Specialty Wood Products Research Partnership

SCION forests · products · innovation

Initial evaluation of genomic selection to improve wood property in *Eucalyptus nitens* breeding population

Authors:
Jaroslav Klápště, Mari Suontama, Emily Telfer,
Natalie Graham, Charlie Low, Toby Stovold,
Russell McKinley, Heidi Dungey

Date: June 2016

Publication No: SWP-T006

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The E. nitens genetic improvement programs are predominantly based on open-pollinated progeny tests. This approach produces high levels of hidden relatedness whose ignorance causes upward bias in genetic parameters. Development of high-throughput genotyping technologies enables generation of large amounts of genomic markers that can provide information to construct matrices of realized genetic relationship. The advantage of marker based relationship matrices is to fill gaps in pairwise relatedness produced by shallow and simple pedigrees commonly present in forest tree genetic evaluations and thus reduce the standard errors of genetic parameters. This in turn results in more precise selection of valuable genotypes.

Our analysis found the marker based approach has improved the accuracy of genetic parameter estimates and also resulted in higher predictive accuracy in cross-validation evaluation. The likely source of improvement is the utilization of all the available information in the populations through complete pairwise relationship matrix compared to very sparse pedigree-based relationship matrix. This besides the faster progress in genetic improvement and delivery are a major benefits to the implementation of genomics in forest tree breeding when generally only shallow and simple pedigrees are available. The marker based approach found generally lower heritability estimates in Tinkers compared to Waiouru which is probably a consequence of a higher selection intensity applied in the Tinkers population compared to Waiouru which resulted in a fixation of part of the genetic variance. Surprisingly, in the pedigree based approach we found the opposite results in several traits such as a15, a16, a17, a39, a40, a41 and ht1 which is probably caused by the smaller sample size used to obtain reliable heritability estimates based on pedigree information. In addition, breeding values were less accurate in Tinkers compared to the Waiouru population. The across seed orchard heritability and breeding values accuracy estimates converted to intermediate values between both population estimates. Surprisingly, a larger sample size did not result in higher accuracy of genetic parameters. This could be a consequence of merging two populations with different selection histories.

We performed cross-validation at both an individual and family. The individual based cross-validation found that Tinkers population produced higher predictive ability compared to Waiouru population which is contrary to the results from heritability and theoretical accuracy estimates. The higher predictive accuracy in the Tinkers population can be explained by larger haploblocks which are built in populations created under higher selection intensity and thus the whole genetic complex can by efficiently captured even by a sparse marker array. The across population cross-validation produced again intermediate predictive accuracies between both populations (Waiouru and Tinkers) and an increase in training population sample size did not help to improve the estimates above the Tinkers population. Therefore, the decrease in effective number of genomic segments through building of larger haploblocks is more efficient than an increase in training population sample size in our population. The family based cross-validation relies purely on linkage disequilibrium between markers and QTLs which is the most stable part of genomic prediction. Generally, we can find higher predictive ability in Tinkers population which is related to the larger haploblocks from more intensive selection.

Generally, it is highly recommended to capture a large proportion of the genetic variability in training populations in order to build robust prediction models, making it important to keep a broad spectra of genetic material in training populations. Therefore, in genomics based breeding programs, the breeding arboretum should be established independently of the production population due to different requirements on genetic diversity vs. genetic gain trade-offs to utilize genomics at maximum efficiency.

INTRODUCTION

Shining gum (*Eucalyptus nitens*) is important forest tree species planted in temperate regions of Southern hemisphere mainly for pulpwood production. The *E. nitens* genetic improvement programs are predominantly based on open-pollinated progeny tests. This approach in eucalypts produces high levels of hidden relatedness due to the fact that the insects, as pollination vectors, limit the gene pool available for the next generation. The ignorance of hidden relatedness in genetic evaluations causes upward bias in genetic parameters such as additive genetic variance and heritability and also changes the ranks of breeding values (El-Kassaby, et al., 2011). Development of highly polymorphic genetic markers such as simple sequence repeats (SSRs) or Single nucleotide Polymorphisms (SNPs) allowed breeders to perform parentage assignment and recover hidden relatedness which in turn improved accuracy of both genetic parameters estimates and breeding values ranking (Doerksen, et al., 2010; El-Kassaby, et al., 2011; Telfer, et al., 2015; Vidal, et al., 2015).

Development of high-throughput genotyping technologies enables generation of large amounts of genomic markers that can provide information to construct matrices of realized genetic relationship (Nejati-Javaremi, et al., 1997; VanRaden, 2008). The advantage of marker based relationship matrices is to fill gaps in pairwise relatedness produced by shallow and simple pedigrees commonly present in forest tree genetic evaluations and thus reduce the standard errors of genetic parameters. This in turn results in more precise selection of valuable individuals (El-Kassaby, et al., 2011; Vidal, et al., 2015). The genomic prediction model generally captures three factors such as 1) share genealogy, 2) co-segregation and 3) linkage disequilibrium between markers and quantitative trait loci (QTL) and the contribution of each factor is affected by trait's genetic architecture, marker density and distribution and effective population size (Habier, et al., 2013). The accuracy of genomic prediction further depends on trait's heritability, training population size and effective number of genomic segments defined as function of a trait's genetic architecture (distribution of QTLs) and decay of linkage disequilibrium along the chromosome (Hayes, et al., 2009). The implementation of genomic prediction can accelerate genetic progress accumulated in breeding programs and its delivery into production plantation. However, it faces the same challenges as conventional breeding such as genotype x environment interaction or age x age correlations (Beaulieu, et al., 2014; El-Dien, et al., 2015; Ratcliffe, et al., 2015; Resende, et al., 2012).

The dense marker arrays also allow the fitting of both additive and non-additive genetic effects even in non-clonally propagated field experiments and provide additional information for family or clonal forestry (Gamal El-Dien, et al., 2016; Nazarian, et al., 2015). This can be done without precise knowledge of the original mating design, required in a normal quantitative-genetic pedigree-based analysis. Therefore, the implementation of genomic selection is promising for the simple reason that pedigree information can be vastly improved, especially in species with breeding programs in initial stages and shallow, simple pedigrees of only a few generations, common in most forest tree species.

The aim of our study is to investigate the potential for implementation of genomic selection in an open-pollinated progeny test of *E. nitens* which is a species with mixed mating and thus higher level of hidden relatedness and inbreeding coming from two parental populations (seed orchards) with different selection history.

METHODS

The *E. nitens* population under study was established as open-pollinated test where families were established from two independent seed orchards (Waiouru (46 OP families) and Tinkers (25 OP families)) based on different genetic resources. While the Tinkers seed orchard individuals came from forward selections in a progeny trial including material of Victorian provenance (showing the best growth (King, et al., 1988), progeny trials at Rotoaira established in 1977 from material coming from two Australian breeding programs and progeny trials based on NSW provenances, the Waiouru seed orchard was designed as a clonal archive and included 123 individuals coming from the same number of families (123).

Genomic DNA was extracted from leaf tissues of 691 individuals from progeny trial using commercial NucleoSpin Plant II kit (Machery-Nagel, Düren, Germany) (Telfer, et al., 2013) and sent to GeneSeek, Inc. (a Neogene company, Lincoln, NE, USA) for genotyping (Telfer, et al., 2015). Genotyping was undertaken using the Illumina Infinium EUChip60K SNP chip (Silva-Junior, et al., 2015) with SNP calling performed on the basis of multi-taxa and/or Maidenaria section reference. The marker data were filtered for genTrain score > 0.5, GenCall > 0.15, minor allele frequency (MAF) > 0.01, call rate > 0.6 and pairwise linkage disequilibrium in terms of a composite estimate (r^2 < 0.9).

Seven-year-old individual trees in open-pollinated progeny trial were assessed (phenotyped) for growth traits such as height, diameter, straightness and malformation in 2014 and wood quality traits such as stiffness, shrinkage and density in 2015 (Table 1) (Suontama, et al., 2016).

Table 1. List of phenotyped traits and their abbreviations.

Trait	Units	Abbreviation
Radial air-dry shrinkage 3 m log	%	a15
Radial reconditioned shrinkage 3 m log	%	a16
Tangential air-dry shrinkage 3 m log	%	a17
Tangential reconditioned shrinkage 3 m log	%	a18
Radial air-dry shrinkage 6 m log	%	a33
Radial reconditioned shrinkage 6 m log	%	a34
Tangential air-dry shrinkage 6 m log	%	a35
Tangential reconditioned shrinkage 6 m log	%	a36
Radial air-dry shrinkage average 3-6 m log	%	a39
Radial reconditioned shrinkage average 3-6 m log	%	a40
Tangential air-dry shrinkage average 3-6 m log	%	a41
Tangential reconditioned shrinkage average 3-6 m log	%	a42
Density	kg/m^3	de6
Diameter at breast height	mm	dbh6
Height	m	htm6
Straightness	score	str6
Malformation	score	mal6
Acceptability	score	ac26
Stiffness 1.4-3 m log	km/s	ht1
Siffness 3-6 m log	km/s	ht2
Growth strain 1.4-3 m log	mm	sp1
Growth strain 3-6 m log	mm	sp2

The genetic parameters were estimated using mixed linear models implemented in the ASReml-R package (Butler, et al., 2009). Two models using pedigree or marker based relationship matrix were investigated and compared. A pedigree-based model (BLUP) was used as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z_1}\boldsymbol{u} + \mathbf{Z_2}\boldsymbol{r} + \mathbf{Z_3}\boldsymbol{r}(\boldsymbol{s}) + \boldsymbol{e} \quad [1]$$

where ***y*** is vector of measurements, ***β*** is vector of fixed terms such as intercept and seed source, ***u*** is vector of additive genetic effects (breeding values) following var(*u*) ~ N(0,σ_a^2**A**), where σ_a^2 is additive genetic variance and **A** is average numerator relationship matrix (Wright, 1922), ***r*** is vector of random replication effects following var(*r*) ~ N(0,σ_r^2**I**) where σ_r^2 is replication variance and **I** is identity matrix, ***r(s)*** is set nested within replication following var(*r(s)*) ~ N(0,$\sigma_{r(s)}^2$**I**) where $\sigma_{r(s)}^2$ is set nested within replication variance, ***e*** is vector of residuals following var(*e*) ~ N(0,σ_e^2**I**), where σ_e^2 is residual variance, **X** and **Z_1, Z_2** and **Z_3** are incidence matrices assigning fixed and random effects to measurements in vector ***y***. The model accommodating marker based relationship matrix (GBLUP) is performed following equation [1] but average numerator relationship matrix **A** is substituted by marker based relationship matrix **G** which was estimated as follows:

$$\boldsymbol{G} = \frac{\boldsymbol{ZZ'}}{tr[\boldsymbol{ZZ'}]/n} \quad [2]$$

Where **Z** is **M** – **P**, **M** is marker matrix with genotypes coded 0, 1 and 2 for alternative allele homozygote, heterozygote and reference allele homozygote and **P** is vector of twice of allele frequency, tr[**ZZ**'] is trace of matrix defined in nominator and n is the number of markers (Forni, et al., 2011). The heritability represents proportion explained by genetic factors and can provide inference about potential efficiency of any improvement. It is estimated following:

$$\hat{h}^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} \quad [3]$$

where σ_a^2 is additive genetic variance and σ_e^2 is residual variance. Standard errors were estimated by Delta method approximation. The accuracy of breeding values represents correlation of their estimates obtained from model (Equation 1) with true breeding values which are commonly unknown and are estimated following:

$$r = \sqrt{1 - \frac{PEV}{G_{ii}\sigma_a^2}} \quad [4]$$

where PEV is prediction error variance (Mrode, 2014) and G_{ii} is diagonal element of realized relationship matrix for ith individual and is substituted by A_{ii} in pedigree based scenario. The 10-fold cross-validation was used as independent evaluation. The folding was performed on individual and family level and within, between and across seed sources. The predictive accuracy represents efficiency of marker based model as prediction tool to predict breeding values based on only marker information. Such scenario is representing main advantage of implementation of genomic selection in breeding programs by leaving testing phase (establishment of progeny trial) out of breeding cycle and perform selection based on only genetic markers. It was estimated as follows:

$$r_p = cor(EBV, GEBV) \quad [5]$$

where EBV is vector of breeding values estimated by pedigree based model and GEBV is vector of predicted breeding values using GBLUP model.

RESULTS

The marker data generated by using EU60K SNP chip produced solid data with very little missing data. The SNPs were called with an algorithm generic to the consensus reference across all 12 species involved in the SNP chip and also specific to *E. nitens* as reference. Both call algorithms produced a similar number of SNPs (58,307 vs. 58,323). The call rates reached 0.8 - 1 for majority of the markers (Figure 1 left). Similarly, sample call rate reached between 0.9 – 1 (Figure 1 right).

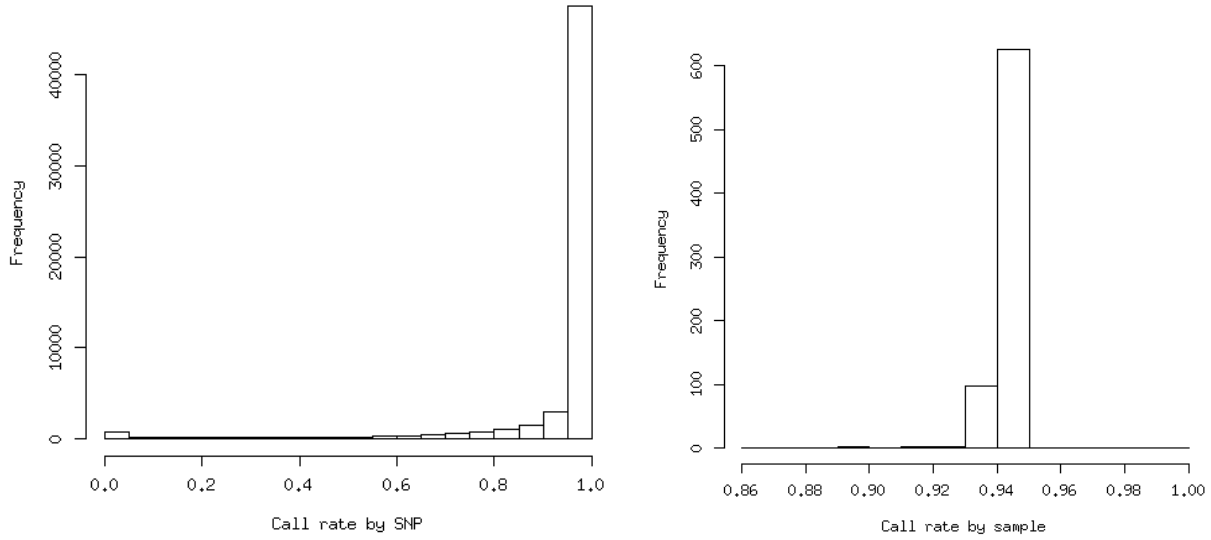

Figure 1. SNP (left plot) and sample (right plot) call rates.

Linkage disequilibrium decreased to 0.2 within 5 kb which is reflecting high genetic variability and is common to many forest tree species (Figure 2).

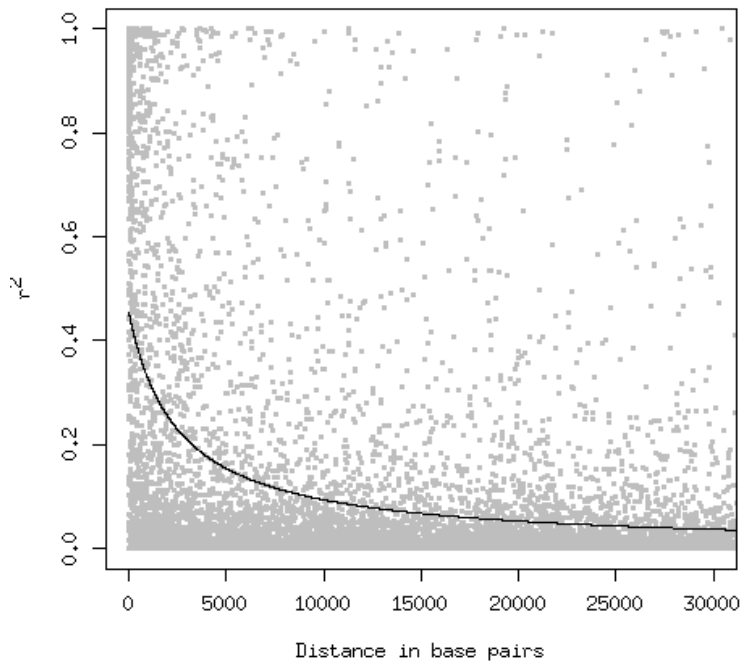

Figure 2. Linkage disequilibrium decay with physical distance in base pairs.

The SNPs were further filtered for MAF, call rate and pairwise linkage disequilibrium and only 12,236 SNPs selected were selected and used to train genomic prediction models. The marker-based analysis results in increased accuracy of breeding values and genetic parameters such as additive genetic variance and heritability. The model was tested for each seed orchard separately as well as across both seed orchards.

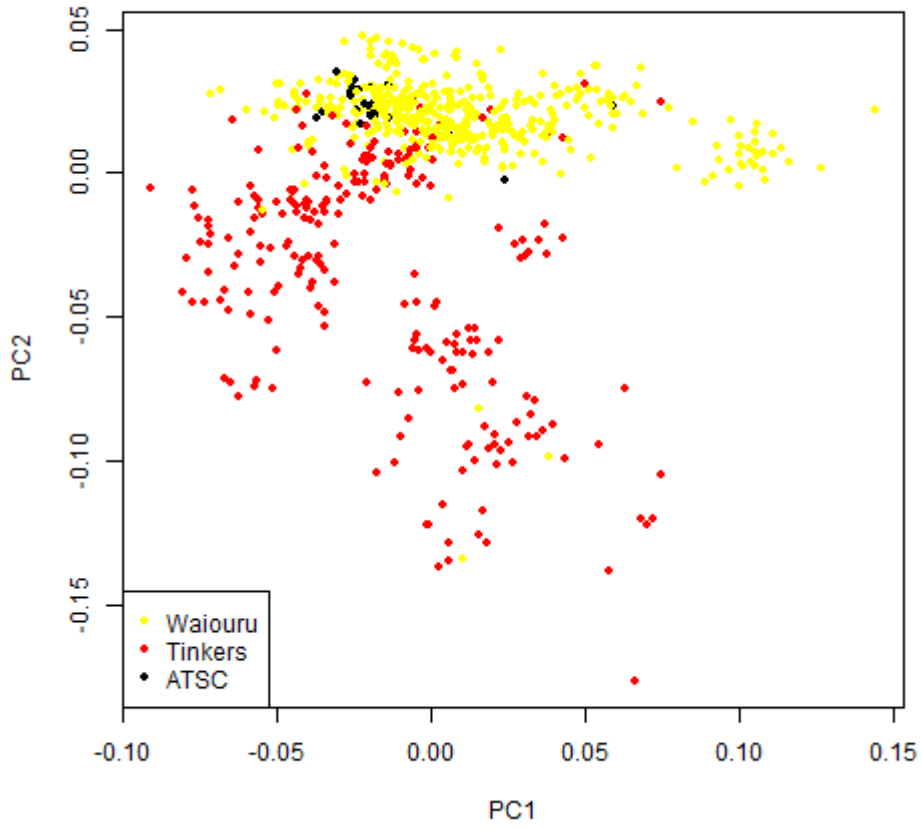

Figure 3. Spectral decomposition of genetic marker based relationship matrix.

The spectral decomposition of realized relationship matrix showed clear segregation of each seed orchard population with ATSC families in between. It reflects the difference in the genetic background due to sampling strategies.

Genetic parameter estimates

Pedigree-based analysis showed generally low heritabilities estimated in progeny from the Waiouru seed orchard population reaching from 0 to 0.29 compared with the Tinkers seed orchard population (0.01 to 0.61). However, there are several exceptions such as a35, str6, a36, a18, a42 de6 where the Waiouru seed orchard population showed much higher heritability, ranging between 0.4 – 0.54 compared with the Tinkers seed orchard population which ranged between 0.11 – 0.49. The lowest heritability estimates were found in malformation and acceptability, which are scale based subjective assessments, reaching values from 0 to 0.07. The marker-based heritability estimates generally produced lower hertiability estimates under increased accuracy (smaller standard errors) and follow similar patterns in both seed orchard populations. High discrepancies between pedigree and marker based estimates were observed in height and diameter where marker based significant heritability was observed in Waiouru but not in Tinkers while opposite pattern was obtained in pedigree based alternative. The breeding value accuracies were higher in marker based analysis reaching from 0.23 to 0.79 within both seed orchard populations (Table 2).

Across seed orchard heritability estimates were generally higher and more accurate in marker based analysis with few exceptions such as str6, a40, a34. Most traits showed statistically significant heritability estimates and only low heritable traits (heritability estimates reaching from 0.00 to 0.08) such as mal6, htm6 (sp_htm6), dbh6 (sp_dbh6) and acc6 showed statistically non-significant estimates. The malformation (mal6) and acceptability (acc6) are scale based subjective criterions which can be highly biased and thus reach very low genetic component while the wood properties in term of shrinkage seem to be measured under high precision and genetic component is mostly statistically significant. The accuracy of breeding value estimates was lower in pedigree based analysis reaching from 0 to 0.68 compared to marker based estimates reaching from 0.23 to 0.77 and was highly dependent on heritability (correlation between heritability and accuracy of breeding values was 0.95 in pedigree based and 0.97 in marker based analysis) (Table 3).

Table 2. Within seed orchard heritability, standard errors and breeding value accuracies.

Trait	Waiouru						Tinkers					
	pedigree			marker			pedigree			marker		
	h^2	se	r	h^2	se	r	h^2	se	r	h^2	se	r
a15	0.27	0.112	0.56	0.36	0.076	0.70	0.41	0.128	0.66	0.22	0.060	0.60
a16	0.26	0.111	0.55	0.26	0.070	0.63	0.32	0.119	0.60	0.27	0.064	0.64
a17	0.00	0.000	0.00	0.06	0.049	0.38	0.12	0.090	0.39	0.16	0.058	0.54
a18	0.47	0.134	0.70	0.45	0.071	0.75	0.37	0.126	0.63	0.33	0.070	0.68
a33	0.15	0.096	0.44	0.23	0.067	0.61	0.35	0.120	0.62	0.16	0.058	0.54
a34	0.13	0.095	0.41	0.25	0.069	0.62	0.61	0.150	0.79	0.20	0.063	0.58
a35	0.40	0.126	0.65	0.37	0.073	0.71	0.14	0.097	0.43	0.21	0.065	0.59
a36	0.44	0.131	0.68	0.42	0.068	0.74	0.49	0.138	0.72	0.35	0.067	0.70
a39	0.26	0.110	0.55	0.35	0.074	0.70	0.52	0.141	0.73	0.22	0.062	0.60
a40	0.25	0.110	0.54	0.33	0.073	0.68	0.58	0.146	0.77	0.24	0.063	0.62
a41	0.22	0.106	0.51	0.20	0.066	0.58	0.31	0.117	0.59	0.37	0.072	0.71
a42	0.51	0.139	0.73	0.50	0.067	0.78	0.43	0.132	0.68	0.39	0.070	0.72
de6	0.54	0.141	0.74	0.52	0.068	0.79	0.34	0.117	0.62	0.42	0.061	0.74
str6	0.41	0.130	0.66	0.19	0.063	0.57	0.11	0.089	0.37	0.12	0.059	0.48
mal6	0.09	0.092	0.35	0.00	0.000	0.23	0.01	0.081	0.10	0.03	0.040	0.30
acc6	0.13	0.092	0.40	0.02	0.042	0.27	0.07	0.083	0.30	0.05	0.045	0.35
ht1	0.28	0.112	0.56	0.34	0.072	0.69	0.35	0.120	0.62	0.24	0.060	0.62
ht2	0.19	0.100	0.48	0.30	0.071	0.66	0.15	0.099	0.44	0.18	0.057	0.56
sp1	NA	NA	NA	0.21	0.064	0.59	NA	NA	NA	0.25	0.061	0.62
sp2	0.25	0.111	0.53	0.23	0.066	0.60	0.24	0.111	0.52	0.30	0.066	0.66
sp_dbh6	0.00	0.000	0.00	0.14	0.062	0.51	0.29	0.112	0.57	0.03	0.039	0.29
sp_htm6	0.00	0.000	0.00	0.14	0.062	0.51	0.29	0.112	0.57	0.03	0.039	0.29

Cross-validation

The individual based folding showed no or very low predictive accuracy to predict breeding values of one seed orchard based on model trained in alternative seed orchard in pedigree based evaluation. The traits showing predictive ability are h1, sp1 and sp2 and achieved higher accuracy when Waiouru is used as training population to predict Tinkers population (0.1 to 0.29) compared to predictions in the opposite scenario which only achieved 0.02 to 0.14. The marker based analysis shows higher predictive accuracy between seed orchards in traits such as a18, a35, a41 and a42 reaching from 0.06 to 0.11 when Waiouru was used as training population and from 0.02 to 0.08 in the opposite scenario. Within seed orchard predictive accuracy reached, in pedigree base analysis from 0 to 0.36 in Waiouru population and from 0.13 to 0.35 in Tinkers while in marker based analysis the accuracy reached from 0.03 to 0.49 in the Waiouru population and from 0.24 to 0.52 in the Tinkers population.

Table 3. Across seed orchard heritability, standard errors and breeding values accuracy.

Trait	Across					
	pedigree			marker		
	h^2	se	r	h^2	se	r
a15	0.28	0.113	0.57	0.31	0.072	0.67
a16	0.29	0.116	0.58	0.26	0.068	0.64
a17	0.33	0.117	0.60	0.43	0.073	0.74
a18	0.44	0.131	0.68	0.47	0.072	0.76
a33	0.21	0.103	0.50	0.22	0.066	0.60
a34	0.34	0.123	0.61	0.25	0.070	0.63
a35	0.28	0.113	0.57	0.41	0.075	0.73
a36	0.42	0.129	0.67	0.42	0.069	0.73
a39	0.31	0.116	0.59	0.33	0.073	0.68
a40	0.41	0.128	0.66	0.31	0.071	0.67
a41	0.37	0.123	0.63	0.50	0.073	0.77
a42	0.44	0.131	0.68	0.49	0.070	0.77
de6	0.44	0.130	0.68	0.46	0.067	0.76
str6	0.28	0.115	0.57	0.19	0.064	0.57
mal6	0.00	0.000	0.00	0.00	0.000	0.23
acc6	0.08	0.085	0.32	0.03	0.042	0.30
ht1	0.24	0.108	0.52	0.29	0.070	0.65
ht2	0.12	0.094	0.39	0.17	0.063	0.55
sp1	NA	NA	NA	0.23	0.065	0.60
sp2	0.16	0.101	0.44	0.24	0.068	0.60
sp_dbh6	0.09	0.085	0.34	0.08	0.052	0.42
sp_htm6	0.09	0.085	0.34	0.08	0.052	0.42

Table 4. Prediction accuracy in analysis based on family folding.

Trait	Family									
	Pedigree based model					Marker based model				
	Waiouru -> Tinkers	Waiouru	Tinkers -> Waiouru	Tinkers	across	Waiouru -> Tinkers	Waiouru	Tinkers -> Waiouru	Tinkers	across
a15	NA	NA	NA	NA	NA	0.00	0.02	0.02	0.04	0.07
a16	NA	NA	NA	NA	NA	0.01	0.01	0.01	0.05	0.08
a17	NA	NA	NA	NA	NA	-0.08	0.00	0.06	0.02	0.02
a18	NA	NA	NA	NA	NA	0.05	0.03	0.07	0.03	0.09
a33	NA	NA	NA	NA	NA	0.01	0.01	-0.04	0.01	0.03
a34	NA	NA	NA	NA	NA	0.04	0.04	0.02	-0.01	0.08
a35	NA	NA	NA	NA	NA	0.09	0.03	0.06	0.02	0.06
a36	NA	NA	NA	NA	NA	NA	0.03	0.01	0.04	0.10
a39	NA	NA	NA	NA	NA	0.03	0.02	0.03	0.03	0.06
a40	NA	NA	NA	NA	NA	0.01	0.02	0.02	0.04	0.09
a41	NA	NA	NA	NA	NA	0.09	0.02	0.08	0.03	0.07
a42	NA	NA	NA	NA	NA	0.05	0.04	0.02	0.03	0.10
de6	NA	NA	NA	NA	NA	0.01	0.10	0.02	0.02	0.18
ht1	NA	NA	NA	NA	NA	0.00	0.01	0.00	0.05	0.10
ht2	NA	NA	NA	NA	NA	-0.01	0.00	-0.06	0.01	0.02
sp1	NA	NA	NA	NA	NA	0.02	-0.04	-0.04	0.02	0.00
sp2	NA	NA	NA	NA	NA	-0.01	-0.04	-0.05	0.06	0.06
dhb6	NA	NA	NA	NA	NA	-0.03	0.05	0.05	-0.01	0.08
htm6	NA	NA	NA	NA	NA	-0.01	0.05	0.06	0.00	0.07
mal6	NA	NA	NA	NA	NA	0.05	0.00	0.06	0.00	0.01
str6	NA	NA	NA	NA	NA	0.04	0.03	-0.05	0.00	0.05
acc6	NA	NA	NA	NA	NA	0.03	0.02	-0.02	-0.02	0.01

Table 5. Prediction accuracy in analysis based on individual folding.

Trait	Individuals									
	Pedigree based model					Marker based model				
	Waiouru -> Tinkers	Waiouru	Tinkers -> Waiouru	Tinkers	across	Waiouru -> Tinkers	Waiouru	Tinkers -> Waiouru	Tinkers	across
a15	0.06	0.26	0.02	0.33	0.30	0.01	0.34	0.02	0.37	0.34
a16	-0.02	0.16	-0.07	0.20	0.17	0.00	0.23	0.01	0.33	0.25
a17	-0.17	0.00	0.05	0.13	0.01	-0.09	0.03	0.05	0.25	0.04
a18	-0.01	0.26	0.01	0.20	0.26	0.08	0.34	0.09	0.37	0.34
a33	-0.02	0.20	-0.03	0.33	0.26	0.01	0.30	-0.06	0.34	0.29
a34	-0.22	0.10	-0.15	0.30	0.23	0.06	0.18	0.02	0.26	0.23
a35	0.00	0.28	0.00	0.16	0.23	0.11	0.33	0.04	0.26	0.31
a36	-0.03	0.24	-0.01	0.21	0.23	0.04	0.33	0.03	0.37	0.34
a39	-0.01	0.22	-0.03	0.35	0.29	0.03	0.33	0.00	0.36	0.34
a40	-0.07	0.12	-0.12	0.26	0.19	0.03	0.21	0.04	0.31	0.26
a41	-0.09	0.13	0.03	0.20	0.16	0.08	0.19	0.08	0.37	0.26
a42	-0.07	0.28	-0.02	0.18	0.27	0.07	0.38	0.07	0.38	0.38
de6	-0.03	0.36	-0.01	0.27	0.32	0.02	0.49	0.03	0.52	0.46
ht1	0.10	0.25	0.02	0.31	0.28	0.04	0.33	-0.02	0.38	0.33
ht2	-0.06	0.28	0.02	0.16	0.21	0.03	0.32	-0.04	0.24	0.25
sp1	0.29	0.22	0.13	0.27	0.24	-0.01	0.30	-0.01	0.36	0.30
sp2	0.23	0.27	0.14	0.19	0.24	-0.05	0.35	-0.07	0.40	0.35
dhb6	0.15	0.13	0.10	0.43	0.28	-0.03	0.26	0.04	0.13	0.22
htm6	0.19	0.13	0.10	0.42	0.27	-0.03	0.26	0.04	0.15	0.21
mal6	0.02	0.03	0.07	0.03	0.01	0.11	0.01	0.07	0.05	0.02
str6	-0.09	0.04	-0.04	-0.01	0.02	-0.01	0.05	-0.04	-0.04	0.02
acc6	0.19	0.12	0.12	0.11	0.12	0.04	0.07	-0.02	0.08	0.08

Across seed orchard predictive accuracy ranged from 0.01 to 0.32 in pedigree based and from 0.04 to 0.46 in marker based analysis.

Family based folding was investigated only in the marker-based analysis due to inability of the pedigree based scenario to predict unrelated individuals. We found higher predictive ability within the Waiouru population reaching from -0.04 to 0.1 compared to the Tinkers population where accuracies ranged from -0.01 to 0.06. Across seed orchard predictive accuracy reached from 0 to 0.18.

CONCLUSION

The forest tree breeding programs are generally at a very early stage compared to species with faster generation times which causes genetic parameter estimates to be less accurate and thus less precise selection of the best genotypes. The development and application of genetic markers improves genetic parameters through pedigree reconstruction which effectively recovers hidden relatedness and corrects pedigree errors (Doerksen, et al., 2010; El-Kassaby, et al., 2011; Telfer, et al., 2015; Vidal, et al., 2015). Such improvement in pedigree accuracies increase the precision of genetic parameters estimation and can be further explored by dense marker arrays to capture Mendelian sampling terms with construction of marker based relationship matrix (Habier, et al., 2007; Hayes, et al., 2009; VanRaden, 2008). The EUChip60K SNP chip (Silva-Junior, et al., 2015) was used in our analysis and provided solid data with minimum of missing values (Figure 1) which is desirable to perform efficient and accurate genomic evaluation. The SNP chip was designed on multiple *Eucalyptus* species which resulted in a significant reduction in the final number of SNPs used in the analysis (~13K) due to lack of polymorphism. However, this still provided a reasonable amount of genomic information with which to perform efficient genomic prediction.

Our analysis found the marker based approach has improved the accuracy of genetic parameter estimates and also resulted in higher predictive accuracy in cross-validation evaluation. The likely source of improvement is the utilization of all the available information in the populations through complete pairwise relationship matrix compared to very sparse pedigree-based relationship matrix. This besides the faster progress in genetic improvement and delivery are a major benefits to the implementation of genomics in forest tree breeding when generally only shallow and simple pedigrees are available. Such dense relationship matrices allow us to dissect not only genetic and environment components more precisely but also additive and non-additive genetic components by construction of non-additive relationship matrices in simple experimental designs where pedigree based analysis does not allow to analyse it (Gamal El-Dien, et al., 2016). Such an option is attractive in species with efficient vegetative propagation when no clonal replications are involved in testing. Also, such a scenario would favour the application of genomics since the number of genotyped individuals in a given training population is more important than phenotyping fidelity when predicting genomic breeding values. The marker based approach found generally lower heritability estimates in Tinkers compared to Waiouru which is probably a consequence of a higher selection intensity applied in the Tinkers population compared to Waiouru which resulted in a fixation of part of the genetic variance. Surprisingly, in the pedigree based approach we found the opposite results in several traits such as a15, a16, a17, a39, a40, a41 and ht1 which is probably caused by the smaller sample size used to obtain reliable heritability estimates based on pedigree information. In addition, breeding values were less accurate in Tinkers compared to the Waiouru population (Table 2). The across seed orchard heritability and breeding values accuracy estimates converted to intermediate values between both population estimates. Surprisingly, a larger sample size did not result in higher accuracy of genetic parameters. This could be a consequence of merging two populations with different selection histories (Table 3).

Genomic prediction relies on three factors: shared genealogy, co-segregation and linkage disequilibrium between markers and quantitative trait loci (QTL) (Habier, et al., 2013). We performed cross-validation at both an individual and family level to dissect the effects of these three factors. The individual based cross-validation captures all of the effects and we found that Tinkers population produced higher predictive ability compared to Waiouru population which is contrary to the results from heritability and theoretical accuracy estimates. The higher predictive accuracy in the Tinkers population can be explained by larger haploblocks which are built in populations created under higher selection intensity and thus the whole genetic complex can by efficiently captured even by a sparse marker array (Ødegård, et al., 2014). However, transferability of such a prediction model is highly reduced and can be seen in the cross-validation between seed populations where the Waiouru training population produces a slightly higher predictive ability but the effect is limited through very small sample size in training population. The across population cross-validation produced again intermediate predictive accuracies between both populations

(Waiouru and Tinkers) and an increase in training population sample size did not help to improve the estimates above the Tinkers population. Therefore, the decrease in effective number of genomic segments through building of larger haploblocks is more efficient than an increase in training population sample size in our population. However, the infusion of new genetic material or changes in selection can cause changes in the haploblocks and thus decrease in prediction accuracy (Table 5).

The family based cross-validation relies purely on linkage disequilibrium between markers and QTLs which is the most stable part of genomic prediction. Generally, we can find higher predictive ability in Tinkers population which is related to the larger haploblocks from more intensive selection. There are a few exceptions with the most obvious contradiction in de6 which could be caused by a fixation of many causal variants through more intensive selection which also resulted in decreased heritability estimate (Table 2). The across population family based cross-validation resulted in higher predictive ability compared to single population evaluation which can be result of larger training population sample size but also capturing historical relatedness as the two populations are clearly split into two clusters (Table 5, Figure 3).

Generally, it is highly recommended to capture a large proportion of the genetic variability in training populations in order to build robust prediction models, making it important to keep a broad spectra of genetic material in training populations. Therefore, in genomics based breeding programs, the breeding arboretum should be established independently of the production population due to different requirements on genetic diversity vs. genetic gain trade-offs to utilize genomics at maximum efficiency.

ACKNOWLEDGEMENTS

We would like to thank to Barbara Geddes and Liz Cunningham for participation in DNA preparation and extraction and Mark Miller and Kane Fleet for help in field operation, assessment and phenotyping.

REFERENCE

Beaulieu, J., Doerksen, T. K., MacKay, J., Rainville, A., & Bousquet, J. (2014). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC genomics, 15*(1), 1.

Butler, D. G., Cullis, B. R., Gilmour, A. R., & Gogel, B. J. (2009). ASReml-R reference manual. *Queensland Department of Primary Industries, Queensland, Australia*.

Doerksen, T. K., & Herbinger, C. M. (2010). Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. *Tree Genetics & Genomes, 6*(4), 591-600.

El-Dien, O. G., Ratcliffe, B., Klápště, J., Chen, C., Porth, I., & El-Kassaby, Y. A. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC genomics, 16*(1), 370.

El-Kassaby, Y. A., Cappa, E. P., Liewlaksaneeyanawin, C., Klápště, J., & Lstibůrek, M. (2011). Breeding without breeding: is a complete pedigree necessary for efficient breeding. *PLoS One, 6*(10), e25737.

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

Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., & El-Kassaby, Y. A. (2016). Implementation of the Realized Genomic Relationship Matrix to Open-Pollinated White Spruce Family Testing for Disentangling Additive from Non-additive Genetic Effects. *G3: Genes|Genomes|Genetics, 6*(3), 743-753. doi:10.1534/g3.115.025957

Habier, D., Fernando, R., & Dekkers, J. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics, 177*(4), 2389-2397.
Habier, D., Fernando, R. L., & Garrick, D. J. (2013). Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics, 194*(3), 597-607. doi:10.1534/genetics.113.152207
Hayes, B. J., Visscher, P. M., & Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research, 91*(01), 47-60.
King, J., & Wilcox, M. (1988). Family tests as a basis for the genetic improvement of *Eucalyptus nitens* in New Zealand. *New Zealand Journal of Forestry Science, 18*(3), 253-266.
Mrode, R. A. (2014). *Linear models for the prediction of animal breeding values*. Cabi.
Nazarian, A., & Gezan, S. A. (2015). Integrating Nonadditive Genomic Relationship Matrices into the Study of Genetic Architecture of Complex Traits. *Journal of Heredity*, esv096.
Nejati-Javaremi, A., Smith, C., & Gibson, J. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of animal science, 75*(7), 1738-1745.
Ødegård, J., & Meuwissen, T. H. (2014). Identity-by-descent genomic selection using selective and sparse genotyping. *Genet. Sel. Evol, 46*(3).
Ratcliffe, B., El-Dien, O., Klápště, J., Porth, I., Chen, C., Jaquish, B., & El-Kassaby, Y. (2015). A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity, 115*, 547-555.
Resende, M., Munoz, P., Acosta, J., Peter, G., Davis, J., Grattapaglia, D., Resende, M., & Kirst, M. (2012). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist, 193*(3), 617-624.
Silva-Junior, O. B., Faria, D. A., & Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species. *New Phytologist, 206*(4), 1527-1540.
Suontama, M., Stovold, G. T., McKinley, R., Miller, M., Fleet, K., Low, C., & Dungey, H. S. (2016). Selection for solid wood properties in *Eucalyptus nitens. SWP Final Report*.
Telfer, E., Graham, N., Stanbra, L., Manley, T., & Wilcox, P. (2013). Extraction of high purity genomic DNA from pine for use in a high-throughput Genotyping Platform. *New Zealand Journal of Forestry Science, 43*(1), 1-8.
Telfer, E. J., Stovold, G. T., Li, Y., Silva-Junior, O. B., Grattapaglia, D. G., & Dungey, H. S. (2015). Parentage Reconstruction in Eucalyptus nitens Using SNPs and Microsatellite Markers: A Comparative Analysis of Marker Data Power and Robustness. *PloS one, 10*(7).
VanRaden, P. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science, 91*(11), 4414-4423.
Vidal, M., Plomion, C., Harvengt, L., Raffin, A., Boury, C., & Bouffier, L. (2015). Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. *Tree Genetics & Genomes, 11*(5), 1-13.
Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist*, 330-338.