RESEARCH ARTICLE

Expected benefit of genomic selection over forward selection in conifer breeding and deployment

Yongjun Li[1,2]*, Heidi S. Dungey[1]

1 Scion (New Zealand Forest Research Institute), Rotorua, New Zealand, **2** Agriculture Victoria, AgriBio Centre, DEDJTR, Bundoora, Victoria, Australia

* Yongjun.Li@ecodev.vic.gov.au

Abstract

Genomic selection is a proven technology in animal and plant breeding to accelerate genetic gain, but as yet is to be fully realised in forest tree breeding. This paper examines, through stochastic simulation, the potential benefits of genomic selection (GS) over forward selection (FS) in a typical conifer breeding program. Methods of speeding the deployment of selected material were also considered, including top-grafting onto mature seed orchard ortets, using additional replicates of clones in archives for crossing, and embryogenesis and clonal propagation. Genetic gain per generation was found to increase considerably when the size of the training population was larger (800 c.f. 3000 clones), or when the heritability was higher (0.2 c.f. 0.5). The largest genetic gain, of 24% was achieved where large training populations (3000 clones) and high heritability traits (0.5) were combined. The accuracy of genomic breeding values (GEBVs) increased with the increase in the number of clones in the training population, the heritability of the trait and the density of the SNP markers. Calculated accuracies of simulated GEBVs and genetic gain per unit of time suggested that 2000 clones in the training population is the minimum size for effective genomic selection for conifers. Compared with forward selection, genomic selection with 2000 clones in the training population, and a 60K SNP panel, an increase of 1.58 mm per year in diameter-at-breast-height (DBH) and 2.44 kg/m^3 per year for wood density can be expected. After one generation (9-years), this would be equivalent to 14.23 mm and 21.97 kg/m^3 for DBH and wood density respectively. Deploying clones of the selected individuals always resulted in higher additional genetic gain than deploying progeny/seedlings. Deploying genetic material selected from genomic selection with top-grafting for early coning appeared to be the best option. Application of genomic selection to conifer breeding programs, combined with deployment tools such as top-grafting and embryogenesis are powerful tools to speed the delivery of genetic gain to the forest.

plan and monitored progress of the research, and approved to publish this manuscript.

Introduction

Many animal and plant breeders now use genomic selection routinely [1, 2, 3]. In forestry, genomic selection is yet to be fully realised. The benefits of genomic selection in forestry are potentially very large through shortening the long generation intervals, speeding the delivery of genetic gain, and reducing the reliance on field testing [4–6]. Benefits also include the estimation of both the additive genetic variance and the non-additive genetic variance [7], which potentially gives further benefits. Partitioning the non-additive genetic variance can add to the accuracy of breeding value estimates due to overall greater accuracy in the variance partitioning process. Trees also have the ability to be vegetatively propagated (cloned), so that non-additive genetic variance can be captured by deploying copies of outstanding genotypes [8, 9].

In practice, the accuracy of genomic selection directly depends on four factors [10, 11]: 1) the level of linkage disequilibrium between SNPs and quantitative trait loci (QTL), 2) the number of varieties/clones with phenotypes and genomic data in the training population from which the SNP effects are estimated, 3) the heritability of the trait in question, and 4) the distribution of QTL effects. The first two factors are under the control of the breeder and the other two are not [10]. Trait heritability can be improved through experimental design and testing strategies [12–15], although the underlying genetic control will remain the same. Linkage disequilibrium is a particularly important variable when considering the prospects of genomic selection in forest trees [4]. The level of linkage disequilibrium can be increased by reducing the effective population size and increasing SNP marker density [4]. A large effective population size leads to low level of linkage disequilibrium and, in turn, reduces the prediction accuracy in genomic selection [4, 16], however, it can also result in more recombination and a greater genetic diversity [17], which is favourable for long-term genetic gain [18].

Genomic tools have been used in research for a number of forestry species now [7, 19–28]. Despite this, some breeding programs are still reluctant to move even to forward selection from a backward selection strategy [29, 30]. Some practical simulations have been undertaken to determine the opportunity that cloning has for delivering genetic gain [9]. Very few simulations for the application of genomics in forestry have been published. Very few simulations have properly considered different deployment pathways that are available to practitioners in order to maximise the opportunity available through each of the different breeding pathways.

Clonal forestry is promising to provide quality planting material to forest growers in the shortest time. It is defined as the commercial production and deployment of plants of field-tested individual genotypes of forest tree species [31]. Clonal forestry is now a commercial reality for many conifer species, such as *Abies, Larix, Picea, Pinus,* and *Pseudotsuga* [32] and contributes approximately 5% of the total current New Zealand planting stock requirements. The advantages of clonal forestry in radiata pine include reducing the rotation age, and efficiently capturing additive and non-additive genetic gains [31, 33–35]. An ability to rapidly multiply valuable crosses is another key advantage [33]. The disadvantages include increases in operational cost [31] and reduction of genetic diversity [36, 37]. Clonal forestry for conifers is achieved mostly by the propagation of cuttings [38, 39] or by somatic embryogenesis [40]. Somatic embryogenesis is the most powerful way to deliver genetic gain to the forest because cell lines can be maintained in a juvenile state indefinitely through cryopreservation.

This paper will simulate the combination of breeding, genomics and deployment methods in order to critically examine the potential benefits of genomic selection in a typical conifer breeding program. Key variables tested include training population sizes, heritability and length of breeding cycles. Genetic gains will be estimated for combinations of variables for tree breeding and genomics approaches. Benefits from the different deployment options employed for the different breeding approaches will be important for the realisation of genetic gain [41].

We will simulate the genetic gain from deployment using embryogenesis from cryopreserved cell lines [40, 42, 43] and seed production, with the option of speeding seed production by top-grafting [44]. These simulations will, for the first time, clearly show the expected genetic gain per year for forest companies to strategically evaluate the opportunity that genomics offers for forest productivity.

Materials and methods

Assumptions and structures—Conifer breeding cycle timelines

We consider two primary conifer breeding approaches; forward selection with clonal testing (FS; Fig 1) and forward selection with genomics (GS, Fig 2). With FS, there are six steps per breeding cycle: 1) crossing parents and generating seeds in a crossing orchard (3 years), 2) sowing seed and preparing stool plants for cuttings (2 years), 3) progeny and field testing, with installation of a clonal crossing archive at the same time (9 years), and 4) scale up the selected individuals with sufficient cones for crossing (3 years). Therefore, the forward selection process has a generation interval of 17 years. Using a genomic selection strategy (GS), progeny field testing is not required. Selection candidates can be ranked based on genomic breeding values (GEBVs) that can be estimated when seedlings are only six months old. The selected trees are kept in a seed orchard for further 5.5 years before crossing can occur. Overall, generation interval can be reduced to 9 years with genomic selection (Fig 2).

One time-saving option for shorter breeding cycle relevant to the FS approach, is the establishment of a clonal archive (CA) at the time of progeny testing, where ramets of each clone under progeny testing are also kept in an archive for crossing on a site known to promote
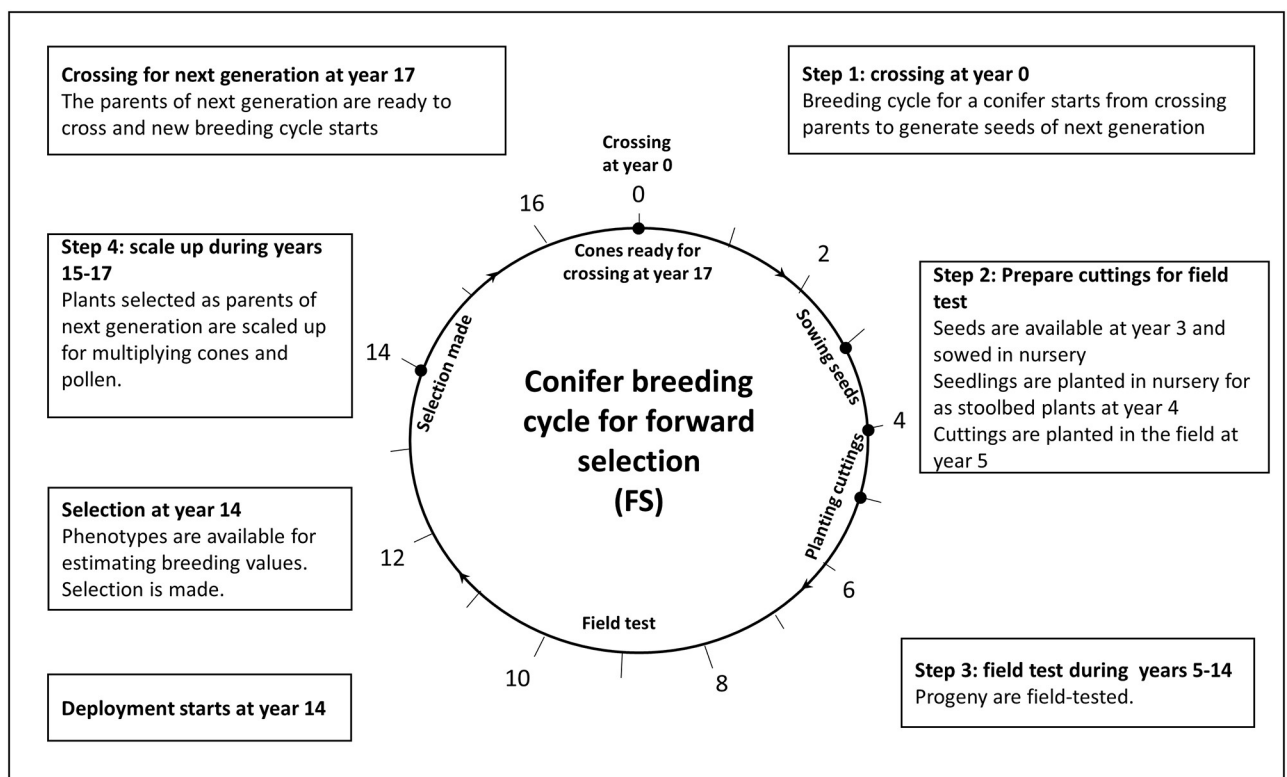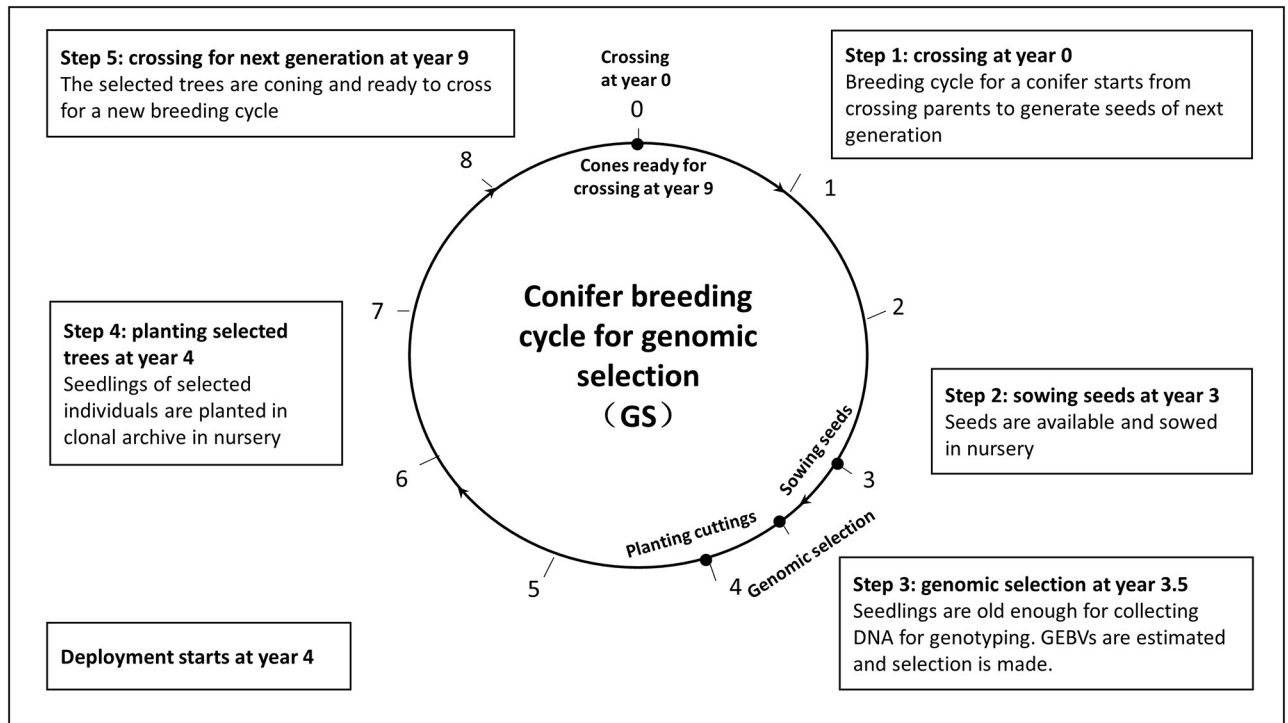

Fig 1. Example of conifer breeding cycle for conventional forward selection with a generation interval of 17 years.

https://doi.org/10.1371/journal.pone.0208232.g001

Fig 2. Example of conifer breeding cycle for genomic selection with a generation interval of 9 years.

good coning. Clones are therefore available with cones for crossing right after selection is made. With this clonal archive option, step 4 is not needed (Fig 1) and the generation interval is reduced to 14 years. This option of forward selection is called forward selection with a clonal archive (FSCA). Top-grafting can also be used to graft juvenile scions of selected individuals onto the top of mature rootstock in order to accelerate coning [45, 46]. This top-grafting process is assumed to reduce the age of coning from 5 years to 3 years and, in turn, reduces the generation interval to 7 years. This option is relevant only to GS and is therefore called genomic selection with top grafting (GSTG).

Assumptions and structures—Deployment pathways

Two deployment pathways were simulated to ensure the best selections reach the forest: deploying progeny or deploying clones (Fig 3). The first option is to generate seeds by crossing the selected trees, establishing individual plants as stool beds (mother plants) from which cuttings are taken and then deploying rooted cuttings into the forest with the following steps: graft the selected trees (three years) → cross and generate seeds (three years) → either i) sow seeds or ii) use embryogenesis to generate trees for planting (one additional year) → stool plants (one year) → deployment in the forest. It therefore takes eight years from selection to deployment in the forest for FS. For the forward selection with clonal archive (FSCA) and genomic selection (GS and GSTG) pathways, traditional grafting is not needed, so that only five years from selection to deployment in the forest are required.

The second option is to deploy clones of selected trees, where immature seeds (embryos) of all progeny are collected during step 1 of the breeding cycle as described in Figs 1 and 2. Somatic embryogenesis is then used to generate plants which are subsequently established as stool beds to produce cuttings. (We have assumed all genotypes are able to go through this
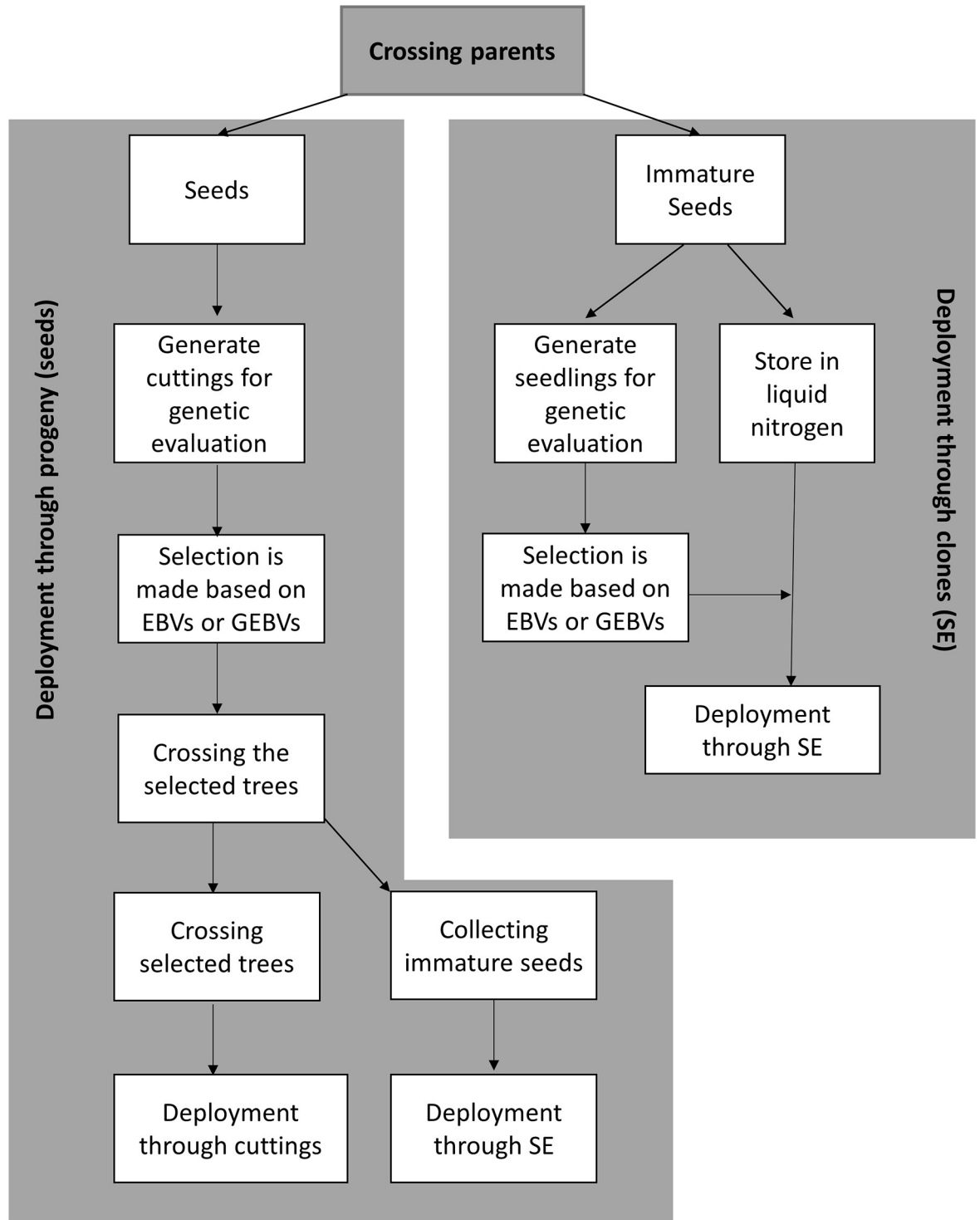
Fig 3. Two deployment options: Deployment of progeny or clones of the selected trees.

https://doi.org/10.1371/journal.pone.0208232.g003

Table 1. Four selection schemes in breeding population, their combinations with two deployment options, generation interval, numbers of years needed from crossing parents to selection and from selection to deployment in the forest.

Selection in breeding population	Material of selected trees to deploy	Deployment abbreviation§	Generation interval	Years from crossing to selection	Years from selection to deployment
FS	Progeny	FS$_s$	17	14	8
FS	Clone	FS$_c$	17	14	1
FSCA	Progeny	FSCA$_s$	14	14	5
FSCA	Clone	FSCA$_c$	14	14	1
GS	Progeny	GS$_s$	9	9	5
GS	Clone	GS$_c$	9	7	1
GSTG	Progeny	GSTG$_s$	7	7	5
GSTG	Clone	GSTG$_c$	7	7	1

§ Subscription $_s$ denotes deploying progeny (seedlings) of the selected individuals and subscription $_c$ denotes deploying clones from tissue culture.

https://doi.org/10.1371/journal.pone.0208232.t001

process). Tissue of all the genotypes tested in the field are preserved in liquid nitrogen (-120˚C) while the trees are grown and can be assessed. Once selections are made from field tests, phenotypes or GEBVs, the preserved cell lines of the selected trees are thawed and cuttings are prepared via stool beds in the nursery, for large-scale clonal deployment. Generating and deploying cuttings to the forest takes one year.

There are eight combinations for deployment between the four breeding and selection schemes (FS, GS, FSCA, GSTG) and two generic deployment options (via seed or clones). Table 1 summarizes these selection combinations, their generation intervals, numbers of years needed from crossing the parents to selection and from selection to deployment in the forest.

Simulation

Population structure.   Three steps were used to simulate a population for conducting forward selection or genomic selection and implemented using the QMSim software [47] based on a forward-in-time process [48]. The objective of these steps was to create initial linkage disequilibrium and establish mutation-drift equilibrium.

The first step created a base population that was named "Historical Generations". This step consisted of running 500 generations with 300 female parents and 300 male parents. Random mating was used to undertake a random union of gametes. No selections were made in this step. This step had a constant population size of 1000. The second step, called "Extended Generations", consisted of 10 generations with 100 female and 100 male parents. This step had a constant population size of 500.

After "Extended Generations", ten "Recent Discrete Generations" were simulated in the third step, by selecting 32 female parents and 32 male parents from the last generation. A circular mating design was implemented and each parent was crossed twice. Fifteen offspring were generated in each family and each offspring had 4 ramets. These parameters used in the recent generations mimicked more closely to a conifer breeding program such as the Elite Population in *P. radiata* [29]. Selections were made based on BLUP EBVs, estimated using Henderson's mixed linear model equations [49] and implemented using ASREML-R [50].

Two quantitative traits with low heritability ($h^2 = 0.2$) and with high heritability ($h^2 = 0.5$) were simulated. Both traits had additive genetic variance of 300 based on similar genetic variances for diameter-at-breast-height (mm) and wood density (kg/m^3), respectively. Three-hundred and fifty bi-allelic QTLs were assumed to control the traits. The true breeding value of an individual was calculated as the sum of the individual QTL additive effects. Non-additive

Table 2. Five SNP panels were simulated with 600, 1,200, 2,500, 4,900 and 8,000 SNPs in one chromosome, which was equivalent to 7K, 14K, 30K, 60K and 90K SNPs.

Panel	SNPs simulated	SNPs included in deriving GEBV	Equivalent at whole genome level (assuming 12 chromosomes)
7K	600	576	6,912
14K	1,200	1,168	14,016
30K	2,500	2,358	28,296
60K	4,900	4,766	57,192
90K	8,000	7,149	85,788

genetic effects between ramets of the same genotype were sampled from a normal distribution with mean of zero and 15% additive genetic variance. Phenotypes were simulated as the sum of the true breeding value, non-additive genetic effects and random residuals.

Groups of 800, 1000, 2000 or 3000 randomly selected clones from generations 4 to 8 (in the "Recent Discrete Generations") were used as the training population. All individuals in generation 10 were used as the prediction population. The prediction equations of SNPs or SNP effects were estimated in the training population. We then tested the power of the prediction equations of SNPs estimated from the training population to predict performance on the prediction population.

Simulation of the conifer genome. In this study, one 150 cM chromosome and five SNP panels of 600, 1,200, 2,500, 4,900 and 8,000 SNP markers randomly distributed on the chromosome were simulated. This was equivalent to a SNP density of 7,000, 14,000, 30,000, 60,000 and 90,000 SNPs at the whole genome level (Table 2). The markers were neutral in their effects on the traits. Initial minor allele frequencies of these markers were larger than 0.10. The three hundred and fifty modelled bi-allelic QTLs were randomly distributed. Additive allelic effects of QTLs were randomly sampled from a gamma distribution with a shape parameter of 0.50 and a scale parameter of 26. Markers and QTLs were simulated to have a recurrent mutation rate of 10^{-5} in the Historical Generations and the Extended Generations but no mutations in markers or QTLs were simulated in the Recent Generations.

Simulation of training and prediction populations. Four training population with a size of 800, 1000, 2000 or 3000 clones were randomly chosen from the Recent Generations set (generations 4 to 8) to reflect conifer breeding populations under selection. The prediction population were composed of 960 clones from the 10th generation.

Genetic evaluation

The EBVs of individual clones were used as input parameters for deriving individual-clone GEBVs, estimated using linear individual tree mixed models [51]:

$$y = \mu + Z_a a + Z_d d + e \tag{1}$$

where y is a vector of trait phenotypes, μ is the overall mean, a is a vector of additive genetic effects, d is a vector of non-additive genetic effects, e is a vector of random residuals. Z_a and Z_d are incidence matrices relating the additive and non-additive genetic effects to phenotype. It was assumed that $\mathrm{var}(a) = A\sigma_a^2$, $\mathrm{var}(d) = D\sigma_d^2$, $\mathrm{var}(e) = I\sigma_e^2$ and $\mathrm{var}(y) = Z_a A Z_a' + Z_d D Z_d' + I\sigma_e^2$, where A is the additive genetic relationship matrix, D is the non-additive genetic relationship matrix and I is the identity matrix, σ_a^2 is the additive genetic variance, σ_d^2 is the non-additive genetic variance and σ_e^2 is the residual variance [51].

GEBV estimation

The allele substitution effect for each SNP was computed using Gensel [52], in which a Bayesian method called BayesC [53, 54] was used for estimating marker effects. BayesC fitted a statistical model assuming a known fraction of SNPs as having zero effects, which in the current analysis was set to be $\pi = 0.95$. Additive SNP effects were fitted to every SNP in the model:

$$y_i = \mu + \sum_{j=1}^{k} X_{ij} a_j + e \tag{2}$$

where y_i is a vector of EBVs estimated from trait phenotypes and pedigree data, X_{ij} is the incidence matrix related to copy number of a given allele of individual i at SNP j, a_j is a vector of additive effects of SNP j, and e is a vector of random residuals. The priors for a_j were a mixture of normal distributions as described by Habier et al. [54]. The priors of all SNP effects have a common variance, which had a scaled inverse chi-square prior with parameters $va = 4.2$ and scale S_a^2. S_a^2 is a univariate student's t-distribution, $t(0, va, S_a^2)$. The effect of a SNP fitted with probability $(1-\pi)$ comes from a mixture of multivariate student's t-distributions, $t(0, va, IS_a^2)$. The prior for the residual effects is normally distributed with mean zero and variance σ_e^2 (based on an arbitrary allocation as an initial value). Gibbs sampling was used to sample the posterior distribution of model parameters. SNP effects were estimated by the mean of the sampled values. The total number of Markov Chain Monte Carlo iterations used for estimating posterior means of marker effects and variances was 40,000. The EBV of a simulated trait was used as response variable and the accuracy of the EBV was used as weighting factor. The weight for the ith individual clone was estimated according to Garrick et al. [55] as

$w_i = \frac{\frac{1-h^2}{h^2}}{c + \frac{1-r_i^2}{r_i^2}}$, where h^2 is the heritability, c is the part of the genetic variance not explained by

SNPs, and r_i is the accuracy of the EBV of the ith clone.

GEBVs ($\hat{g}$) were predicted as the linear combination of the SNP substitution effects as

$$\hat{g} = X'\hat{a} \tag{3}$$

where X is the matrix of SNP genotypes for each individual clone in the prediction population and $\hat{a}$ is the vector of estimated SNP effects. The accuracy of the GEBVs was estimated as the correlation between the GEBVs and the true simulated breeding value.

Linkage disequilibrium

The underlying assumption of genomic selection is that haplotypes at some loci are in linkage disequilibrium with QTL alleles that affect the traits that are subject to selection. Linkage disequilibrium was created through the Historical Generations and Extended Generations to mimic level of linkage disequilibrium of a conifer population, based on knowledge of the radiata pine breeding population. Linkage disequilibrium at the beginning of the Recent Generations was measured as R^2, the squared correlation coefficient between pairs of loci:
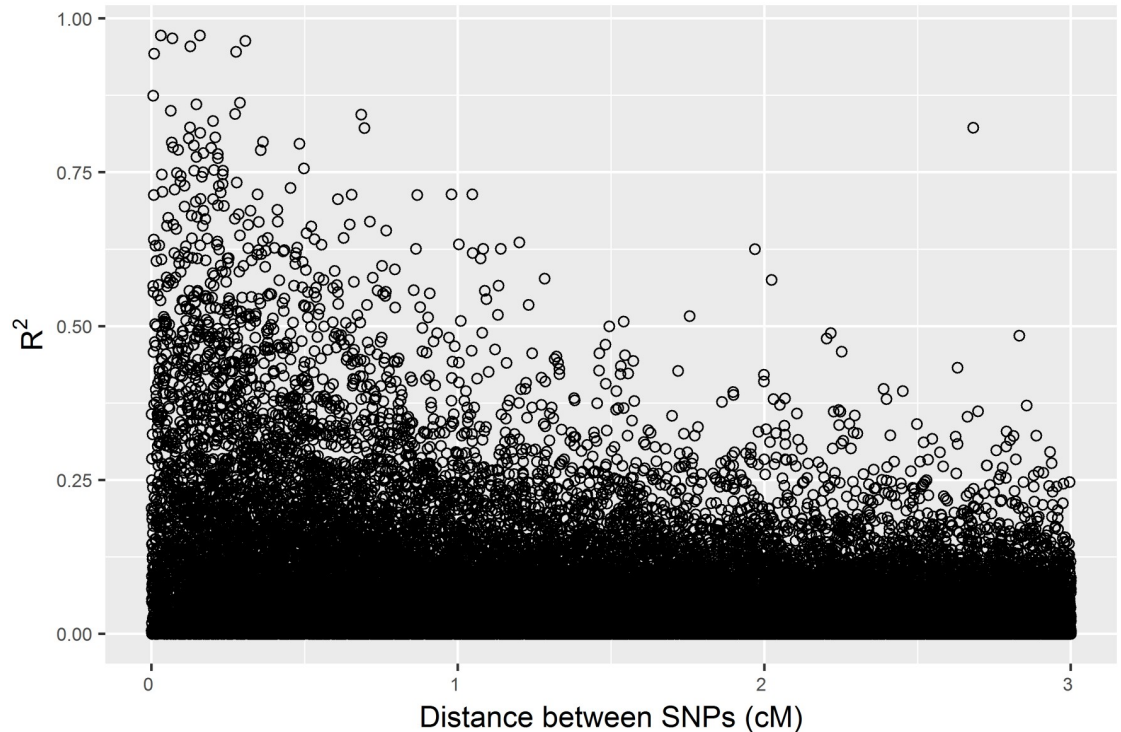
$$R^2 = \frac{D^2}{P_A P_a P_B P_b} \tag{4}$$

where D is the deviation of the observed frequency of a haplotype from the expected with

$$D = P_{AB} - P_A P_B, \tag{5}$$

where P_{AB}, P_a, P_A, P_B and P_b are observed frequencies of haplotypes AB and of alleles A, a, B, b,

Fig 4. Plot of the squared correlations of allele frequencies (R^2) versus distance in cM between closely located SNPs in a panel with 7K SNPs across 150 cM linkage map.

respectively. R^2 is a more informative measure of LD as it measure the overall departure from complete independence between pairwise combinations of polymorphisms [56].

This simulation assumed that the conifer genome has low level of linkage disequilibrium and that linkage disequilibrium rapidly decayed within 1 centimorgan. The average LD was 0.17, 0.20, 0.22, 0.22 and 0.24 for 7K, 14K, 30K, 60K and 90K SNP panels, respectively. Fig 4 shows that the linkage disequilibrium (R^2) rapidly decays with increasing distance between closely located SNPs in the 7K SNP panel.

Benefit of genomic selection over forward selection

Genetic gain in breeding population. In the simulated breeding population, nine clones were selected from 960 clones in the prediction population, which was equivalent to a selection proportion of 1%. Genetic gain was given as the average true breeding value of the selected clones, representing the increase above the population mean of zero. Genetic gain per generation and per year was compared among selections based on their EBVs and GEBVs. Additional genetic gain from selection methods of FSCA, GS and GSTG over FS was calculated in the same manner.

Genetic gain in deployment. Genetic gain from deployment was estimated as the total genetic gain from a plantation of 8 rotations in a hectare of forest. It was assumed that one rotation started immediately after the previous rotation had finished. Each rotation was assumed to be 25 years. The first rotation was assumed to start at the same time that the breeding cycle was started in the breeding population, either for GS or FS. At the start of a new rotation, it was assumed that any deployment could only use the genetic material available from the latest breeding cycle. Genetic merit was the average true breeding value of individual trees
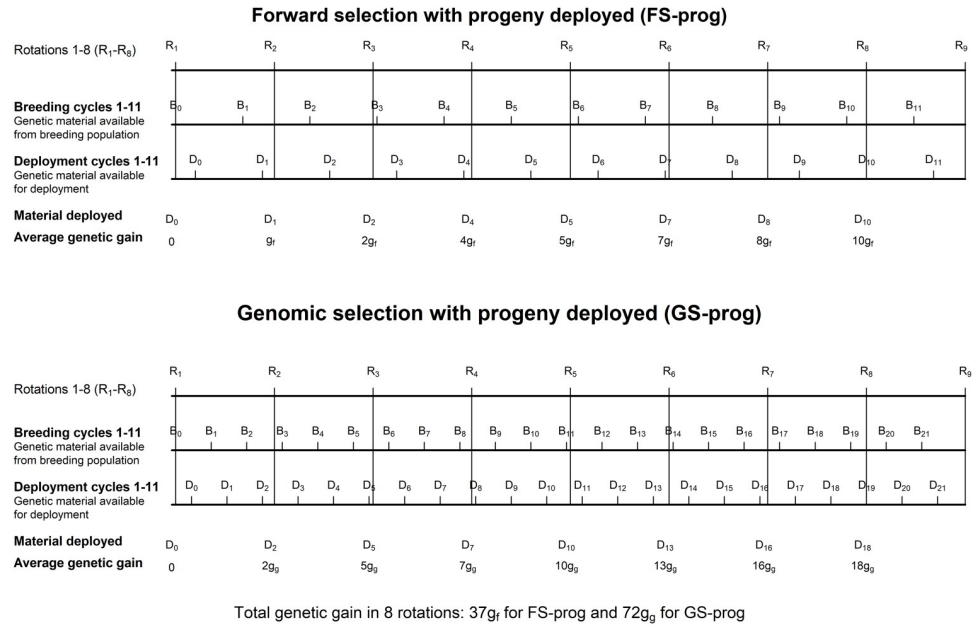
Fig 5. Diagram of calculating total genetic gain from deployment of progeny of the selected individuals for 8 rotations for forward selection (FS$_s$) and genomic selection (GS$_s$). R$_i$ is the start of i rotation with 25 years long, B$_i$ is the time when selection is made in breeding cycle i, D$_i$ is the time when genetic material is available for deployment in breeding cycle i, which is 5 years after selection in breeding cycle i. g_f and g_g are the genetic gain per generation from FS$_s$ and GS$_s$, respectively.

selected to be used as parents in the breeding population. The increment of genetic merit in each breeding cycle was assumed to be constant. Genetic merit in breeding cycle i was calculated as $g(b_i - 1)$, where g was the average true breeding value of selected parents and b_i was ith breeding cycle. The genetic merit of ith rotation was the genetic merit of genetic material that was ready for deployment of the latest breeding cycle. The total genetic gain from a deployment for 8 rotations was calculated as the summation of the average genetic merit of genetic material deployed in 8 rotations. Fig 5 shows an example of how the total genetic gain in the deployment of progeny was calculated for forward selection (FS$_s$) and for genomic selection (GS$_s$). For instance, the genetic gain in rotation 3 for FS$_s$ ($2g_f$) indicates that genetic material from the latest breeding cycle (breeding cycle 2) was deployed. The total genetic gain from a deployment for 8 rotations was estimated as $37g_f$ for FS$_s$ and $72g_g$ for GS$_s$. Additional genetic gain from deployment options of FSCA, GS and GSTG over FS was calculated for two deployment options, deployed as progeny and as clones.

Results

Accuracy of GEBVs

The accuracy of GEBVs increased with the increase in the number of individuals in the training population, heritability of the trait of interest, and the density of SNP markers (Table 3). The accuracy of GEBVs increased from 0.21 to 0.51 for a trait with low heritability and from 0.67 to 0.75 for a trait with high heritability, when the training population size increased from 800 to 3000. The accuracy of GEBVs in genomic selection was close to the accuracy of FS for both traits with either low or high heritabilities when the training population size was 3000. The accuracy of GEBVs increased from 0.28 to 0.45 for a trait with low heritability and from

Table 3. Accuracies of GEBVs using a 60K SNP panel and a different number of individuals (800, 1,000, 2,000 and 3,000) in the training population for low ($h^2 = 0.2$) and high ($h^2 = 0.5$) heritability traits, compared with the accuracy of equivalent EBVs.

No. of individuals in the training population	$h^2 = 0.2$		$h^2 = 0.5$	
	GEBV	EBV	GEBV	EBV
800	0.21	0.52	0.67	0.75
1000	0.43	0.52	0.69	0.75
2000	0.44	0.52	0.73	0.75
3000	0.51	0.52	0.75	0.75

https://doi.org/10.1371/journal.pone.0208232.t003

Table 4. Accuracies of GEBVs for different numbers of makers in different sizes of SNP panel: 7K, 14K, 30K, 60K or 90K SNPs used in the prediction population for low ($h^2 = 0.2$) and high ($h^2 = 0.5$) heritabilities, compared with the accuracy of equivalent EBVs. Marker effects were estimated using 1,000 clones in the training population.

SNP density	$h^2 = 0.2$		$h^2 = 0.5$	
	GEBV	EBV	GEBV	EBV
7K	0.28	0.52	0.51	0.75
14K	0.37	0.52	0.58	0.75
30K	0.40	0.52	0.60	0.75
60K	0.43	0.52	0.69	0.75
90K	0.45	0.52	0.70	0.75

https://doi.org/10.1371/journal.pone.0208232.t004

0.51 to 0.70 for a trait with high heritability, when the SNP density increased from 7K to 90K across the simulated chromosome (Table 4). The accuracy of the GEBVs for a trait with a low heritability was much lower than that for a trait with a high heritability for SNPs panels with various numbers of SNPs.

Genetic gain in the breeding population

Genetic gain was used to evaluate the efficiency of genomic selection when compared with forward selection using phenotypic data and field testing. Genetic gain per generation increased from 4.22% to 16.02% for a trait with low heritability and 18.90% to 24.20% for a trait with high heritability when the number of clones used in the training population increased from 800 to 3000 (Table 5). For a trait with low heritability, the genetic gain per generation from GS was always lower than that from FS across all training population sizes examined. For traits with a high heritability, genetic gain from GS was higher than that from FS when the training population size was 2000 or more. Additional genetic gain was observed in traits with higher heritability.

Table 5. Genetic gain per generation obtained in the prediction population from GS with 60K SNP panel for different training population sizes (800, 1,000, 2,000 and 3,000 clones) for traits with low ($h^2 = 0.2$) and high ($h^2 = 0.5$) heritabilities in the breeding population, compared with genetic gains obtained from FS.

Training population	$h^2 = 0.2$		$h^2 = 0.5$	
size	GS	FS	GS	FS
800	4.22	19.13	18.90	21.30
1000	10.47	19.13	19.21	21.30
2000	14.23	19.13	21.97	21.30
3000	16.02	19.13	24.20	21.30

https://doi.org/10.1371/journal.pone.0208232.t005

Table 6. Genetic gain per year obtained in the breeding population from genomic selection with (GSTG) or without (GS) top grafting and from forward selection with (FSCA) and without (FS) clonal archive for traits with low ($h^2 = 0.2$) or high ($h^2 = 0.5$) heritabilities for various training population sizes of 800, 1,000, 2,000 and 3,000. GEBVs were estimated using a 60K SNP panel.

Heritability	Training population size	Genetic gain/year				Benefit over FS		
		FS	FSCA	GS	GSTG	FSCA	GS	GSTG
$h^2 = 0.2$	800	1.13	1.37	0.47	0.60	21%	-58%	-47%
	1000	1.13	1.37	1.16	1.50	21%	3%	33%
	2000	1.13	1.37	1.58	2.03	21%	40%	80%
	3000	1.13	1.37	1.78	2.29	21%	58%	103%
$h^2 = 0.5$	800	1.25	1.52	2.10	2.7	22%	68%	116%
	1000	1.25	1.52	2.13	2.74	22%	70%	119%
	2000	1.25	1.52	2.44	3.14	22%	95%	151%
	3000	1.25	1.52	2.69	3.46	22%	115%	177%

Table 6 shows the simulated genetic gain expected in the breeding population per year from FSCA, GS and GSTG relative to the genetic gain achieved from FS. The largest genetic gain was obtained from GSTG. Genetic gain per unit of time in FS was 1.13 for trait with a low heritability and 1.25 for trait with a high heritability. The use of a clonal archive in FS (FSCA) increased gains/year by 21–22%. Genetic gain also increased with the increase of the size of the training population. For the low heritability, when the training population size was 1000, yearly genetic gain from GS was similar to that from forward selection (FS) but lower than the gain from forward selection with a clonal archive (FSCA). When the training population was 2000–3000 clones, genomic selection (GS) led to 40–58% more genetic gain/year than FS. Accelerating coning in genomic selection with top-grafting (GSTG) led to 33%, 80% and 103% additional genetic gain/year over FS for training population sizes of 1000, 2000 and 3000, respectively. For the highly heritable trait, GS resulted in 68–115% additional genetic gain/year and 116–177% more for training population sizes of 800–3000, compared with forward selection (FS).

Genetic gain in deployment

Deploying clones through somatic embryogenesis after FS (FS_c) led to 8.11% additional genetic gain over deploying progeny (seedlings) of the selected individuals (FS_s) (Fig 6). Use of a clonal archive in combination with forward selection led to considerably higher genetic gain in deployment than forward selection alone. Additional genetic gain available from deploying progeny of the selected individuals through a clonal archive ($FSCA_s$) was 21.62%. Deploying individuals via a clonal archive followed by clonal propagation ($FSCA_c$) provided an additional 24.32% compared with FS_s.

Compared with the baseline of deploying progeny of the selected individuals from FS (FS_s), the additional genetic gains obtained from deploying genetic material selected from GS increased with increasing heritability and training population size (Table 7). Deploying genetic material selected with GS was always superior to deploying genetic material selected from FS except where the training population had 800 individuals and the heritability was low (0.2). Increasing the training population size always increased the additional benefit obtained from deploying genetic material selected from GS over FS. Deploying genetic material selected from GS was inferior to deploying material selected from FS when the heritability was low and the training population size was 800. This suggested that a bigger training population size was needed for a trait with low heritability.
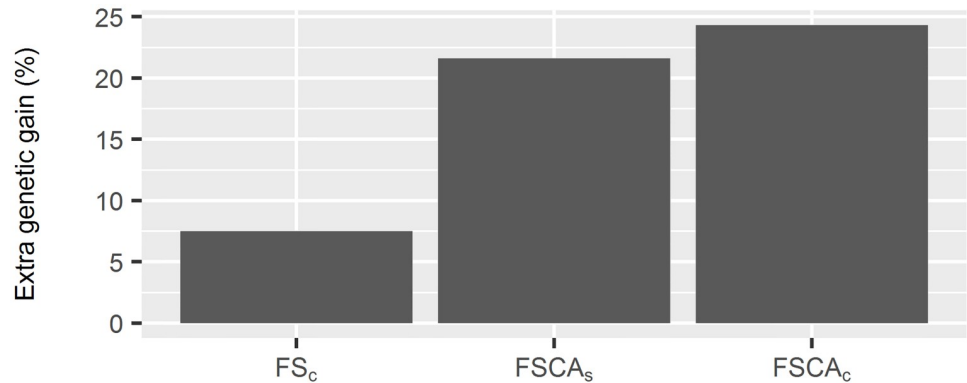
Fig 6. Additional genetic gain per year from FSCA when deployed as seedlings (FSCA$_s$), or clones (FSCA$_c$), when compared with deployment of selected individuals from FS via seedlings (FS$_s$).

https://doi.org/10.1371/journal.pone.0208232.g006

Table 7. Additional genetic gains (percentage) across 8 rotations obtained from deploying genetic material selected from GS over a baseline of deploying progeny of the selected individuals from FS (FS$_s$) for different deployment options: Deploying progeny (seedlings) of the individuals selected from GS (GS$_s$), deploying clones of the individuals selected from GS (GS$_c$), deploying progeny of the individuals selected through GSTG (GSTG$_s$) or deploying clones of the individuals selected through GSTG (GSTG$_c$).

Heritability	Training population size	GS$_s$	GS$_c$	GSTG$_s$	GSTG$_c$	GS$_c$ vs GS$_s$	GSTG$_c$ vs GSTG$_s$	GSTG$_s$ vs GS$_s$	GSTG$_c$ vs GS$_c$
$h^2 = 0.2$	800	-57	-56	-45	-42	1	3	12	14
	1000	7	9	38	43	2	5	31	34
	2000	45	49	87	95	4	8	42	46
	3000	63	67	110	120	4	10	47	53
$h^2 = 0.5$	800	73	77	123	133	4	10	50	56
	1000	76	80	127	136	4	9	51	56
	2000	101	106	159	170	5	11	58	64
	3000	121	127	186	198	6	12	65	71

https://doi.org/10.1371/journal.pone.0208232.t007

With genomic selection (GS and GSTG), deploying clones was always superior to deploying progeny (seedlings) of the selected individuals (Table 7). The additional genetic gain obtained from deploying clones of the selected individuals (GS$_c$) was 1–4% and 4–6% higher than that from deploying progeny of the selected individuals (GS$_s$) for a trait with low heritability and for a trait with high heritability, respectively. The additional genetic gain obtained from deploying clones of the individuals selected from GSTG (GSTG$_c$) was 3–10% and 9–12% higher than that from deploying progeny of the individuals selected from GSTG (GSTG$_s$) for a trait with low heritability and high heritability, respectively.

When comparing genetic gains obtained from deploying genetic material selected from GS, the use of top-grafting to accelerate coning remarkably increased genetic gains from deployment (Table 7). Deploying progeny selected from GSTG$_s$ led to 12–47% and 50–65% higher additional genetic gain than deploying progeny selected from GS$_s$ for a trait with low heritability and high heritability, respectively. Deploying clones selected from GSTG$_s$ led to 14–53% and 56–71% higher additional genetic gain than deploying clones selected from GS$_s$ for a trait with low heritability and high heritability, respectively.

Discussion

The objective of this study was to simulate the effectiveness of genomic selection over traditional selection using phenotypes and pedigree information in conifer species. There are

several key components to achieve this: increasing the accuracy of estimated breeding values, shortening the generation interval of the breeding population and shortening the time frame from the point of selection to deployment in the forest. The accuracy of genomic breeding values increased with the training population size and the number of markers included in developing genomic breeding values. These accuracies were in the range of those achieved in genomic selection in the literature. The accuracy of GEBVs for growth and wood density was 0.30–0.83 in loblolly pine (*Pinus taeda* L.) [57]; 0.52–0.69 for growth and 0.71–0.79 for wood density in white spruce (*Picea glauca*) [58, 59] and; 0.42–0.65 for wood quality traits and 0.63–0.76 for growth in black spruce (*Picea mariana*) [17]. The simulated accuracy achieved from genomic selection was lower or equivalent to that achieved in forward selection (Tables 3 and 4), which reflected the population size and the number of markers used in most genomic selection projects in conifers [17, 57–59].

While this study focussed on increasing accuracies with increasing marker densities, there are other considerations when designing genomic breeding programmes. This includes designing the genotyping tools to be tailored to the population and traits. In this study higher marker density and higher heritabilities gave higher accuracies of estimated breeding values. Other genomic selection simulations have shown that increasing marker density will better describe the genetics of traits when compared with traditional genetic analysis, for example the recent simulation by [60]. High marker density is the most effective where the heritability is low and where the effective population size is larger. Although this trend is the same for higher heritability traits, marker density is not as important. Of course, this is largely intuitive, but an important point that has not been covered in this study.

The heritability of a trait is an important as it effects the efficiency of genomic selection over traditional selection [10]. It has been assumed that genomic selection is more beneficial where traits have a low heritability [5, 11]. The current results do not support this assumption, as higher additional genetic gain from genomic selection was observed with higher heritabilities over forward selection both in breeding and deployment pathways. It is possible that where heritability is low, more markers or a bigger training population is needed in order to obtain reliable accuracy of GEBVs. Two quantitative traits with $h^2 = 0.2$ and $h^2 = 0.5$ were simulated. These two traits were representative of the growth trait diameter-at-breast-height (DBH) and wood density. Extrapolating from the current simulation under forward selection, and single-trait selection, we can expect an increase of 1.13 mm per year for DBH and 1.25 kg/m^3 per year for wood density. Over a 17-year breeding cycle, the expected increases would be 19.13 mm for DBH and 21.3 kg/m^3 for wood density. With genomic selection, using 2000 individuals in the training population and 60K SNPs for developing GEBVs, we can expect an increase of 1.58 mm per year in DBH and 2.44 kg/m^3 per year for wood density. After one generation (9-years), this would be equivalent to 14.23 mm for DBH and 21.97 kg/m^3 for wood density.

This simulation is a generic study for evaluating the efficiency of genomic selection in conifer breeding, assuming very simple situations, such as simulating a single chromosome, no missing phenotypic and genomic data, and discrete recent generations. In reality, wood density has a high heritability and is expensive to measure. It has not been normal, historically, for example to assess wood density for all individuals in a trial. The level of linkage disequilibrium (LD) in a genome is also a key factor that affects the accuracy of genomic selection. This simulation therefore attempted to mimic linkage disequilibrium in conifers using knowledge from radiata pine. Low levels of LD and a rapid decay of LD have been observed in conifer genomes. LD decays about 50% from $r^2 \approx 0.5$ to $r^2 \approx 0.25$ over 2000 base pair segments in *Pinus taeda* [61] and from $r^2 \approx 0.25$ to $r^2 \approx 0.10$ over 2000 base pair segments in *Pseudotsuga mensiesii* [62]. The average LD in this simulation was 0.17 to 0.24 for 600 SNPs to 8000 SNPs over one

chromosome, which was equivalent to a density of 7K to 90K on the whole radiata pine genome, as a model genome. Chromosome length used in this study was 150cM based on data from previous molecular genetics studies at Scion, which implied 1800 cM in length for the whole radiata pine genome. The possible number of base pairs is over 20 Gb in conifers and 25 Gb in radiata pine [63]. However, due to the low recombination rate in conifers [64], a genetic linkage map might not be larger than other species that have a smaller number of base pairs [65]. However, the genome length used in this simulation is still short compared with genome lengths in some other species. For example, the genome length is 2730 cM in humans [66] and 2300 cM for cattle [67].

Juvenility is generally defined as the period during which a plant cannot be induced to produce cones [68, 69]. Some species of conifers initiate coning only a year or so after being established through grafting or rooted cuttings in a nursery, for example, 10–12 months for *Pinus mugo* [70]. *Pinus* species such as *P. banksiana* Lamb., *P. virginiana* Mill., and *P. contorta* Dougl. var. *contorta* begin coning within 3–5 years of germination [71]. In most forest tree species, the time from seed germination to the onset of coning is about 10 years whereas, in some species, coning does not begin until 25–35 or more years after germination, such as *Fagus sylvatica* L. and *Picea abies* (L.) Karst [72]. Research has been initiated to develop accelerated breeding techniques to reduce the number of years between selection and crossing [73]. The current study showed that top-grafting has the potential to effectively reduce generation interval of conifers and potentially increase genetic gain per unit of time. Examples of using top-grafting as a tool to accelerate coning for conifers include *P. taeda* L., where generation interval can be shortened to a minimum of three years and breeding cycle to five years [74–76].

The onset of sexual reproduction will no doubt be a major barrier to shortening the generation interval even further than simulated in this study. The current study assumed that generation interval can be reduced by two years. The time reduced through the use of top-grafting might be different for different conifer species. Coning can, however, be accelerated in conifers using a number of methods which could also be considered. Coning can be accelerated through wide spacing, fertilization and irrigation, the development of full tree crowns and rapid early growth are promoted so that trees reach coning size as soon as possible [77]. Gibberellins, plant hormones that accelerate growth and induce or promote coning of plants, have also been used to accelerate coning in conifers. Gibberellins were used to induce coning from 4–5 years to 3 months for Arizona cypress (*Cupressus arizonica* Greene), from 20–25 years to 12 years for Norway spruce (*Picea abies* (L.) Karst), from 10–20 years to 6 years for Douglas-fir (*Pseudotsuga menziesii*, (Mirb.) Franco) from 20–30 years to 3 years for western hemlock (*Tsuga heterophylla* (Raf.) Sarg.) [78]. Coning of jack pine (*Pinus banksiana* Lamb) can be induced from 5–10 years in naturally regenerated stands under open-grown conditions to 12 months under near optimum growing conditions in the greenhouse and nursery [79, 80]. Time to coning was effectively reduced by application of top-grafting and gibberellins in scots pine (*Pinus sylvestris* (L.)) [81]. Early-coning in conifers can be also achieved through transformation of the gene Coning Locus T (FT), an important coning regulator, which causes plants to cone at a very young age when the plants are quite small [82–84].

One impediment that prevents somatic embryogenesis to be applied commercially in clonal forestry for most of conifers is its high costs [85]. This problem has been largely solved for some conifer species, for example radiata pine. It has been become a reliable and cost-effective commercial clonal forestry development and delivery system in radiata pine [31]. About four million of radiata pine clonal plants generated through embryogenesis have been deployed annually in New Zealand and this number is expected to be expanded substantially. Production costs for somatic embryogenesis clonal forestry is about two times of that for control-

pollinated seedlings with a 100% of successful rate per cross (Carson, personal communication, 2018). With the development of this technology, production costs will be reduced in the future for most of conifer species. Fluidic systems are also under development that increase yield of somatic embryos, which should further increase the competitiveness of clonal systems [86].

Therefore, somatic embryogenesis was simulated as a pathway to deploy genetic gain to the forest. There are likely to be at least two ways of using somatic embryogenesis in conifer breeding programmes, outlined in Fig 3. One pathway is described in this study, where immature seeds of every individual to be tested in the breeding programme are collected, cell lines developed and then used to generate plants for testing or genomic evaluation while also being cryopreserved in liquid nitrogen. Once a selection is made, embryos of the selected individuals can be grown for propagation and deployment. A second pathway is to use somatic embryogenesis when deploying progeny of the selected individuals. Immature embryos are collected after crossing selected parents and replicated through tissue culture for mass propagation via stoolbeds. The time frame from selection to deployment for this option in the current study is equivalent to that for rooting of cuttings. The advantage of this somatic embryogenesis deployment option is that a very large number of copies of a clone can be generated.

Due to the high cost of genotyping, other studies have often only genotyped part of a training population using a high SNP density assay. The remainder of the population is genotyped using a low SNP density assay. Imputation is often used to 'fill in the blanks' through SNP associations, and predict genotypes to the levels of a high SNP density assay [87]. We have not considered this strategy here. A possible next step in the simulation would be to consider a more complicated situation of high- and low-density SNP assays with imputation to determine the validity of this approach in conifers. Only two scenarios—low and high heritability traits—were considered. Investigation of more heritability options and multiple traits will be beneficial in informing conifer breeding programs how to best optimise the implementation of GS in their populations.

Conclusions

The accuracy of GEBVs increased with the increase in the number of individuals in the training population, heritability of the interested trait and the density of SNP markers. Accuracy of GEBVs and genetic gain per unit of time for the trait with lower heritability was a key factor determining the minimum training population size, suggesting that 2000 clones in the training population is the minimum size for effective genomic selection for conifers. Deploying clones of the selected individuals always resulted in higher additional genetic gain than deploying progeny. Deploying genetic material selected from genomic selection with top-grafting for early coning resulted in 12–71% additional genetic gain in the deployment than genomic selection without an early coning option. Application of genomic selection to conifer breeding programs, combined with deployment options such as top-grafting and somatic embryogenesis are powerful tools to speed the delivery of genetic gain to the forest.

Acknowledgments

interval of selection strategies, and Mari Suontama for discussion of selection and deployment strategies.

Author Contributions

Conceptualization: Heidi S. Dungey.

Data curation: Yongjun Li.

Formal analysis: Yongjun Li.

Funding acquisition: Heidi S. Dungey.

Methodology: Yongjun Li, Heidi S. Dungey.

Project administration: Yongjun Li.

Resources: Heidi S. Dungey.

Writing – original draft: Yongjun Li, Heidi S. Dungey.

Writing – review & editing: Yongjun Li, Heidi S. Dungey.

References

1. Shi F, Tibbits J, Pasam RK, Kay P, Wong D, Petkowski J, et al. Exome sequence genotype imputation in globally diverse hexaploid wheat accessions. Theor Appl Genet. 2017. Epub 2017/04/06. https://doi.org/10.1007/s00122-017-2895-3 PMID: 28378053.

2. Brito LF, McEwan JC, Miller SP, Pickering NK, Bain WE, Dodds KG, et al. Genetic diversity of a New Zealand multi-breed sheep population and composite breeds' history revealed by a high-density SNP chip. BMC Genet. 2017; 18(1):25. Epub 2017/03/16. https://doi.org/10.1186/s12863-017-0492-8 PMID: 28288558.

3. Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K, Clark MD. Genomic innovation for crop improvement. Nature. 2017; 543(7645):346–54. https://doi.org/10.1038/nature22011 PMID: 28300107

4. Grattapaglia D, Resende MDV. Genomic selection in forest tree breeding. Tree Genet Genomes. 2011; 7(2):241–55.

5. 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.

6. Isik F. Genomic selection in forest tree breeding: the concept and an outlook to the future. New For. 2014; 45. https://doi.org/10.1007/s11056-014-9422-z

7. Zapata-Valenzuela J, Isik F, Maltecca C, Wegrzyn J, Neale D, McKeand S, et al. SNP markers trace familial linkages in a cloned population of *Pinus taeda*-prospects for genomic selection. Tree Genet Genomes. 2012; 8. https://doi.org/10.1007/s11295-012-0516-5

8. Shaw DV, Hood JV. Maximizing gain per effort by using clonal replicates in genetic tests. Theor Appl Genet. 1985; 71(3):392–9. https://doi.org/10.1007/BF00251178 PMID: 24247443

9. Shelbourne CJA, Kumar S, Burdon RD, Gea LD, Dungey HS. Deterministic simulation of gains for seedling and cloned main and elite breeding populations of *Pinus radiata* and implications for strategy. Silvae Genet. 2007; 56(6):259–70.

10. 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. https://doi.org/10.3168/jds.2008-1646 PMID: 19164653

11. Goddard M. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica. 2009; 136(2):245–57. https://doi.org/10.1007/s10709-008-9308-0 PMID: 18704696

12. Williams ER, John JA, Whitaker D. Applications: example of block designs for plant and tree breeding trials. Aust N Z J Stat. 1999; 41(3):277–84. https://doi.org/10.1111/1467-842X.00082

13. Williams ER, John JA. Construction of row and column designs with contiguous replicates. J R Stat Soc Ser C Appl Stat. 1989; 38(1):149–54. https://doi.org/10.2307/2347689

14. Williams ER, Matheson AC, Harwood CE. Experimental design and analysis for tree improvement. 2nd ed. Australia: CSIRO Pub.; 2002.

15. Butler DG, Smith AB, Cullis BR. On the design of field experiments with correlated treatment effects. J Agric Biol Environ Statistics. 2014; 19(4):539–55. https://doi.org/10.1007/s13253-014-0191-0

16. Lin Z, Hayes BJ, Daetwyler HD. Genomic selection in crops, trees and forages: a review. Crop and Pasture Sci. 2014; 65(11):1177–91. https://doi.org/10.1071/CP13363

17. Lenz PRN, Beaulieu J, Mansfield SD, Clément S, Desponts M, Bousquet J. Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). BMC Genomics. 2017; 18(1):335. https://doi.org/10.1186/s12864-017-3715-5 PMID: 28454519

18. Falconer D, Mackay T. Introduction to quantitative genetics. 4th ed. Harlow, Essex, England: Longman Group Ltd. Pearson Education Ltd; 1996.

19. 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/s11295-016-1019-6

20. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, et al. The genome of Eucalyptus grandis. Nature. 2014; 510(7505):356–62. Epub 2014/06/12. https://doi.org/10.1038/nature13308 PMID: 24919147.

21. Bartholome J, Van Heerwaarden J, Isik F, Boury C, Vidal M, Plomion C, et al. Performance of genomic prediction within and across generations in maritime pine. BMC Genomics. 2016; 17(1):604. Epub 2016/08/16. https://doi.org/10.1186/s12864-016-2879-8 PMID: 27515254.

22. Di Pierro EA, Mosca E, Rocchini D, Binelli G, Neale DB, La Porta N. Climate-related adaptive genetic variation and population structure in natural stands of Norway spruce in the South-Eastern Alps. Tree Genet Genomes. 2016; 12(2):16. https://doi.org/10.1007/s11295-016-0972-4

23. Gamal El-Dien O, Ratcliffe B, Klapste J, Porth I, Chen C, El-Kassaby YA. Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. G3 (Bethesda). 2016; 6(3):743–53. Epub 2016/01/24. https://doi.org/10.1534/g3.115.025957 PMID: 26801647.

24. Hedrick PW, Hellsten U, Grattapaglia D. Examining the cause of high inbreeding depression: analysis of whole-genome sequence data in 28 selfed progeny of *Eucalyptus grandis*. New Phytol. 2016; 209(2):600–11. https://doi.org/10.1111/nph.13639 PMID: 26356869

25. Mousavi M, Tong C, Liu F, Tao S, Wu J, Li H, et al. De novo SNP discovery and genetic linkage mapping in poplar using restriction site associated DNA and whole-genome sequencing technologies. BMC Genomics. 2016; 17:656. Epub 2016/08/20. https://doi.org/10.1186/s12864-016-3003-9 PMID: 27538483.

26. Porth I, Bull G, Ahmed S, El-Kassaby YA, Boyland M. Forest genomics research and development in Canada: Priorities for developing an economic framework. For Chron. 2015; 91(01):60–70. https://doi.org/10.5558/tfc2015-011

27. Thistlethwaite FR, Ratcliffe B, Klápště J, Porth I, Chen C, Stoehr MU, et al. Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. BMC Genomics. 2017; 18(1):930. https://doi.org/10.1186/s12864-017-4258-5 PMID: 29197325

28. Klápště J, Suontama M, Telfer E, Graham N, Low C, Stovold T, et al. Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of Eucalyptus nitens. PLOS ONE. 2017; 12(9):e0185137. https://doi.org/10.1371/journal.pone.0185137 PMID: 28938023

29. Dungey HS, Brawner JT, Burger F, Carson M, Henson M, Jefferson P, et al. A new breeding strategy for *Pinus radiata* in New Zealand and New South Wales. Silvae Genet. 2009; 58(1–2):28–38.

30. Jayawickrama KJS, Carson MJ. A breeding strategy for the New Zealand Radiata Pine Breeding Cooperative. Silvae Genet. 2000; 49(2):82–90.

31. Carson MJ, Carson SD, Te Riini C. Successful varietal forestry with radiata pine in New Zealand. NZ J For. 2015; 60(1):8–11.

32. Park Y-S, Klimaszewska K, editors. Archievements and challenges in conifer somatic embryogenesis for clonal forestry. the XII World Forestry Congress; 2003; Quebec City, Canada.

33. Carson M. Advantages of clonal forestry for *Pinus radiata*—Real or imagined? NZ J For Sci. 1986; 16(3):403–15.

34. Carson MJ, Carson SD, editors. Clonal forestry of radiata pine. Lessons for Europe. Vegetative Propagation & Deployment of Varieties—The Scope for Europe; 2009 21–23 April 2009; Scotland, UK.

35. Rezende GD, editor Clonal forestry of Eucalyptus—in Europe and South America. Vegetative Propagation and Deployment of Varieties—The Scope for Europe; 2009 21–23 April 2009; Scotland, UK.

36. Gullberg U. The vegetative propagation of forest trees Sveriges Skogsvardsforbunds Tidskrift. 1980; 1/2:104–11.

37. Burdon RD, Aimers-Halliday J. Risk management for clonal forestry with *Pinus radiata*—Analysis and review. 1. Strategic issue and risk spread. NZ J For Sci. 2003; 33(2):156–80.

38. Park YS. Implementation of conifer somatic embryogenesis in clonal forestry: technical requirements and deployment considerations. Ann For Sci. 2002; 59:651–6.

39. Menzies MI, Faulds T, Dibley MJ. Production of radiata pine cuttings for plantation forestry in New Zealand. Acta Hortic. 1992; 319:359–64.

40. Hargreaves CL, Reeves CB, Find JI, Gough K, Menzies MI, Low CB, et al. Overcoming the challenges of family and genotype representation and early cell line proliferation in somatic embryogenesis from control-pollinated seeds of *Pinus radiata*. NZ J For Sci. 2011; 41:97–114.

41. Balocchi CE, editor Realised versus operational gain: the role of propagation strategies. IUFRO working party S20219 '97 Genetics of radiata pine FRI Bulletin No 203, 253–255; 1997.

42. Hargreaves CL, Foggo MN, Smith DR, Gordon ME. Development of protocols for the cryopreservation of zygotic embryos of *Pinus radiata* and subsequent plant regeneration. NZ J For Sci. 1999; 29:54–63.

43. Cyr DR, Attree SM, El-Kassaby YA, Ellis DD, Polonenko DR, Sutton BCS. Application of somatic embryogenesis to tree improvement in conifers. In: Noriyuki M, Atsushi K, editors. Progress in Biotechnology. Volume 18: Elsevier; 2001. p. 305–12.

44. Almqvist C. Survival and strobili production in topgrafted scions from young *Pinus sylvestris* seedlings. Scand J For Res. 2013; 28(6):533–9.

45. Lott LH, Lott LM, Stine M, Kubisiak TL, Nelson CD, editors. Top grafting longleaf x slash pine F1 hybrids on mature longleaf and slash pine interstocks. the 27th Southern Forest Tree Improvement Conference; 2003 24-27th June; Stillwater, Okla.

46. Goading GD, Bridgwater FE, Bramlett DL, Lowe WJ. Top grafting loblolly pine in the western gulf region. the 25th Biennial Southern Forest Tree Improvement Conference; 11–14 July; New Orleans, Louisiana, USA. 1999.

47. Sargolzaei M, Schenkel FS. QMSim: a large-scale genome simulator for livestock. Bioinformatics. 2009; 25(5):680–1. https://doi.org/10.1093/bioinformatics/btp045 PMID: 19176551

48. Carvajal-Rodríguez A. Simulation of genomes: A review. Curr Genomics. 2008; 9(3):155–9. https://doi.org/10.2174/138920208784340759 PMID: 19440512

49. Henderson C. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975; 31:423–49. https://doi.org/10.2307/2529430 PMID: 1174616

50. Butler DG, Cullis BR, Gilmour AR, Gogel BJ. ASReml-R reference manual. NSW Department of Primary Industries & VSN International Ltd.; 2009.

51. Mrode R, A. Linear models for the prediction of animal breeding values. 3rd ed. London, UK: Cambridge: CABI; 2014.

52. Fernando R, Garrick D. GenSel—User manual for a portfolio of genomic selection related analyses. Animal Breeding and Genetics, Iowa State University, Ames 2016 [cited 2016 01 January]. http://bigs.ansci.iastate.edu/bigsgui/lo?gin.html.

53. 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. GSE. 2014; 46(1):50.

54. Habier D, Fernando R, Kizilkaya K, Garrick D. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics. 2011; 12:186. https://doi.org/10.1186/1471-2105-12-186 PMID: 21605355

55. Garrick D, Taylor J, Fernando R. Deregressing estimated breeding values and weighting information for genomic regression analyses. GSE. 2009; 41(1):55.

56. Oraguzie NC, Wilcox PL, Rikkerink EHA, Nihal De Silva H. Linkage disequilibrium. In: Oraguzie NC, Rikkerink EHA, Gardiner SE, Nihal De Silva H, editors. Association mapping in plants. New York, USA: Springer; 2007.

57. Isik F, Whetten R, Zapata-Valenzuela J, Ogut F, McKeand S. Genomic selection in loblolly pine—from lab to field. BMC Proceedings. 2011; 5(Suppl 7):I8. https://doi.org/10.1186/1753-6561-5-S7-I8

58. 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). https://doi.org/10.1186/1471-2164-15-1048 PMID: 25442968

59. 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. 2014. https://doi.org/10.1038/hdy.2014.36 PMID: 24781808

60. Stejskal J, Lstibůrek M, Klápště J, Čepl J, El-Kassaby YA. Effect of genomic prediction on response to selection in forest tree breeding. Tree Genet Genomes. 2018; 14(5):74. https://doi.org/10.1007/s11295-018-1283-8

61. Brown G, Gill G, Kuntz R, Langley C, Neale D. Nucleotide diversity and linkage disequilibrium in loblolly pine. PNAS. 2004; 101:15255–60. https://doi.org/10.1073/pnas.0404231101 PMID: 15477602

62. Krutovsky KV, Neale DB. Nucleotide diversity and linkage disequilibrium in cold hariness- and wood quality-related candidate genes in Douglas-fir. Genetics. 2005; 171:2029–41. https://doi.org/10.1534/genetics.105.044420 PMID: 16157674

63. De La Torre AR, Birol I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJ, et al. Insights into conifer giga-genomes. Plant Physiol. 2014; 166(4):1724–32. Epub 2014/10/29. https://doi.org/10.1104/pp.114.248708 PMID: 25349325.

64. Moritsuka E, Hisataka Y, Tamura M, Uchiyama K, Watanabe A, Tsumura Y, et al. Extended linkage dis-equilibrium in noncoding regions in a conifer, *Cryptomeria japonica*. Genetics. 2012; 190(3):1145–8. https://doi.org/10.1534/genetics.111.136697 PMID: 22209904

65. Neale DB, Savolainen O. Association genetics of complex traits in conifers. Trends Plant Sci. 2004; 9(7):325–30. http://dx.doi.org/10.1016/j.tplants.2004.05.006.

66. Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, et al. Comparison of human genetic and sequence-based physical maps. Nature. 2001; 409(6822):951–3. http://www.nature.com/nature/journal/v409/n6822/suppinfo/409951a0_S1.html.

67. Snelling WM, Chiu R, Schein JE, Hobbs M, Abbey CA, Adelson DL, et al. A physical map of the bovine genome. Genome Biol. 2007; 8(8):R165. Epub 2007/08/19. https://doi.org/10.1186/gb-2007-8-8-r165 PMID: 17697342.

68. Meilan R. Floral induction in woody angiosperms. New Forests. 1997; 14(3):179–202. https://doi.org/10.1023/a:1006560603966

69. Zimmerman RH, Hackett WP, Pharis RP. Hormonal aspects of phase change and precocious flowering. In: Pharis RP, Reid DM, editors. Hormonal regulation of development III: Role of environmental factors. Berlin, Heidelberg: Springer Berlin Heidelberg; 1985. p. 79–115.

70. Chalupka W, Cecich RA. Control of the first flowering in forest trees. Scand J For Res. 2008; 12(1): 102–11.

71. USDA. Seeds of woody plants in the United States. Agriculture Handbook No. 450, 1974.

72. Cecich RA, Kang H, Chalupka W. Regulation of early flowering in *Pinus banksiana*. Tree Physiol. 1994; 14(3):275–84. PMID: 14967702

73. White TL, Adams WT, Neale DB. Chapter 17: Advanced-generation breeding strategies—breeding population size, structure and management. In: White TL, Adams WT, Neale DB, editors. Forest genet-ics. Wallingford and Cambridge: CAB International; 2007.

74. Bramlett DL, Burris LC, editors. Topworking young scions into reproductive mature loblolly pine. 23rd Southern Forest Tree Improvement Conference; 1995 20–22 June; Asheville, NC, USA.

75. Gooding GD, Bridgwater FE, Bramlett DL, Lowe WJ, editors. Top grafting loblolly pine in the western gulf. 25th South Forest Tree Improvement Conference; 1999; New Orleans, LA.

76. Resende MFR, Muñoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, et al. Accelerating the domes-tication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol. 2012; 193(3):617–24. https://doi.org/10.1111/j.1469-8137.2011.03895.x PMID: 21973055

77. White TL, Adams WT, Neale DB. Forest genetics. Cambridge, MA, USA: CAB International; 2007.

78. Chalupka W. Usefulness of hormonal stimulation in the production of genetically improved seeds. Silva Fennica. 1991; 25(4):235–40.

79. Rudolf PO. Jack pine (Pinus banksiana Lamb.). In: Fowells HA, editor. Silvics of forest trees of the United States, US Department of Agriculture, Agriculture Handbook 271. Washington, DC.1965. p. 338–48.

80. Rudolph TD. Female strobili on 12-month-old jack pine. Tree Planters' Notes. 1979; 30:24–6.

81. Almqvist C, Ekberg I. Interstock and GA4/7 effects on flowering after topgrafting in *Pinus sylvestris*. For Genet. 2002; 8(4):279–84.

82. Klocko A, Ma C, Esfandiari E, Dow M, Strauss SH, editors. Precocious flower induction by FT overex-pression in Eucalyptus. Tree Biotechnology 2013 Conference—Forest Biotechnology: Meeting the Needs of a Changing World; 2013 26 May—1 June; Asheville, NC, United States.

83. Karlgren A, Gyllenstrand N, Clapham D, Lagercrantz U. Flowering locus T/Terminal flower1-lke genes affect growth rhythm and bud set in Norway spruce. Plant Physiol. 2013; 163(2):792–803. https://doi.org/10.1104/pp.113.224139 PMID: 23958861

84. Klintenas M, Pin PA, Benlloch R, Ingvarsson PK, Nilsson O. Analysis of conifer flowering locus T/Termi-nal flower1-like genes provides evidence for dramatic biochemical evolution in the angiosperm FT line-age. New Phytol. 2012; 196(4):1260–73. https://doi.org/10.1111/j.1469-8137.2012.04332.x PMID: 23020222

85. Hazubska-Przybył T, Bojarczuk K. Tree somatic embryogenesis in science and forestry. Dendrobiology. 2016; 76:105–16.

86. Mamun NHA, Aidun CK, Egertsdotter U. Improved and synchronized maturation of Norway spruce (Picea abies (L.) H.Karst.) somatic embryos in temporary immersion bioreactors. In Vitro Cell Dev Biol Plant. 2018. https://doi.org/10.1007/s11627-018-9911-4 PMID: 30459490

87. Ratcliffe B, El-Dien OG, Klapste J, Porth I, Chen C, Jaquish B, et al. 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.