant Sci. 2016;242:108–19.
4. Isik F, Whetten R, Zapata-Valenzuela J, Ogut F, McKeand S. Genomic selection in loblolly pine - from lab to field. BMC Proc. 2011;5(Suppl 7):I8.
5. Resende MDV, Resende MFR Jr, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, et al. Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol. 2012;194(1):116–28.
6. Beaulieu J, Doerksen T, Clement S, MacKay J, Bousquet J. Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. Heredity (Edinb). 2014;113(4):343–52.
7. Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J. Genomic selection accuracies within and between environments and small breeding groups in white spruce. BMC Genomics. 2014;15(1):1048.
8. Li Y, Telfer E, Wilcox PL. New Zealand forestry enters the genomics era - applications of genomics in tree breeding. NZ J For. 2015;60(1):23–5.
9. Schefers JM, Weigel KA. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Anim Frontiers. 2012;2(1):4–9. https://doi.org/10.2527/af.2011-0032 .
10. Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genet Genomes. 2011;7(2):241–55.
11. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6(5):e19379.
12. Neves LG, Davis JM, Barbazuk WB, Kirst M. Whole-exome targeted sequencing of the uncharacterized pine genome. Plant J. 2013;75(1):146–56.
13. Jorde LB. Linkage disequilibrium as a gene-mapping tool. Am J Human Genet. 1995;56(1):11–4.
14. Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol. 2002;5(2):94–100.

15. Jorde LB. Linkage disequilibrium and the search for complex disease genes. Genome Res. 2000;10(10):1435–44.

16. Meuwissen TH, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001;157(4):1819–29.

17. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: Progress and challenges. J Dairy Sci. 2009; 92(2):433–43.

18. Li Y, Dungey HS. Expected benefit of genomic selection over forward selection in conifer breeding and deployment. PLoS One. 2018;13(12): e0208232.

19. Fernando R, Garrick D. Bayesian methods applied to GWAS. Methods Mol Biol. 2013;1019:237–74.

20. Fernando R, Dekkers J, Garrick D. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet Sel Evol. 2014;46(1):50.

21. Habier D, Fernando R, Kizilkaya K, Garrick D. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics. 2011;12:186.

22. Van Raden PM, Tooker ME, Cole JB. Can you believe those genomic evaluations for young bulls? J Dairy Sci. 2009;92(E-Suppl. 1):314.

23. Ratcliffe B, El-Dien OG, Klapste J, Porth I, Chen C, Jaquish B, El-Kassaby YA. A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* x *glauca*) using unordered SNP imputation methods. Heredity (Edinb). 2015;115(6):547–55.

24. Li Y, Dungey H, Yanchuk A, Luis A. Improvement of non-key traits in radiata pine breeding programme when long-term economic importance is uncertain. In: Forest Genetics for Production Conference; 2016. 14-15 March 2016, Rotorua New Zealand.

25. Kennedy SG, Yanchuk AD, Stackpole DJ, Jefferson PA. Incorporating non-key traits in selecting the *Pinus radiata* production population. NZ J For Sci. 2014;44(1):12–20.

26. Li Y, Dungey H, Yanchuk A, Apiolaza LA. Improvement of non-key traits in radiata pine breeding programme when long-term economic importance is uncertain. PLoS One. 2017;12(5):e0177806 doi: 0177810.0171371/journal. pone.0177806.

27. Jayawickrama KJS, Carson MJ. A breeding strategy for the New Zealand Radiata pine breeding cooperative. Silvae Genet. 2000;49(2):82–90.

28. Cown DJ, Donaldson LA, Downes GM. A review of resin features in radiata pine. NZ J For Sci. 2011;41:41–6.

29. Blackburn DP, Hamilton MG, Harwood CE, Baker TG, Potts BM. Assessing genetic variation to improve stem straightness in *Eucalyptus globulus*. Ann For Sci. 2013;70:461–70.

30. Wu HX, Ivković M, Gapare WJ, Matheson AC, Baltunis BS, Powell MB, McRae TA. Breeding for wood quality and profit in radiata pine: a review of genetic parameters. NZ J For Sci. 2008;38(1):56–87.

31. Kumar S. Genetic parameter estimates for wood stiffness, strength, internal checking, and resin bleeding for radiata pine. Can J For Res. 2004;34(12): 2601–10.

32. Kumar S, Cown DJ, Ivković M, Burdon RD. Susceptibility to intra-ring checking in *Pinus radiata*: potential for genetic improvement. NZ J For Sci. 2010;40:61–70.

33. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol. 2007;16(5):1099–106.

34. Slate J, Marshall T, Pemberton J. A retrospective assessment of the accuracy of the paternity inference program cervus. Mol Ecol. 2000;9(6):801–8.

35. Lynch M, Walsh JB. Genetics and analysis of quantitative traits. Sunderland: Sinauer Associates Inc. Publishers; 1998.

36. Mrode RA. Linear models for the prediction of animal breeding values. 3rd ed. London: Cambridge: CABI; 2014.

37. Long TE, Johnson RK, Keele JW. Effects of errors in pedigree on three methods of estimating breeding value for litter size, backfat and average daily gain in swine. J Anim Sci. 1990;68(12):4069–78.

38. Banos G, Wiggans GR, Powell RL. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. J Dairy Sci. 2001;84(11):2523–9.

39. Doerksen TK, Herbinger CM. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. Tree Genet Genomes. 2010;6(4): 591–600.

40. Zhang Z, Luo Y, Li Q, He J, Gao N, Zhang H, Ding X, Zhang Q, Li J. Developing and applying of a parentage identification approach based on high density genetic markers. Hereditas (Beijing). 2014;36(8):835–41.

41. Visscher PM, Woolliams JA, Smith D, Williams JL. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J Dairy Sci. 2002;85(9):2368–75.

42. Geldermann H, Pieper U, Weber WE. Effect of misidentification on the estimation of breeding value and heritability in cattle. J Anim Sci. 1986;63(6): 1759–68.

43. Israel C, Weller JI. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. J Dairy Sci. 2000;83(1):181–7.

44. Muñoz PR, Resende MFR, Huber DA, Quesada T, Resende MDV, Neale DB, Wegrzyn JL, Kirst M, Peter GF. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. Crop Sci. 2014;54(3):1115–23.

45. Li Y, Wilcox P, Telfer E, Graham N, Stanbra L. Association of single nucleotide polymorphisms with form traits in radiata pine in the presence of genotype by environment interactions. Tree Genet Genomes. 2016;12:63. https://doi.org/10.1007/s11295-11016-11019-11296 .

46. Wu HX, Matheson AC. Genotype by environment interactions in an Australia-wide radiata pine diallel mating experiment: implications for regionalized breeding. For Sci. 2005;51(1):29–40.

47. Burdon RD, Banister MH, Low CB. Genetic survey of *Pinus radiata*. 4. Variance structures and heritabilities in juvenile clones. NZ J For Sci. 1992;22: 187–210.

48. Carson MJ. Control-pollinated seed orchards of best general combiners - a new strategy for radiata pine improvement. Plant Breed Sym. 1986;5:144–9.

49. Carson SD. Genotype × environment interaction and optimal number of progeny test sites for improving *Pinus radiata* in New Zealand. NZ J For Sci. 1991;21:32–49.

50. Li Y, Suontama M, Burdon RD, Dungey HS. Genotype by environment interaction in the forest tree breeding: review methodology and perspectives on research and application. Tree Genet Genomes. 2017;13(60):1–18.

51. Cullis BR, Jefferson P, Thompson R, Smith AB. Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. Theor Appl Genet. 2014;217:2193–210.

52. Lindgren D, Gea L, Jefferson P. Loss of genetic diversity monitored by status number. Silvae Genet. 1996;45:52–9.

53. Lindgren D, Gea LD, Jefferson PA. Status number for measuring genetic diversity. For Genet. 1997;4:69–76.

54. Telfer E, Graham N, Macdonald L, Li Y, Klápště J, Resende M Jr, Neves LG, Dungey H, Wilcox P. A high density exome capture genotype-by-sequencing panel for forestry breeding in *Pinus radiata*. PLoS One. 2019; 14(9):e0222640.

55. Wright S. Coefficients of inbreeding and relationship. Amer Nat. 1922;56: 330–8.

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

57. Telfer EJ, Stovold GT, Li Y, Silva-Junior OB, Grattapaglia DG, Dungey HS. Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: a comparative analysis of marker data power and robustness. PLoS One. 2015;10(7):e0130601.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.