

GENOMIC SELECTION IN DOUGLAS-FIR

by

Frances Rebecca Thistlethwaite

B.Sc., Cardiff University, 2011

M.Sc., The University of Edinburgh, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2019

© Frances Rebecca Thistlethwaite, 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

GENOMIC SELECTION IN DOUGLAS-FIR

submitted by Frances Rebecca Thistlethwaite in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Forestry

Examining Committee:

Dr. Yousry El-Kassaby, Forestry

Supervisor

Dr. Michael Stoehr, BC MFLNRO

Supervisory Committee Member

Dr. Richard Hamelin, Forestry

Supervisory Committee Member

Dr. Tongli Wang, Forestry

University Examiner

Dr. Quentin Cronk, Botany

University Examiner

Additional Supervisory Committee Members:

Dr. Jaroslav Klápště, New Zealand Forest Research Institute Ltd., Scion

Supervisory Committee Member

Supervisory Committee Member

Abstract

Conventional tree breeding productivity (especially in conifers) is primarily constrained by late expression of commercially important traits, and late onset of sexual maturity. These characteristics conjointly correspond to lengthy testing phases prior to selection, which in turn restrains the accumulation of genetic gain. GS has the potential to increase genetic gain per unit time by allowing for the prediction and selection of traits at an earlier age.

This dissertation investigates some facets of GS, specifically in relation to Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)), and ‘real-world’ applications. Expressly: to compare pedigree based ABLUP and two GS methods; assess GS prediction accuracy over spatial and temporal deviations; validate the use of exome capture as a cost-effective genotyping platform for use in GS; use of GS to predict breeding values in the next generation; investigate the effect of relatedness on GS prediction accuracy; and to assess the number of makers required for GS in conifers.

Chapter 2 utilizes exome capture as a genotyping platform to assess height and wood density GS prediction accuracies across space and time. A cross-generational GS analysis was performed in Chapter 3 using a progeny generation as an independent validation set. Chapter 4 investigates the effect of marker density on GS predictive accuracy in Douglas-fir and Interior spruce (*Picea glauca* (Moench) Voss x *Picea engelmannii* Parry ex Engelm.).

The overriding conclusion, is that while some of the GS models’ prediction accuracies were high, the main driving force was the tracking of relatedness rather than LD. However with regard to the ability of the available markers to track pedigree, exome capture was found to be very competent. Knowing this, the following trends were observed: GS models performed similarly and were comparable to ABLUP; genotype x environment interactions are an important consideration for GS spatial analyses; height at 12 years was deemed an acceptable age at which accurate predictions can be made concerning future height and wood density; moderate to high cross-generational GS prediction accuracies were obtained, but were influenced by the relationship between training and validation sets; and increasing marker number increases GS prediction accuracy, for Douglas-fir and Interior spruce.

Lay Summary

Traditionally, forest tree breeding is a protracted endeavor. It can take decades to complete a selection cycle in which desired characteristics (phenotypes), are enhanced. Genomic selection (GS) is a DNA based method which allows the prediction of phenotypes, without waiting for the trees to grow. Essentially speeding up the traditional method, which requires phenotypic selection, by selecting based on DNA markers instead.

The genetic material (genomes) of conifers are very large and complex and this makes the application of GS difficult. In this dissertation are studies which explore GS and its application to conifer breeding. Including: comparing two methods of GS and a traditional approach; assessing GS efficiency over spatial and temporal deviations; validating the use of a cost-effective genotyping method; use of GS to predict phenotypes in the next generation; investigating the effect of relatedness on GS; and assessing the number of DNA makers required for GS in conifers.

Preface

This work represents original research by F. Thistlethwaite. I conceived of the research questions and objectives through much consultation of my supervisory committee. I collaborated with colleagues in the data collection stages. I performed the data analysis, interpretation of results, and prepared the manuscripts. Chapters of this dissertation have been, published or are under review by peer-reviewed journals, listed below:

• Chapter 2

Thistlethwaite, F.R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M.U., and El-Kassaby, Y.A. (2017). Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genomics* 18, 930.

• Chapter 3

Thistlethwaite, F.R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M.U., and El-Kassaby, Y.A. (2019). Genomic selection of juvenile height across a single-generational gap in Douglas-fir. *Heredity*, DOI: 10.1038/s41437-018-0172-0

• Chapter 4

Thistlethwaite, F.R., Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M.U., Ingvarsson, P.K., El-Kassaby, Y.A. (2019). Linkage disequilibrium vs. pedigree: genomic selection prediction accuracy in two conifer species. (Submitted).

Table of Contents

Abstract.....	iii
Lay Summary	iv
Preface.....	v
Table of Contents	vi
List of Tables	x
List of Figures.....	xi
List of Abbreviations	xiii
Acknowledgements	xv
1 Introduction.....	1
1.1 Traditional Selection Methods and Statistical Analysis.....	1
1.2 Marker Based Selection Methods	3
1.3 Genomic Selection Validation and Predictive Accuracy	5
1.4 Genomic Selection Statistical Approaches	7
1.5 Principal Work in Forestry.....	8
1.6 Research Objectives.....	10
1.7 Dissertation overview	11
1.7.1 Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform.....	11
1.7.2 Genomic selection of juvenile height across a single generational gap in Douglas-fir	11
1.7.3 Linkage disequilibrium vs. pedigree: genomic selection prediction accuracy in two conifer species.....	11

2	Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform.....	15
2.1	Introduction.....	15
2.2	Materials and Methods.....	17
2.2.1	Experimental population.....	17
2.2.2	Tissue sampling, DNA extraction and genotyping.....	18
2.2.3	Phenotyping.....	19
2.2.4	Estimated breeding values (EBVs) and deregressed breeding values (DEBVs).....	19
2.2.5	Cross-validation across time and space.....	20
2.2.6	Genomic selection analysis.....	21
2.2.7	Genomic estimated breeding values (GEBVs) and genomic estimated deregressed breeding values (GDEBVs).....	21
2.3	Results.....	23
2.3.1	Traits' heritabilities and EBV accuracy.....	23
2.3.2	Cross-validation across space and time.....	23
2.4	Discussion.....	28
2.4.1	Exome capture.....	28
2.4.2	Heritability.....	29
2.4.3	Genomic selection.....	30
2.4.4	Single and multi-site cross-validation.....	33
2.4.5	Cross-site validation.....	34
2.4.6	Time- time and trait- trait correlations.....	35

2.4.7	Genomic selection and forest tree breeding	36
2.5	Summary	37
3	Genomic selection of juvenile height across a single generational gap in Douglas-fir.....	46
3.1	Introduction.....	46
3.2	Materials and Methods.....	49
3.2.1	Experimental population	49
3.2.2	Phenotyping, deregression, tissue sampling, DNA extraction and genotyping	50
3.2.3	Estimated breeding value (EBV) prediction and accuracy	52
3.2.4	Genomic selection analysis.....	54
3.3	Results.....	56
3.3.1	Heritability and EBV accuracy	56
3.3.2	Validation in the progeny generation.....	57
3.4	Discussion.....	61
3.4.1	Heritability of juvenile height.....	61
3.4.2	Pedigree vs. marker based models	61
3.4.3	GS across generations	63
3.4.4	Main factors that affect GS: relatedness and LD	67
3.5	Summary	69
4	Linkage disequilibrium vs. pedigree: genomic selection prediction accuracy in two conifer species.....	76
4.1	Introduction.....	76
4.2	Material and Methods	79

4.2.1	Experimental populations	79
4.2.2	Phenotyping and genotyping.....	80
4.2.3	Estimated breeding value (EBV) calculation and ABLUP accuracy	80
4.2.4	Genomic selection and Cross-validation.....	82
4.2.5	Effective population size estimation	83
4.2.6	Random marker sampling	83
4.3	Results.....	83
4.3.1	Marker Number Effect.....	83
4.3.2	Pedigree and Relatedness Effect	84
4.4	Discussion	84
4.5	Summary	88
5	Conclusion	90
5.1	Research novelties and applications	90
5.2	Conclusions regarding research objectives	91
5.3	Future research directions	94
5.4	Strengths and limitations.....	95
	References.....	97

List of Tables

Table 2.1 Within single-site prediction accuracies and their standard errors for ABLUP and genomic selection (ridge regression (RR-BLUP) and generalized ridge regression (GRR)) models for EBVs and GEBVs of heights (HT12 and HT35) and wood density (WD_{res}).....	39
Table 2.2 Within multi-site genomic selection prediction accuracies and their standard errors for ridge regression (RR-BLUP) and generalized ridge regression (GRR) models), estimating GEBVs and GDEBVs for heights (HT12 and HT35) and wood density (WD_{res}).....	39
Table 2.3 Genomic selection prediction accuracies for time- time correlations for HT12 for RR-BLUP and GRR models (standard errors). Prediction accuracies based on correlations between GEBVs at age 12 and EBVs at age 38; and correlations between GDEBVs at age 12 and DEBVs at age 38. Traits key: HT35 = height at 35 years (cm); WD_{res} = wood density calculated from resistance drilling to drilling (g/cm^3); HT12 = height at 12 years (cm).....	40
Table 3.1 Summary of the number of individuals per site, and to which models they contributed.....	70
Table 3.2 Genomic selection analyses of four models using three GS statistical methods (RR-BLUP, GRR, and Bayes-B).....	71
Table 3.3 Corresponding predictive abilities for GS analyses in table 3.2, calculated as the correlation between the raw phenotype (juvenile height: HTJ) and their genomic estimated breeding values (GEBVs) or deregressed genomic estimated breeding values (GDEBVs).....	73

List of Figures

Figure 1.1 Diagram showing traditional tree improvement pathway, including selection, seed orchard establishment and forestation (Neale and Kremer, 2011).....	13
Figure 1.2 Visualization of metadata concerning height prediction accuracy, from various sources of GS studies in forestry of conifer species. Filled points represent studies using full-sib populations; empty points represent studies using half-sib populations; empty points with crosses represent studies using full and half-sib populations. Point diameter is a function of sample size. Marker density was calculated based on genetic map lengths estimated in the following studies: <i>Pinus taeda</i> (Eckert et al., 2010), <i>Picea glauca</i> (Pavy et al., 2017), <i>Picea glauca</i> x <i>engelmannii</i> (based on white spruce data) (Pavy et al., 2008, 2017; Pelgas et al., 2011), <i>Pinus pinaster</i> (Ritter et al., 2002), <i>Picea mariana</i> (Kang et al., 2010), <i>Picea abies</i> (Acheré et al., 2004).	14
Figure 2.1 Distribution of estimated breeding values (EBVs) and deregressed estimated breeding values (DEBVs) for (a) height at 12 years (cm), (b) height at 35 years (cm), and (c) wood density (g/cm^3) calculated from resistance to drilling.	41
Figure 2.2 Heritabilities and GS prediction accuracies for models trained on EBVs and predicting GEBVs for each of the traits. Showing the results of within-site, cross-site, combined-sites, and multi-site to single-site cross-validation (Top (a), using RR-BLUP, Bottom (b), using GRR). The direction of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance drilling to drilling (g/cm^3). Sites are Adam, Fleet River, Lost Creek, and multi/combined-site (ALL).....	42
Figure 2.3 Heritabilities and prediction accuracies of the ABLUP model for each of the traits. Showing the results of within-site, cross-site, combined-sites, and multi-site to single-site cross-validation. The direction	

of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance drilling to drilling (g/cm^3). Sites are Adam, Fleet River, Lost Creek, and multi/combined-site (ALL).....43

Figure 2.4 Heritabilities and GS prediction accuracies for models trained on EBVs and predicting GEBVs for each of the traits. Showing the results of two sites predicting one site cross-validation (Top (a) using RR-BLUP, Bottom (b), using GRR). The direction of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance to drilling (g/cm^3). Sites are Adam, Fleet River, Lost Creek.....44

Figure 2.5 Heritabilities and prediction accuracies of the ABLUP model for each of the traits. Showing the results of two sites predicting one site cross-validation. The direction of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance to drilling (g/cm^3). Sites are Adam, Fleet River, and Lost Creek.....45

Figure 3.1 GS correlation in the validation set (Correlation in the validation set of: EBVs and genomic estimated breeding values (GEBVs) for juvenile height (HTJ) using a) RR-BLUP, b) GRR, and c) Bayes-B; deregressed estimated breeding values (DEBVs) and genomic deregressed estimated breeding values (GDEBVs) using d) RR-BLUP, e) GRR, and f) Bayes-B; EBVs for HTJ vs. GEBVs for mature height age 35 years (HT35) using g) RR-BLUP, h) GRR, and i) Bayes-B; and DEBVs for HTJ vs. DGEBVs for HT35 using j) RR-BLUP, k) GRR, and l) Bayes-B)..... 75

Figure 4.1 Effect of marker density on RR-BLUP prediction accuracy for multi-site data for height and wood resistance, with ABLUP for comparison, for: a) Douglas-fir (HT at 35 years, WD_{res} at 38 years), and b) Interior spruce (HT at 38 years, WD_{x-ray} at 38 years) 89

List of Abbreviations

ABLUP	additive best linear unbiased predictor
BLUE	best linear unbiased estimator
BLUP	best linear unbiased predictor
BV	breeding value
DArT	diversity array technology
DBH	diameter at breast height
DEBV	deregressed estimated breeding value
EBV	estimated breeding value
GDEBV	genomic deregressed estimated breeding value
GEBV	genomic estimated breeding value
GRR	generalised ridge regression
GS	genomic selection
GxE	genotype by environment
HT	height
HT12	height at 12 years
HT35	height at 35 years
LD	linkage disequilibrium

MAF	minor allele frequency
MAS	marker assisted selection
MM	mixed model
N_e	effective population size
OLS	ordinary least squares
QTL	quantitative trait loci
REML	restricted maximum likelihood
RR-BLUP	ridge regression best linear unbiased predictor
SNP	single-nucleotide polymorphism
WD_{res}	wood density measured via Resistograph®
WD_{x-ray}	wood density measured via X-ray densitometry

Acknowledgements

Thank you to everyone who made this journey possible, and especially my supervisor Dr. Yousry El-Kassaby. You have been dazzlingly knowledgeable, endlessly patient and overwhelmingly kind.

Thank you to my esteemed lab mates:

Susan Song, Blaise Ratcliffe, Dr. Omnia Gamal El-Dien Ibrahim, Dr. Yang Liu, and Dr. Faisal Alharbi.

Thank you to my mentors:

Dr. Ilga Porth, Dr. Jaroslav Klápště, Dr. Michael Stoehr, and Dr. Charles Chen

Thank you Mum and Dad, thank you family

Thank you friends and teammates

Thank you Lucy

1 Introduction

1.1 Traditional Selection Methods and Statistical Analysis

Traditionally, tree improvement practices follow a three phase structure: breeding, testing, and selection. The aim is to produce trees with enhanced economically desirable traits. To do this the existing natural genetic variation is captured and favorably enhanced through various principles of selection and recombination. Not only is commercial success a concern for breeders but ultimately, the continued ability to produce high quality stock is just as important. To do this genetic diversity must be maintained within the population in order that adaptability is not impaired (Zobel and Talbert, 1984).

Tree improvement programs for commercial forest trees were first adopted in the latter half of the 20th century, as a direct consequence of the success of animal and crop breeders. Up until this point there was much debate about the role and importance of genetics versus environment in tree improvement (Sharma, 2000). Prior to this a simple form of mass selection was adopted but breeders quickly found that although initial gains were good (20-30%) (Sharma, 2000), the gains were not as large as predicted in subsequent generations largely due to a lack of planning in the selection process. Early breeding strategies focused on 'plus tree' selection on phenotypic attributes, progeny testing and following this the establishment of seed orchards and seed production areas (Figure 1.1 (Neale and Kremer, 2011)) (Sharma, 2000).

Since the inception of tree improvement programs there has been an accelerated improvement in techniques required for success. Namely in the application of statistical analysis of genetic improvement and experimental design aspects. These success have been largely driven by progress made by animal breeders and their early uptake of such procedures (Goddard et al., 2010; VanRaden et al., 2009). The aim of most traditional programs is to ascertain population wide estimates of genetic attributes: heritability, trait variation (both phenotypic and genetic), genetic by environmental interactions, as well as multi-trait and

juvenile-mature correlations (White et al., 2007). This is achieved through progeny testing of families with structured mating designs. The progeny are planted according to an experimental design (incorporating randomization and replication) from which statistical analysis can provide these estimates of genetic parameters (Falconer and Mackay, 2009; White et al., 2007). From this information, selection decisions for valuable traits, and approximate estimates of gains over time can be made. In addition, high performing individuals (forward selection) or their parents (backward based selection) may be selected based on their breeding values and incorporated into advanced generations. Breeding values (BV) are a measure of deviation from a population mean of zero (White et al., 2007). It is the average of heritable allelic effects, generally progeny gain one half of the breeding value of each parent (Falconer and Mackay, 2009).

In 1953 mixed model analysis (MM) proved to be one of the greatest advancements in the field of quantitative genetics. It was developed by an animal breeder, Charles Henderson, from Cornell University (Henderson, 1953). Previously ordinary least squares analysis was used, but experimenters found there to be various inadequacies with this method and the underlying assumptions. Whilst both methods aim to partition phenotypic variation into genetic and environmental effects using linear models (Mrode, 2014), only MM allows for both fixed and random terms to be incorporated into the analysis. Genetic effects could now be treated as random terms. In addition MM is more suited to the data sets available, i.e. unbalanced and containing multiple generations with complex relatedness (White et al., 2007). The treatment of genetic effects as random allows for the prediction of these effects to be based on best linear unbiased predictions (BLUP). Meanwhile the fixed elements are estimated using generalized least squares method, giving best linear unbiased estimates (BLUE) (McCulloch and Searle, 2001). The iterative method of restricted maximum likelihood (REML) is used to produce the estimates of variance components according to the 'best fit' of the information that are weighted according to the quality and quantity of observations (White et al., 2007). It is a restricted version of maximum likelihood (ML) devised by Patterson and Thompson (1971, 1974). It differs from ML in that it accounts for degrees of freedom used in estimating the fixed effects and is thus less biased. This is computationally intensive and various software resources are

available which implement these methods in a user-friendly way, a popular tool being ASReml-R (Butler et al., 2017).

ABLUP was another major development for breeding programs. The introduction of the animal model devised by Henderson (1976) used information from all relatives to produce BLUP breeding values. An additive relationship (A) matrix is produced from pedigree information and this is used as a design matrix for the random terms, also allowing for correlated random effects (Henderson, 1976; Thompson, 2008). The A matrix describes the extent to which individuals in the analysis are expected to be genetically related to each other (additively). This method (ABLUP) is still widely used today in tree breeding programs where pedigrees are maintained. An extension of this was the introduction of a realized relationship matrix (G), in which the values of the design matrix are based on observed allele frequencies and captures deviations away from the average expected values of A (which occur due to random segregation). G matrices represent the proportion of the genotyped genome identical by descent between pairs (Hayes et al., 2009a). Hayes et al. (2009a) concluded that using G increases the accuracy of artificial selection since BLUP estimated BVs using G are more accurate, and that this method can even predict BVs for individuals with no recorded phenotype. Obtaining an accurate assessment of relatedness is especially important in tree breeding where open-pollination is common and regularly results in unintended crosses (e.g. selfs, half-sibs in a full sib mating design and vice versa).

1.2 Marker Based Selection Methods

Genomic selection (GS) enables for complex quantitative traits to be selected using genomic marker data alone. The method does not require a priori knowledge regarding the specific genetic architecture of the trait in question. Instead, markers at regular intervals throughout the entire genome (or in this case exome) are incorporated into the estimate. Their effect sizes being simultaneously estimated (Meuwissen et al., 2001).

The adoption of genomic methodologies into animal breeding programs has been successful in garnering substantial gains in important economic traits. Though to date the focus of such methods has been

to engineer those alleles with large effect sizes in linkage with a known marker (e.g., Zikhali et al. (2014)). As it stands, those more complex traits still rely largely on field testing for selection decisions (Edwards and Page, 1994; Moreau et al., 2004). Marker assisted selection (MAS) has been attempted in forestry in the past, with the aspiration of replicating the success in animal breeding.

Underlying MAS, is the assumption that the chosen markers are in linkage disequilibrium (LD) with alleles which have a known effect with regards to the trait in question (Lorenz et al., 2011). Whilst this is discernible for large-effect genes, the significance of smaller-effect genes is harder to detect statistically and therefore many go undetected (Edwards and Page, 1994; Lorenz et al., 2011; Moreau et al., 2004). Herein lies the issue for most forest tree complex traits, whose genetic architecture is more in keeping with the infinitesimal model (Hill et al., 2008). Whereby many loci contribute small effects to the overall phenotype (Cappa et al., 2013; Hill et al., 2008). In addition to this MAS becomes ineffectual when traits are susceptible to influences from the environment (e.g. genotype x environment interactions) (Moreau et al., 2004; Resende et al., 2012a), and from the effects of recombination on marker-allele LD. This again is especially pertinent to forest trees, which have observed low levels of LD (Neale and Kremer, 2011), it would require a large number of markers to find causative SNPs (Jaramillo-Correa et al., 2015).

It was especially the need for statistical methodologies that could detect small-effect loci which became the driving force behind genomic selection (GS). Since this was a major obstacle for MAS in forest tree breeding: because there were very few markers with significant effects (White et al., 2007). Methods that could accurately generate genomic predictions for quantitative traits. It was Meuwissen et al. (2001) who proposed a solution to this problem by estimating all marker loci effects simultaneously (without prior knowledge of genetic architecture or assumptions), and it was they who coined the term 'genomic selection'. This includes all loci effects previously considered insignificant by MAS (and the more markers throughout the genome, the more loci effects that can be estimated). In this way GS has the potential to capture more variance due to the large number of markers used (Resende et al., 2012a).

GS for use on forest trees was initially investigated by Grattapaglia and Resende (2011). They used simulations of deterministic equations to assess the effect of linkage disequilibrium, training population

size, trait heritability, and QTL number on the predictive accuracy of GS. Their results were positive and recommended further experimentation on the matter. Since then some proof of concept investigations have been able to realize this initial success. Notably Resende et al. (2012b) produced data which concluded that indeed GS could be used to generate increased genetic gain per generation in comparison with pedigree based selection, and that there was no significant difference in the predictive accuracy between different methods of GS (Resende et al., 2012c).

The process of GS resembles the following basic procedure: To begin with, random samples from the population(s) under investigation are chosen and these represent the training set. This training set is then both phenotyped and genotyped. The data is in turn incorporated into statistical model, in which alleles at all loci have their effects simultaneously estimated. Once a suitable model has been designed based on the training population, genomic estimated breeding values (GEBVs) can be calculated for the selection candidates (other individuals that are genotyped only) (Grattapaglia, 2014; Meuwissen et al., 2001). With regards to this project, the statistical analyses will be cross validated within the training data and also have a separate validation population. The outputted GEBVs provided a basis, upon which selection decisions are made. The repercussion of this, is a paradigm shift in which the model unit of these breeding analyses shifts from being the line of descent to the allele. The upshot being that GS could potentially bypass the long testing phase currently required in forest tree breeding for the estimation of BVs via traditional means. Resulting in greater genetic gain per unit time.

1.3 Genomic Selection Validation and Predictive Accuracy

Predictive accuracy refers to the ability of the GS model to accurately assign breeding values to (genotyped) individuals in the validation set. These are genomic estimated breeding values (GEBVs). It is the correlation between the GEBVs and experimentally derived estimated breeding values (EBVs). Validation can be done in a number of ways, most commonly in tree breeding k-fold cross validation is used. Where the data is assigned into k number of subsets or folds, the model is trained on k-1 folds and the remaining fold is used as a validation set. The process is looped until all individuals have a GEBV which are correlated with their

EBVs to assess predictive accuracy (Refaeilzadeh et al., 2009; Resende et al., 2012b). Another option is to use a completely independent validation population. One in which the individuals are not related to the training set. This is useful in ascertaining the wider applicability of the model i.e. across populations, sites, generations etc.

Grattapaglia (2014) describes four features which both independently and collectively influence the success of GS as measured by the predictive accuracy. 1) The extent of LD between markers and QTL and number of markers, 2) Training population size, 3) trait heritability, 4) genetic architecture of trait (number of loci and effect size). Effective population size is also an important influencing factor as pointed out by Lorenz et al. (2011). The first two factors and effective population size can be controlled by the experimenter. Grattapaglia and Resende (2011) tested the importance of these factors in determining an optimum experimental design. Firstly they demonstrated that the effective population size and the number of markers used, together dictate the extent of LD between markers and QTL. Along with established theory, the lower the effective population size the stronger the effects of genetic drift which increases non-random association (LD) of markers with trait QTLs so subsequently fewer markers are required. However in more outbred populations (with higher effective population sizes), LD is balanced out by recombination and so LD is gradually lost over generations. In fact this is one of the limitations, when GS models which are trained on older generations are used to predict GEBVs of advanced generations, predictive accuracy will be lost since ideally the estimate is largely based on marker-QTL LD. Though as Gianola et al. (2009) and Habier et al. (2007) ascertained, some degree of familial relationship is also captured during the process, inflating predictive accuracy. Yet it is the portion of accuracy due to LD that will be most informative. Since recombination events should occur at a lower rate between markers and QTL in LD, it (the portion of accuracy due to LD) is more persistent over generations than co-segregation information from kinship (Habier et al., 2013). Given this, models must be validated and updated according to the distance between generations.

Two things can be garnered from the above: effective population size must dictate the scale of marker density; and LD can be controlled through choosing an effective population size (Grattapaglia,

2014). In their simulations Grattapaglia and Resende (2011) estimated that for an effective population of 60, 2-3 markers/cM would produce results in the range of traditional selection methods. This interaction is thought to have the most impact on GS in trees (Grattapaglia, 2014). For effective populations approaching 100, the density required is more like 10-20 markers/cM. Due to the large size of the Douglas fir genome this would require between 20000 and 50000 informative markers (Grattapaglia, 2014).

Increasing the training population size increased accuracy up to a point. Beyond 1000 individuals the gains were insignificant. Also it was noted that increasing the training population size negated, somewhat, the effects of low trait heritability (Meuwissen et al., 2001). Which in turn had little effect on predictive accuracy when traits were more complex (50-100+ QTL) (Grattapaglia and Resende, 2011).

1.4 Genomic Selection Statistical Approaches

The desired outcome of GS is to produce unbiased marker effect estimates (Meuwissen et al., 2001). This is in contrast to MAS where markers that are deemed insignificant (according to an arbitrary threshold) are set to have zero effect, thus leaving the significant marker effects to be overestimated (Lande and Thompson, 1990). This was coined the Beavis effect (Xu, 2003). In order to reduce this bias, it was necessary to estimate all marker effects instead of selecting markers based on a significance threshold. In effect this caused a “large p, small n” problem, in which there are more predictor effects (p, markers in this case) to be estimated than there are observations (n, samples) (Lorenz et al., 2011). Least squares cannot be used to estimate all the effects at once since there are not enough degrees of freedom. In addition multicollinearity of markers would cause any model to be over fitted (Lorenz et al., 2011).

To address the issue, various statistical models have been proposed. They generally fall into the following categories: shrinkage models, variable selection models, and kernel methods. Shrinkage models (e.g. ridge regression BLUP [RR-BLUP], (Whittaker et al., 2000)) fit all marker effects which are all shrunk to the same degree. With RR-BLUP, it is assumed that the trait in question more closely resembles the infinitesimal model proposed by Fisher (1918) (many loci with small effects), and marker effects are samples from a normal distribution (with equal variance). Variable selection (e.g. generalized ridge-

regression [GRR], (Shen et al., 2013)) however reduces the number of markers used, and in doing so assumes that the trait is controlled by fewer, strong effect loci (Lorenz et al., 2011). Kernel methods convert the predictor variables to distances in effect creating a matrix similar to an additive genetic matrix. The kernel matrix quantifies the distance between individuals but also smoothing parameters are added (Lorenz et al., 2011). This method is flexible and can incorporate complex relationships between markers, for this reason it is useful in cases where non-additive effects are suspected to occur (Gianola et al., 2006). Since different traits have differing genetic architecture, there is no one model that is necessarily the best for all traits or populations (Lorenz et al., 2011).

Berger et al. (2015) compared the effectiveness of shrinkage and variable selection models and indeed found that variable selection models gave higher predictive accuracies in less complex traits. But also found that in highly complex traits both models performed poorly. Thus various models should be tested for any new data set to find the most appropriate. Habier et al. (2007) also note that with shrinkage models the predictive accuracy is determined largely by the relatedness between individuals since all markers are used. However the success of variable selection models is largely based on marker-QTL LD, which is less bias.

1.5 Principal Work in Forestry

Initially due to the success of GS applied to animal breeding, deterministic simulations of GS in forest trees were carried out by Grattapaglia and Resende (2011) to assess potential gains. Due to the positive feedback from these simulations, empirical studies were carried out to see if GS had real-world applications to forest tree breeding. Predictive accuracy was used as a means of determining successful models, and a range of results were uncovered.

Work on Eucalyptus by Resende et al. (2012a) found GS to be at least as accurate as traditional methods with predictive accuracies ranging from 0.55 to 0.88. They focused on two separate populations with $N_e=11$ and 51, and based predictions on 3000 DArT markers. Their findings also indicated that initially population specific prediction models will be the primary source of significant gain, due to the

effects of differing population structure and its effects on LD. The differing genetic architecture between populations combined with GxE interactions means that cross-population predictive accuracy is low. This is something to bear in mind when applying GS to real-world data. Work by Zapata-Valenzuela et al. (2012) agreed with this finding, genomic prediction accuracy (which ranged from 0.30 to 0.83) based on 3406 SNPs, was in line with pedigree based predictions for wood quality traits in a cloned population of *Pinus taeda* L. (training population: N=149, total population: N=526). But they warned that the accuracy was likely more to do with tracing kinship rather than historical marker-QTL LD.

Within-population analysis was carried out by Resende et al. (2012b), they attempted to discern what accuracy is maintained when models are used across environments and ages. A population of clonally replicated *Pinus taeda* L. (N \approx 800) grown on four separate sites was used. Predictive accuracies based on 4825 SNP markers ranged from 0.63 to 0.75. Although the authors found that cross-site validation predictive accuracy was lower than within-site validation (due to GxE interactions), they showed that models retained some accuracy so long as they were applied to sites within the same breeding zone. Accuracies decreased with increasing distance between sites. However with regards to models used across ages the predictive accuracy was low. Suggesting that models trained on young trees cannot necessarily predict future phenotypes of those same individuals. This study reported a selection efficiency estimate per unit time of 53-112%, higher than traditional methods.

In a study which compared GS methods, Resende et al. (2012c) found there to be little difference between the predictive abilities of shrinkage and variable selection methods (4 in total) even considering they have different a priori assumptions. They used a *Pinus taeda* training population with 951 individuals genotyped for 4853 SNPs. Predictive accuracies for 17 traits (including growth, disease resistance and development) ranged from 0.17 to 0.51. Only one trait (fusiform rust resistance) showed any significant difference between the models. Higher predictive accuracies were obtained using variable selection methods for this trait. This reflects the genetic architecture of the trait which is controlled by few, large effect loci (Resende et al., 2012c).

Since these initial findings, there have been a number of more recent empirical studies concerning conifer species (Fig. 1.2) specifically; loblolly pine (*Pinus taeda* L.) (Munoz et al., 2014; Zapata-Valenzuela et al., 2013), white spruce (*Picea glauca*) (Beaulieu et al., 2014a, 2014b), interior spruce (*Picea glauca x engelmannii*) (Gamal El-Dien et al., 2015; Ratcliffe et al., 2015), maritime pine (*Pinus pinaster* Aiton) (Bartholomé et al., 2016; Isik et al., 2016), black spruce (*Picea mariana* (Mill.) BSP) (Lenz et al., 2017), and Norway spruce (*Picea abies* (L.) Karst.) (Chen et al., 2018). The most commonly assessed trait in forestry GS is height, therefore Fig. 1.2 summarizes the metadata from these previous studies, as well as Thistlethwaite et al. (2017), in terms of height prediction accuracy.

1.6 Research Objectives

The main objectives of this dissertation are:

- To compare ABLUP and two GS methods; ridge regression best linear unbiased predictor (RR-BLUP) and generalised ridge regression (GRR) (Chapter 2).
- Assay GS prediction accuracy across environments/ spatial divisions and time (Chapter 2).
- Review the potential of exome capture for use as a genotyping platform in GS studies concerning conifer species (Chapters 2, 3, and 4).
- Determine intergenerational prediction accuracy using progeny as a validation population (Chapter 3).
- Investigate the effect of kinship on GS prediction accuracy (Chapter 3).
- Study the effect of marker density on GS prediction accuracy in two conifer species (Chapter 4).

1.7 Dissertation overview

The identified objectives were investigated in the following chapters:

1.7.1 Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform

In this chapter, the study objectives were to: 1) compare two GS methods i) ridge regression best linear unbiased predictor (RR-BLUP) and ii) generalised ridge regression (GRR), in relation to their predictive accuracies; and 2) to test the GS prediction accuracy for within-, cross-, pooled multi-site, and time- time (age-age /trait-trait) between age 12 and 38 years. Results of ABLUP cross-validations were used as a reference for comparison. In addition, the potential of exome capture as a genotyping platform for conifers was reviewed. Exomic information was collected from 1372, 38-year-old coastal Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)) trees. The samples represented 37 full-sib families with replications over three sites in coastal BC. This chapter presents the first use of exome capture as a genotyping platform in a large-scale genomic tree improvement study.

1.7.2 Genomic selection of juvenile height across a single generational gap in Douglas-fir

Here a cross-generational GS analysis was performed on coastal Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)), in an attempt to gain insight into how GS would perform in a more realistic breeding application. GS predictive models were trained on parental generation (F_1 , $n = 1321$) data, and validated using the progeny generation (F_2 , $n = 136$) with shared pedigree. A total of 69,551 SNPs were used in the GS analyses to produce a predictive model for juvenile height, which was compared to pedigree based (ABLUP) predictions. In addition, EBVs were deregressed in order to disentangle LD from pedigree effects.

1.7.3 Linkage disequilibrium vs. pedigree: genomic selection prediction accuracy in two conifer species

This chapter explores the effect of marker density on GS prediction accuracy in two conifer species with differing pedigree structures. A random sampling method (repeated 10 times) was used to choose SNP sets

with totals in the range of 200- 50,000, from total SNP ‘pools’ (Douglas-fir = 56,454; Interior spruce = 62,190). A 10-fold cross-validation process using RR-BLUP was performed on each SNP set for the traits height (age 35) and wood density (age 38). This chapter explicitly explores the idea that in species with very large genomes (i.e. conifers), the role of short-range marker-QTL LD on GS prediction accuracy is imperceptible compared to the effect of relatedness.

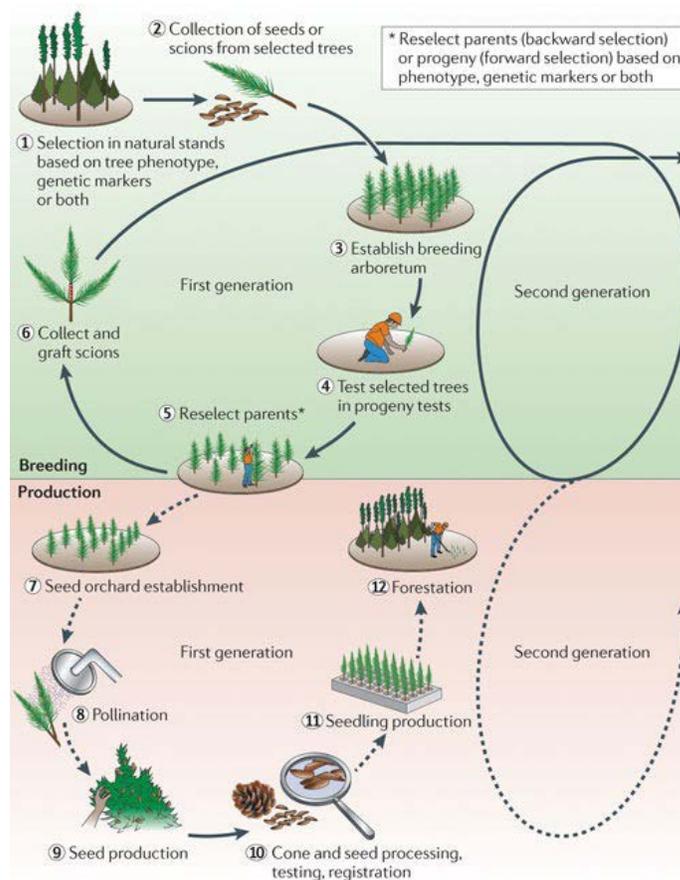


Figure 1.1 Diagram showing traditional tree improvement pathway, including selection, seed orchard establishment and forestation (Neale and Kremer, 2011).

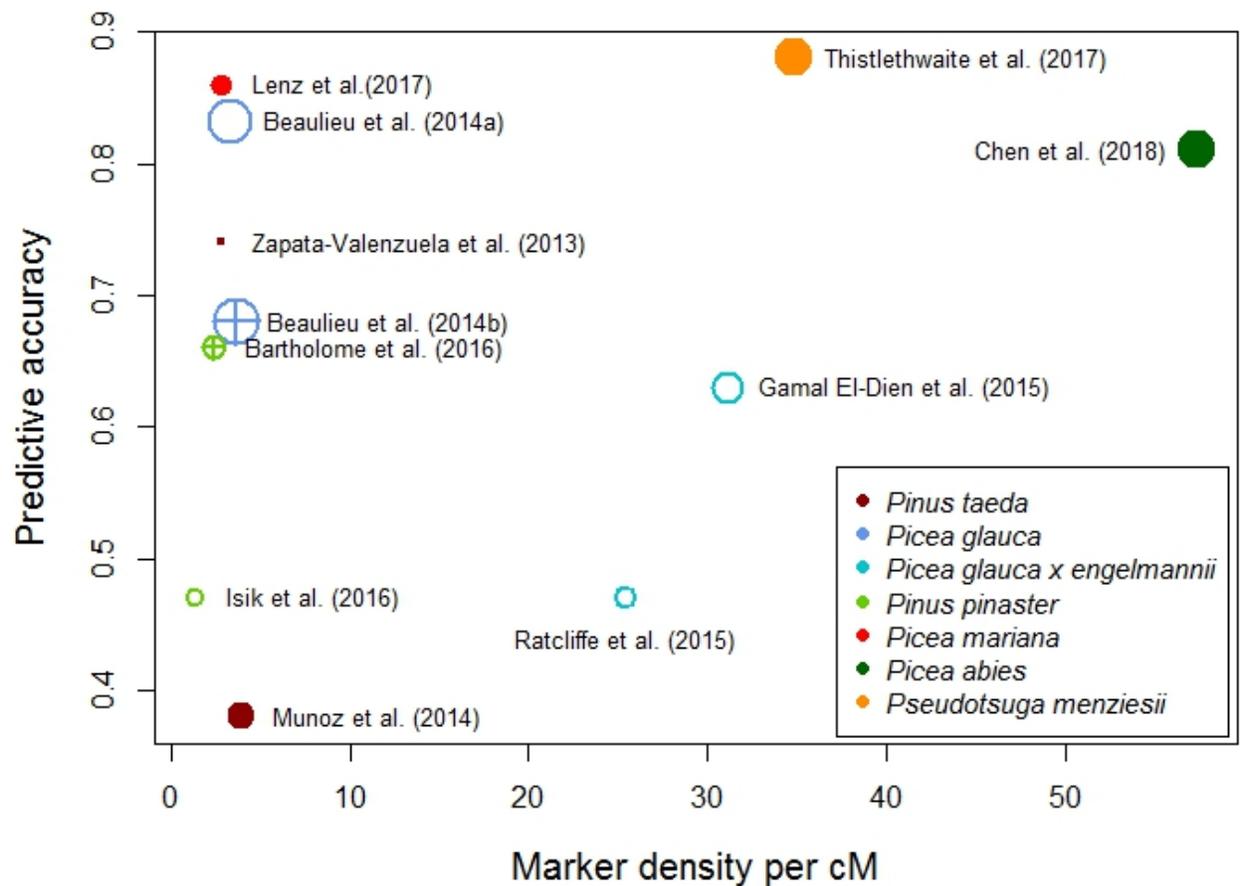


Figure 1.2 Visualization of metadata concerning height prediction accuracy, from various sources of GS studies in forestry of conifer species. Filled points represent studies using full-sib populations; empty points represent studies using half-sib populations; empty points with crosses represent studies using full and half-sib populations. Point diameter is a function of sample size. Marker density was calculated based on genetic map lengths estimated in the following studies: *Pinus taeda* (Eckert et al., 2010), *Picea glauca* (Pavy et al., 2017), *Picea glauca x engelmannii* (based on white spruce data) (Pavy et al., 2008, 2017; Pelgas et al., 2011), *Pinus pinaster* (Ritter et al., 2002), *Picea mariana* (Kang et al., 2010), *Picea abies* (Acheré et al., 2004).

2 Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform

2.1 Introduction

Novel advancements in genomics technologies and statistical genetics, have paved the way for an increasingly prosperous environment for breeding industries. Notably in the dairy sector, the traditional phenotype-dependent selection has been successfully replaced by genotype-dependent selection (aka genomic selection, GS) which reduces the time required for evaluating genetic traits (Hayes et al., 2009b). Tree selective breeding (aka tree improvement) challenges are somewhat similar to those of the livestock industry; namely, long generation times and low heritability and late expressing traits. GS, if successful, can offer unprecedented gains in forestry through reducing trait evaluation time, allowing faster breeding generations turn-over, and hence, an increased genetic gain per unit time can be reached. Additionally, the implementation of GS to forestry would offer a certain resilience to spontaneous market or environmental, and/or extraneous influences (e.g., disease/pest resistance, climate change); as breeding programmes adapt accordingly in a shorter time frame (Grattapaglia, 2014).

Selective breeding has traditionally focused on phenotypic selection, and more recently the linkage disequilibrium (LD) based indirect method of marker-assisted selection (MAS) (Lorenz et al., 2011). MAS is suited for major-gene traits (such as resistance to white-pine blister rust (*Cronartium ribicola*)) and proven to be less effective for predicting complex quantitative traits that closely reflect Fisher's infinitesimal model (Fisher, 1919). MAS models could not effectively describe a complex trait, since small effect loci are not readily discovered and considered 'not significant', thus leaving a substantial proportion of genetic control unaccounted for (i.e., missing heritability). Meuwissen et al. (2001) proposed the method of 'genomic selection' to alleviate this problem by simultaneously considering all marker effects, and in doing so, all genetic contributions are captured regardless of size (significance). GS enables complex quantitative trait selection using genomic marker data alone and the method does not require

a priori knowledge regarding the specific genetic architecture of the trait in question. Instead, markers throughout the entire genome (or in this case exome) are incorporated into the estimate. The resulting genomic estimated breeding values (GEBVs) for each individual derived from the GS models provide a basis, upon which selection decisions are made. The effect of this is a paradigm shift, in which the model unit of these breeding analyses shifts from being the *line of descent* to the *allele*. This means that the phenotypic values of individuals are determined from genotypic data, enabling early selection of traits, leading to a significantly shorter breeding cycle and higher selection differential, particularly for the “difficult to assess” attributes.

The feasibility of applying GS to forest trees was initially assessed through deterministic simulations and the results indicated that GS has the potential to radically improve the efficiency of tree selective breeding (Grattapaglia and Resende, 2011). Grattapaglia and Resende (2011) also recommended further experimentation and proof of concept investigations. Since then, several forest tree species “proof-of-concept investigations” have been conducted with encouraging results (Bartholomé et al., 2016; Beaulieu et al., 2014a, 2014b; Fuentes-Utrilla et al., 2017; Gamal El-Dien et al., 2015; Isik et al., 2016; Müller et al., 2017a; Ratcliffe et al., 2015; Resende et al., 2012a, 2012b, 2012c, 2017).

Neves et al. (2013) recognized that one of the major barriers to the application of genomic technologies to tree selective breeding was the large size of the genome of many forest species. This is particularly true of conifers, and although conifers have a large genome (Neale and Kremer, 2011; Neale et al., 2017; Suren et al., 2016), their transcriptomes are comparable with other plants such as *Arabidopsis* whose genome is more than 100 times smaller (Nystedt et al., 2013). Therefore, as an alternative to the prohibitive costs in both monetary terms and time, and the complexity of sequencing the whole genome, Neves et al. (2013) focused on the coding region. This they refer to as “sequence capture”, and proposed that it would enable more efficient genetic variant identification in conifers; as it had previously been done in human and maize genomic studies (Fu et al., 2010; Gnirke et al., 2009; Walsh et al., 2010). Although sources of variation may not be exclusively found in the exome, the reduced cost and time compared to that of whole genome sequencing, as well as the ability to still capture a significant proportion of variants and

rare variants, makes this method desirable as it is harder to find functional variants in non-coding regions (Kiezun et al., 2012; Mertes et al., 2011). Exome capture has been recognized as an effective method for capturing rare variants in the field of medicine (Choi et al., 2009), and for increasing knowledge of unmapped large genomes (Mertes et al., 2011; Neves et al., 2013). In addition, Suren et al. (2016) have shown it to be a cost effective method for reducing complexity in large genomes, such as those of conifers. The present GS study was based on the exomic information collected from 1372, 38-year-old coastal Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)) trees. The samples represented 37 full-sib families with replications over three sites in coastal BC. The study objectives were: 1) to compare two GS methods; ridge regression best linear unbiased predictor (RR-BLUP) and generalised ridge regression (GRR), and 2) to test the GS prediction accuracy for within-, cross-, pooled multi-site, and time- time (age-age /trait-trait) between age 12 and 38 years. Two phases to this analysis were carried out, firstly the two GS models were trained on estimated breeding values (EBVs). This represents an analysis in which model predictions are based on pedigree (both historical and contemporary) and marker-QTL LD information. Secondly the models were trained on deregressed breeding values (DEBVs). In this analysis the pedigree information (parental average) is removed, resulting in model predictions based on marker-QTL LD and co-segregation. Results of ABLUP cross-validations are provided as a reference for comparison.

2.2 Materials and Methods

2.2.1 Experimental population

Predictive models were trained on a replicated 38-year-old pedigreed coastal Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)) progeny testing population. The trial was established by the Ministry of Forests, Lands and Natural Resource Operations of British Columbia, Canada in 1975. The breeding scheme and mating design was a partial, disconnected diallel cross; using 6 parents per diallel, with no selfs. The experimental population consists of 9 diallels, 165 full-sib families (54 parents), of which 37 families were selected for sampling from three test sites (**Adams** (Lat. 50 24'42" N, Long. 126 09' 37"W, Elev. 576 mas), **Fleet River** (Lat. 48 39'25" N, Long. 128 05'05" W, Elev. 561 mas), and **Lost Creek** (Lat. 49 22'15"

N, Long. 122 14'07" W, Elev. 424 mas)) with a total of 1372 trees ($N \approx 500$ per site) and effective population size (N_e) of 21.

2.2.2 Tissue sampling, DNA extraction and genotyping

Cambial tissue was collected using a hammer and hollow punch tool (approx. 2 cm diameter) to remove two small circular pieces of bark/ cambium and developing tissue from each tree. The cambium disks were separated from the bark layer and immediately placed in a 2 ml collection tube with 1 ml storage buffer (10 mM EDTA pH 8.0, 10 mM Na_2SO_3) and kept at 4 °C until DNA extraction. DNA extraction followed the modified procedure developed by (Ivanova et al., 2008) (R. Whetten, unpublished, North Carolina State University, personal communications). Genotyping was done using the exome capture method in a commercial facility (RAPiD Genomics©, Florida, US). A total of 40 K probes were designed using the available Douglas-fir transcriptome assembly (Howe et al., 2013). From the Douglas-fir reference transcriptome, a total of 325,372 non-overlapping 120-bp probes were initially designed. After filtering out redundant and organelle matching probes, this number was reduced to 117,135 probes. Of the remaining probes, we selected 7464 that contained 17,096 SNPs previously reported (Howe et al., 2013). A further 32,536 probes were selected bringing the total to 40 K. Selection was performed by randomly sampling the remaining probes restricting the selection to a maximum of two probes per transcript. These 40 K (7464 + 32,536) probes cover a total of 21,187 transcripts. The raw sequenced reads were demultiplexed in each individual barcodes. Low quality bases with less than 20 quality score in the 3' end were trimmed out followed by a low quality filter that removed reads with more than 10% of the read with less than 20 quality score. The filtered reads were aligned against the reference transcriptome using Mosaik v2.2.3 (Lee et al., 2014) with the following parameters -mmp 0.05 -m all -a all -hs 15. SNP markers were identified at a population level using Freebayes (Garrison and Marth, 2012) without considering indels, multi-nucleotide polymorphisms and complex events. This analysis resulted in approximately 550,000 SNPs. These SNPs were filtered to identify the highest quality sites. These included only biallelic SNPs with less than 40% missing data. Further filtering was applied so that data was located on contigs with a mean read depth less

than or equal to 60. Additional filters included minor allele frequency (MAF) >5%, Hardy-Weinberg disequilibrium cut-off $P \leq 0.05$, and maximum site depth < 60. This process resulted in a total of 1372 samples with 69,551 SNPs for use in the study, and mean imputation was used for the missing data. (for more details on exome capture see, (Neves et al., 2013)).

2.2.3 Phenotyping

Early- (1988) and mid- (2011) rotation growth trait measurements of the studied trees had been assessed for height (HT: in meters) and mid-rotation (age 38 years, 2014) wood density (WD_{res}) was assessed indirectly using the average resistance measurements recorded with a Resistograph® (Instrumenta Mechanik Labor, Germany). Resistograph readings were subsequently converted to wood density indices (g/cm^3) by scaling them by the DBH measurements, as performed by (El-Kassaby et al., 2011).

2.2.4 Estimated breeding values (EBVs) and deregressed breeding values (DEBVs)

EBVs were fitted in ASReml 3.0 (Gilmour et al., 2009), using the following mixed linear model for a single site:

$$y = X\beta + Za + e \quad (1)$$

Where y is the phenotypic trait measurement, β is a fixed effect vector (overall mean), a is a random effects vector (additive genetic) which are normally distributed ($\sim N(0, A \sigma_a^2)$) where A is average numerator relationship matrix and σ_a^2 is additive genetic variance, e is the random residual effects which are normally distributed ($\sim N(0, I \sigma_e^2)$) where I is identity matrix and σ_e^2 is residual variance. X and Z are incidence matrices relating the fixed and random effects to the observations. Since GxE plays an important role in forestry (Cappa et al., 2016), the combined site EBVs were estimated with terms accounting for site, replication, and family structure (an additive model with site by family effect). Thus minimising biases in BV calculation caused by environmental variations between and within sites, and full-sib genetic effects.

Narrow-sense heritability was calculated as $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$, where σ_a^2 and σ_e^2 are the variances of additive genetic and residual effects, respectively. Combined site heritability estimates included the GxE model term in the denominator. The breeding values (\hat{a}) are fitted using BLUP as follows:

$$\hat{a} = \mathbf{AZ}' \sigma_a^2 \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}) \quad (2)$$

Where \mathbf{V} is the variance- covariance matrix of \mathbf{y} obtained by:

$$\mathbf{V} = \mathbf{ZAZ}' \sigma_a^2 + \mathbf{I} \sigma_e^2 \quad (3)$$

Breeding values were deregressed using the deregression procedure of Garrick et al. (2009). This adjusts the BV data to account for family means, resulting in DEBVs that contain information regarding individuals only, without parental BV influence.

The theoretical accuracy of the EBVs (\hat{r}) was calculated following the procedure of Dutkowski et al. (2002).

$$\hat{r} = \sqrt{1 - \frac{SE_i^2}{(1+F_i)\hat{\sigma}_a^2}} \quad (4)$$

Where SE_i is the standard error of breeding value, and F_i is the inbreeding coefficient of the i^{th} individual.

2.2.5 Cross-validation across time and space

The two GS models (RR-BLUP and GRR) were compared to assess their application to various traits, and in addition they were cross-validated in order to assess their prediction accuracy across environments/ spatial divisions and time. The intention was to give an indication to which extent (if any) GS increases gain per unit time over traditional methods.

The validation processes used was a replicated randomly assigned 10-fold cross validation. Models were trained on 9/10 of these folds, with the remaining 1/10 fold used as a validation set. Prediction accuracy was determined as the mean of the replications of the Pearson product-moment correlation between EBVs of the validation set and their predicted GEBVs. Or in the case where DEBVs were used to

train the models, the correlation between DEBVs and predicted GDEBVs (genomic estimated deregressed breeding values). The following combinations were used: 1) within-site, using information from a single site to estimate the GEBVs within that same site, 2) Cross-site/ between sites in all combinations, with information from single sites used to predict other sites, 3) Combined sites, pooled information from all the sites, 4) Multi-site to single site, the pooled information from all three sites used in the estimation of single sites only, 5) Two sites to predict one site, pooled information from two sites used in the estimation of the remaining site only, and 6) Time-time (age-age)/ trait-trait, using information from individual trees to obtain correlations between GEBVs at age 12 and EBVs at ages 35 for height and 38 for wood density. In addition, the same 10-fold cross-validation process was used to assess the predictive accuracy of the ABLUP model for all the spatial cross-validation analyses for all the attributes (HT12, HT35, and WDres) in the study. In this case predictions made were based on relatedness as given by the pedigree.

2.2.6 Genomic selection analysis

Two GS methods were compared: ridge regression (RR-BLUP) and generalized ridge regression (GRR) (Lorenz et al., 2011). The performance of each of the methods was assessed according to their predictive accuracy determined by the correlation between GEBVs and EBVs, or DEBVs and GDEBVs.

2.2.7 Genomic estimated breeding values (GEBVs) and genomic estimated deregressed breeding values (GDEBVs)

RR-BLUP

Ridge regression (RR-BLUP: (Whittaker et al., 2000)) was proposed for use as a selection tool based on marker information. The model was fitted using the R package ‘RRBLUP’ (Endelman, 2011), the GEBV (or GDEBV if using deregressed BVs) is obtained by the sum of p marker effects:

$$g(x_i) = \sum_{k=1}^p x_{ik}\beta_k \quad (5)$$

Where x_{ik} is the score (genotype) of SNP k in individual i , β_k is the marker effect of k . This method assumes that the marker effects are normally distributed (mean = 0) and so the BLUP solutions for marker effects

are derived from solving mixed linear model equations of (Henderson, 1976) so that λ is optimised in the following Eq.:

$$\hat{\beta} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1} \mathbf{Z}'\mathbf{y} \quad (6)$$

Where \mathbf{Z} is an incidence matrix relating markers to individuals, \mathbf{I} is an identity matrix, and \mathbf{y} is the vector of EBVs fitted in ASReml. The shrinkage parameter (λ) can be written as $\lambda = \sigma_e^2 / \sigma_{\beta}^2$ (residual variance / common marker effect variance). Since marker effects are assumed to be identically distributed, all effects are shrunk equally towards zero. This method is equivalent to using lines (as opposed to markers) as random effects in a mixed model analysis, where the covariance is modelled by a kinship matrix calculated from the marker data (a genomic relationship matrix [\mathbf{G} matrix]) (this is sometimes referred to as GBLUP).

GRR

Generalized ridge regression (GRR) is a two-step variable selection method, the first step obtains estimates in the same way that RR-BLUP does using linear mixed model analysis to solve for optimum λ , and in the second step the BLUP for $\hat{\beta}$ is subjected to an alternative shrinkage parameter which is marker specific, using GRR to solve a heterogeneous error model which replaces $\mathbf{I}\lambda$ in (5) with $diag(\lambda)$:

$$\hat{\beta} = (\mathbf{Z}'\mathbf{Z} + diag(\lambda))^{-1} \mathbf{Z}'\mathbf{y} \quad (7)$$

In this case λ is a vector of p shrinkage parameters. For the k^{th} element: $\lambda_k = \widehat{\sigma}_e^2 / \widehat{\sigma}_{\beta k}^2$, is the parameter, where $\widehat{\sigma}_{\beta k}^2$ is the variance of marker effect k ($\widehat{\sigma}_{\beta k}^2 = \hat{\beta}_k^2 / (1 - h_{kk})$). $\hat{\beta}$ is from step 1 (the BLUP marker effect) and h_{kk} is the influence of the dependant variable on the fitted value for observation k . In other words, h_{kk} represents the diagonal element ($n + k$) of the influence matrix $\mathbf{H} = \mathbf{T}(\mathbf{T}'\mathbf{T})^{-1} \mathbf{T}'$, and:

$$\mathbf{T} = \begin{pmatrix} \mathbf{X} & \mathbf{Z} \\ 0 & diag(\lambda) \end{pmatrix} \quad (8)$$

2.3 Results

To assess the studied attributes' variation, boxplots were produced showing the variance of the estimated (EBVs) and the deregressed (DEBVs) breeding values (Fig. 2.1). It is interesting to note that the deregression process maintained the within family variation for the three studied attributes (HT12, HT35, and WD_{res}); however, it virtually eliminated among family variation resulting in similar family means (Fig. 2.1).

2.3.1 Traits' heritabilities and EBV accuracy

Pedigree-based relationship matrix (ABLUP) heights heritability estimates varied among sites ranging between 0.13 (Lost Creek) and 0.24 (Adam), and 0.05 (Adam) and 0.23 (Fleet) for age 12 and 35 years, respectively (Fig. 2.2). Multi-site height heritability estimates were similar to the average of the single-site estimates at age 12 (0.17 vs. 0.19); and was slightly higher than the average single site estimate by age 35 (0.17 vs. 0.14) (Fig. 2.2). Pedigree-based relationship wood density heritability estimates generally were higher than those obtained for height and substantially varied among sites (range: 0.22 (Lost Creek) and 0.45 (Fleet)), with higher multi- than single-site average estimates (0.43 vs. 0.37) (Fig. 2.2). The average theoretical accuracies for the EBVs were: 0.61, 0.68, and 0.76 for HT12, HT35, and WD_{res} respectively.

2.3.2 Cross-validation across space and time

2.3.2.1 Within-site prediction accuracy

Within-site prediction accuracies, determined based on the correlations between EBVs and GEBVs for the two genomic selection (RR-BLUP and GRR) models, generally produced very similar results (correlations and standard errors) across EBV traits and sites (Table 2.1). For all EBV traits, Adam produced the highest prediction accuracies, with Fleet River second, and Lost Creek always producing the lowest accuracies. The two models, RR-BLUP and GRR, produced almost identical results in analysed EBV traits. The only differences occurring in the prediction of GEBV for HT35, in which the GRR method produced slightly lower prediction accuracies than RR-BLUP (0.84 vs. 0.86, and 0.80 vs. 0.82) for Fleet River and Lost

Creek, respectively. The same was found in the analysis of WD_{res} for Adam (GRR = 0.86, RR-BLUP = 0.87). In contrast the WD_{res} , GRR results for Lost Creek (0.84) were slightly more accurate than in the RR-BLUP model (0.79). It is worthy to mention that the overall predictive accuracy of the studied genomic selection models was generally high across sites and EBV traits (RR-BLUP: average = 0.86 and range of 0.79–0.91 and GRR: average = 0.86 and range of 0.80–0.91) (Table 2.1). Generally, all prediction accuracies' standard error estimates were small reflecting good model fit.

Using the deregressed breeding values to train the two GS models, we obtained predictive accuracy results approximating 0.0 for WD_{res} for within-site cross-validation (RR-BLUP: Adam = -0.10 ± 0.055 , Fleet River = -0.04 ± 0.046 , and Lost Creek = -0.06 ± 0.049 ; GRR: Adam = -0.05 ± 0.074 , Fleet River = -0.03 ± 0.045 , and Lost Creek = -0.04 ± 0.046). The other models for HT12 and HT35 failed to converge.

The ABLUP within-site cross-validation, provided results of a similar nature as the GS models trained on EBVs. The average prediction accuracy for ABLUP within-site was 0.87 (both RR-BLUP and GRR had averages of 0.86), with a range of 0.77–0.96. WD_{res} was predicted with the highest accuracy of all three traits, and surpassed the accuracy of the GS models: 0.94 ± 0.0009 , 0.96 ± 0.0009 , and 0.94 ± 0.001 , for Adam, Fleet River, and Lost Creek, respectively (Fig. 2.3).

2.3.2.2 Cross-sites prediction accuracy

The average predictive accuracy of the RR-BLUP and GRR genomic selection models was very similar for cross- and within-site analyses. However, some trends in the predictive ability of the sites did occur (Fig. 2.2a and b). The sites Adam and Fleet River always produced the same or higher prediction accuracies for Lost Creek than the within site estimate for Lost Creek. In addition, Lost Creek cross-site prediction accuracies for Adam and Fleet River were also always higher than for itself. Fleet River always produced higher prediction accuracies for Adam than itself. This was true for all traits using EBVs. This may be an indication that whilst there may be some GxE occurring as expected, it may not be significant enough to employ single-site testing and breeding. The accuracy of cross-site predictions varied and ranged from 0.78 (GRR for WD_{res} : Adam - Lost Creek) to 0.92 (RR-BLUP for HT12: Lost Creek - Adam), and both selection

models provided comparable results (Fig. 2.2a and b). Overall, across sites prediction accuracy decreased from HT12 to HT35 to WD_{res} with averages of 0.90, 0.86, and 0.83, respectively (all from RR-BLUP).

When the studied attributes' deregressed breeding values (DEBVs) were used to train the two GS models, surprising outcomes were obtained and the resulting predictive accuracies were extremely low approximating 0.0 for WD_{res} , while the remaining models for HT12 and HT35 failed to converge.

The results of the ABLUP cross-sites validation provided evidence of stronger GxE interaction than was predicted by the GS models (Fig. 2.3). This particular trend can occur as site-specific GxE interactions tend to over-estimate within-site prediction accuracy due to the sharing of a common environment, whilst seemingly inhibiting the efficacy of cross-site analyses. The average cross-sites prediction accuracies for ABLUP were as follows: 0.68, 0.70, and 0.79 for HT12, HT35, and WD_{res} , respectively (Fig. 2.3). A drop from the accuracies obtained by the GS models trained on EBV data, and a complete reversal of the order of those accuracies.

2.3.2.3 Within multi-site (combined) prediction accuracy

The RR-BLUP and GRR models gave almost identical results, with HT12 having the highest combined-site prediction accuracy, followed by HT35 and lastly WD_{res} (0.91, 0.88, and 0.83, respectively for both RR-BLUP and GRR) (Fig. 2.2a and b, Table 2.2). In general, the combined site prediction accuracies were higher than the average of the within-site accuracies, except for WD_{res} which produced the same accuracy for both (RR-BLUP: 0.91 vs. 0.89, 0.88 vs. 0.86, 0.83 vs. 0.83; for HT12, HT35, and WD_{res} , respectively) (Fig. 2.2a). Finally, it is also interesting to note the exceedingly small standard error values associated with all within multi-site prediction accuracies, highlighting the model(s) fit (Table 2.2).

Again, using the deregressed breeding values (DEBVs) to train the two GS models, we obtained results approximating 0.0 for all within multi-site cross-validation analyses for all traits considered (HT12, HT35, and WD_{res}) (Table 2.2).

The multi-site cross-validation accuracies for the ABLUP model were: 0.88 ± 0.002 , 0.86 ± 0.003 , and 0.84 ± 0.003 , for HT12, HT35, and WD_{res} , respectively (Fig. 2.3). Similar in magnitude and sequence to the GS models trained with EBVs.

2.3.2.4 Multi-site to single-site prediction accuracy

On average the multi- to single-site predictability for each trait was slightly higher than the within multi-site average estimates (for RR-BLUP, HT12: 0.91 vs. 0.89, HT35: 0.87 vs. 0.86, and WD_{res} : 0.85 vs. 0.83) (Fig. 2.2a). In most cases the multi- to single-site accuracy predictions were the same or higher, than the corresponding single-site predictions. Except for HT35 at Lost Creek for both RR-BLUP and GRR which gave slightly lower prediction accuracies for multi- to single-site than within site (Lost Creek) analyses (RR-BLUP: 0.79 vs. 0.82; GRR: 0.79 vs. 0.80) (Fig. 2.2a and b). In most cases Adam was the most predictable site, and on average, Lost Creek was the least predictable.

Again, using the deregressed breeding values (DEBVs) to train the two GS models, we obtained results approximating 0.0 for all multi- to single-site cross-validation analyses.

The multi- to single-site predictions for the ABLUP model closely followed the pattern of predictions from the GS models trained on EBV data. In each case (for HT12, HT35, and WD_{res}) Adam was the most predictable site (joint first with Fleet River for WD_{res}), Fleet River the second, and Lost Creek the least predictable site (Fig. 2.3). Average multi- to single-site predictability for the traits were again the same or slightly higher than the within multi-site predictions for ABLUP (HT12; 0.88 vs. 0.88, HT35: 0.86 vs. 0.86, and WD_{res} : 0.85 vs. 0.84) (Fig. 2.3).

2.3.2.5 Two sites predicting one site accuracy

The RR-BLUP and GRR models for this analysis produced similar results overall (Fig. 2.4a and b). Generally, the highest prediction accuracies were obtained for HT12 (average = 0.90), followed by HT35 (average = 0.88), and lastly WD_{res} (average = 0.83) for both RR-BLUP and GRR. There were only minor differences between the two models (Fig. 2.4a and b). Similar to the multi- to single-site (above), the prediction accuracy of two sites to one site indicated that, in all cases, Adam was the most predictable site using this two to one analysis, despite the discrepancy in heritabilities, and on average Lost Creek was the least predictable in most cases.

The deregressed breeding values produced similar results to those obtained from within, cross- and multi-sites GS models, with two sites to one site cross-validation analyses prediction accuracy approximating 0.0 for HT12, HT35, and WD_{res} .

The ABLUP cross-validation for two sites predicting one site resembled the results of the GS model trained on EBVs. Prediction accuracies for HT12 and HT35 were lower than the GS models using EBVs (HT12 average = 0.81 vs. 0.90, and HT35 average = 0.81 vs. 0.88) (Fig. 2.5). However, the average predictability for WD_{res} remained the same for ABLUP as the GS models using EBVs (0.83) (Fig. 2.5). In general Adam was the most predictable site for all three traits (HT12, HT35 and WD_{res}), followed by Fleet River and lastly Lost Creek in this ABLUP analysis (Fig. 2.5). This is the same site predictability trend displayed by the GS models trained on EBV data.

2.3.2.6 Time- time prediction accuracy (age- age/ trait-trait correlation)

In order to test the theory that the target time for forward selection can be reduced, prediction models were assessed on their accuracy when trained on younger trees (12 years: HT12) followed by validation on the same trees for height at age 35 (HT35) and wood density at age 38 (WD_{res}) (i.e., correlations between GEBVs at age 12 and EBVs at ages 35 (HT12-HT35) and 38 (HT12- WD_{res}). The GEBVs values for HT12 have significant positive correlations with EBVs for HT35 for both RR-BLUP (0.71 ± 0.0004) and GRR (0.71 ± 0.0004) (Table 2.3). Trait- trait correlation between HT12 and WD_{res} (recorded at 38 years) produced significant negative correlations (RR-BLUP: -0.46 ± 0.0005 ; GRR: -0.46 ± 0.0006) (Table 2.3). These are the most useful correlations to make as they reflect the direction in which selections will be made.

Using the deregressed breeding values to train the two GS models, we obtained results approximating 0.0 for all time-time cross-validation analyses (Table 2.3).

Time-time and trait-trait cross-validation of ABLUP resembles closely that of the GS models trained on EBV data. EBVs for HT12 have a medium to strong positive correlation with EBVs at HT35 (0.74 ± 0.001), and a significant negative correlation with EBVs of wood density WD_{res} at age 38 (-0.48 ± 0.002).

2.4 Discussion

2.4.1 Exome capture

Exome capture is a target enrichment method for sequencing the protein coding regions in a genome. This makes analysis much more efficient. Through targeting this region, the focus is immediately resolved to those areas that are likely to contain sources of variation for the phenotype. The reduced cost and time compared to that of whole genome sequencing, as well as the ability to still capture a significant proportion of variants makes exome capture desirable as it is harder to find functional variants in noncoding regions (Kiezun et al., 2012; Mertens et al., 2011; Suren et al., 2016). Added to which, forest trees already tend to have large and complex genome sizes (Neale and Kremer, 2011; Neale et al., 2017; Suren et al., 2016). Exome capture has been recognized as an effective method for increasing knowledge of unmapped large genomes (Mertens et al., 2011; Neves et al., 2013).

Whilst exome capture is expected to produce some missing marker information, the method surpasses whole genome shotgun sequencing or genotyping methods using genome complexity reduction, in depth of coverage. Given a restricted budget (with no option for re-sequencing) the depth achieved using exome sequencing is expected to be much greater than genotyping-by-sequencing (Bodi et al., 2013), with the added benefit of obtaining more reliable calls. After the sequencing process, those markers with a large proportion of missing information can be filtered out, however Rutkoski et al. (2013) admit that this step is unnecessary since it affects the GS prediction accuracy little. In order to maintain power in the subsequent analyses, marker imputation should be used to infer missing data in those records not filtered out, conserving the majority of the original SNPs. This was carried out in the current study, although the ratio of missing data was very low and therefore the impact of this procedure is expected to be minimal. Given that currently there is no genetic map available for Douglas-fir, imputation methods in this case are restricted to those which support unordered data. There have been various imputation methods proposed for unordered markers however their efficacy and suitability is yet to be fully determined in genomic selection (Rutkoski et al., 2013). Though some methods do show some promising results when compared

to mean imputation, notably random forest regression produced greater GS prediction accuracy than its counterparts in both Rutkoski et al. (2013) and Poland et al. (2012) studies.

Although we managed to capture significant family effects using this type of SNP data, we found little evidence that the GS models were able to capture marker-QTL LD using this genotypic data set (see below for discussion). It is highly likely that substantially more SNPs will be required to capture significant effects for these traits (i.e., LD). In addition to this, our genotyping efforts were focused on a restricted portion of the genome, the exome, which is limited to the available 40 K probes used in the present study. In humans, the exome constitutes a mere 1% of the total genome (Ng et al., 2009), thus considering the unique genome size and complexity of conifers (De La Torre et al., 2014) we expect that the population of SNPs used in this study represents a very small fraction of the Douglas-fir genome. Additionally, in this case the revealed exome which represents functional genes that, by default, are under selection and thus are conserved. By focusing the present study on this region we were not able to capture the variation among families, as this was not represented in the exome. In order to seize this variation in intergenic regions, an alternative, whole genome approach must be used for genotyping, for example genotyping-by-sequencing as it uncovers unordered SNPs across the entire genome. These intergenic regions are not under the same selection pressure as the exome, and may contain important regulatory sequences which correspond to the control of the traits we investigated. Another approach used by Fuentes-Utrilla et al. (2017), was to use restriction-site associated DNA sequencing (RADseq) technology. They found that in a species with no whole genome assembly (in this case Sitka spruce (*Picea sitchensis* (Bong.) Carr)), a SNP panel could be constructed from randomly located markers generated from RADseq. However, it must be noted that their GS analysis was performed strictly within a single family.

2.4.2 Heritability

The use of ABLUP instead of GBLUP is likely falsely inflating the heritability estimate. Though the impact of heritability on the predictive accuracy seems to be low in this study. Our results show that even with modest heritability, predictive accuracies can be high. As shown by the combined site analyses, the

heritability for HT12, HT35, and wood density were modest (0.17, 0.17, and 0.43, respectively), yet the prediction accuracies were 0.91 (± 0.004), 0.88 (± 0.006), and 0.83 (± 0.009), respectively (RR-BLUP) (Table 2.2). The large sample size and low effective population size of the present study ($N_e = 21$) likely helped in negating the effect of low trait heritability (Meuwissen et al., 2001). Märtens et al. (2016) provided evidence that increased relatedness between training and validation populations leads to higher prediction accuracy in their study on yeast. Furthermore, the use of correlation between pedigree-based and marker-based breeding values, approximates correlation between unknown true breeding value and genomic breeding value. The accuracy can go above heritability in this case because both values are representing only genetic effects, this is in line with results obtained by Gamal El-Dien et al. (2015).

2.4.3 Genomic selection

The desired outcome of genomic selection is to produce unbiased marker effect estimates (Meuwissen et al., 2001), and to avoid the Beavis effect (Xu, 2003) which hinders MAS causing marker effects to be overestimated (Lande and Thompson, 1990). Instead of selecting markers based on a significance threshold, GS estimates all marker effects simultaneously causing a different problem; there are more predictor effects (p , markers in this case) to be estimated than there are observations (n , samples) (Lorenz et al., 2011). Least squares cannot be used to estimate all the effects at once since there are not enough degrees of freedom. In addition, multi-collinearity of markers would cause any model to be over fitted (Lorenz et al., 2011).

To address this issue (large p , small n), various statistical models have been proposed. They generally fall into the following categories: shrinkage models, variable selection models, and kernel methods. Shrinkage models (e.g., ridge regression BLUP [RR-BLUP], (Whittaker et al., 2000)) fit all marker effects which are all shrunk to the same degree. With RR-BLUP, it is assumed that the trait in question more closely resembles Fisher's infinitesimal model (many loci with small effects), and marker effects are samples from a normal distribution (with equal variance). Variable selection (e.g., generalized ridge-regression [GRR], (Shen et al., 2013)) however, reduces the number of markers used, and in doing so assumes that the trait is controlled by fewer, strong effect loci (Lorenz et al., 2011). Kernel methods

convert the predictor variables to distances in effect creating a matrix similar to an additive genetic relationship matrix. The kernel matrix quantifies the distance between individuals but also smoothing parameters are added (Lorenz et al., 2011). This method is flexible and can incorporate complex relationships between markers, for this reason it is useful in cases where non-additive effects are suspected to occur (Gianola and van Kaam, 2008). Indeed, Douglas-fir height and wood density have proven to have non-additive genetic variance component (Yanchuk, 1996); however, its unpredictability has driven the species' advanced generation breeding and selection methods to be additive genetic gain-dependent (El-Kassaby and Park, 1993; Krakowski et al., 2006; Yanchuk, 1996). Since different traits have differing genetic architecture, there does not exist one model that is necessarily the best for all traits or populations (Lorenz et al., 2011).

In the present study, we assessed the prediction accuracies of two GS models (RR-BLUP and GRR) in two phases; first when trained on EBVs, and second when trained on DEBVs. In the first instance, using EBVs and by virtue of retaining family means, our data contained both contemporary and historical pedigree information as well as marker-QTL LD information. Without further adjustment, all this information was parsed into the GS models and resulted in high prediction accuracies (for both model types). Subsequent to this, in phase two, we deregressed the EBVs, removing the parent average effect in order to disassociate the pedigree information from the marker-QTL LD. This resulting in DEBVs that contained the marker-QTL LD information only without the between family genetic variance. Using these DEBVs to train the GS models, we obtained prediction accuracies which for all practical purposes were 0.0. The success of the first phase can be attributed to the large within and among family genetic variances. This pedigree-driven approach dominated the analysis and produced the observed high prediction accuracies, which were comparable to the ABLUP accuracies. When this influence of pedigree was removed, the generated models were no longer able to provide useful predictions. A similar effect was seen in interior spruce (*Picea glauca* x *Picea engelmannii*) where cross-validation using family folding resulted in decreased prediction accuracy, because between family variance was dominating the analysis (El-Dien et al., 2018). In addition to this Fuentes-Utrilla et al. (2017), when studying GS in Sitka spruce, found that

within family predictions cannot be extrapolated to between families. Furthermore, when examining marker transferability between families in white spruce, Beaulieu et al. (2011) also found within family predictions to be more precise than between family predictions. Similarly, to El-Dien et al. (2018) they also found that when a family had no representation in the training group, the accuracies obtained for that family were very small, occasionally negative, and often not statistically significant from zero. Much like using the DEBVs here, which try to predict between family variance, but have family means stripped away. Fuentes-Utrilla et al. (2017) concluded that species with large effective population sizes (notably conifers), have a reduced ability to make predictions across families. With this in mind, GS may best be employed to produce GEBVs for within large full-sib families in conifers as it captures the Mendelian sampling term.

This discrepancy in the results from the two GS phases, is indicative that we simply were not able to capture the marker-QTL LD with the available SNPs. Whilst other studies have had some success in this area, it is important to note that these particular investigations have focused on tree species with much smaller genomes, for example Resende et al. (2012a). Resende et al. (2017) make this point in their study of *Eucalyptus*, that although other studies may have failed to produce significant predictions between unrelated populations, this may be due to low marker density. This is likely inhibiting the ability to capture short-range LD, the result being that prediction models rely almost entirely on relatedness.

In their study, Resende et al. (2012a) used 7680 DArT markers on *Eucalyptus* which is estimated to have a genome size of 609 Mbp (Resende et al., 2012a), equivalent to 12.61 markers/Mbp. In contrast, here we used 69,551 SNP markers on the Douglas-fir genome which is estimated at 18,700 Mbp, giving 3.72 markers/Mbp. To obtain a similar coverage as Resende et al. (2012a) would require approximately 235,800 markers, many more than we currently have. With such a large genome, it is likely that many more SNPs will be required ($\approx 235,800+$), with greater genomic coverage in order to capture this, so far elusive, LD. In a more recent study, also using *Eucalyptus*, Müller et al. (2017) managed to capture some short-range historical LD using 5000–10,000 SNPs. Yet they too concluded that the genomic prediction in this case was largely driven by relatedness.

The similar predictions given by the RR-BLUP and GRR methods, and the similarly high ABLUP accuracies, was again the result on heavy reliance on between family variance, and thus we have gained no new information as to the genetic control or architecture of the traits in question. Although similar findings have been noted in more successful GS studies. For example, Resende et al. (2012c) when comparing GS methods, found there to be little difference between the predictive abilities of shrinkage and variable selection methods (4 in total) even considering they have different *a priori* assumptions. They used a *Pinus taeda* training population with 951 individuals and 4853 SNPs in the analyses. Prediction accuracies for 17 traits (including growth, disease resistance and development) ranged from 0.17 to 0.51. Only one trait (fusiform rust resistance) showed any significant difference between the models. Higher prediction accuracies were obtained using variable selection methods for this trait. This reflects the genetic architecture of the trait which is controlled by few, large effect loci (Resende et al., 2012c).

2.4.4 Single and multi-site cross-validation

The combined site GS analysis produced higher prediction accuracies than the single site analyses on average (Fig. 2.2a and b). The combined site training population having more individuals than the single site populations. This is in agreement with what the literature states should happen; Grattapaglia (2014) noted that increasing the training population size increases accuracy up to a point (around 1000 individuals). In addition to increased sample size for predictive accuracy improvement, the multi-site approach incorporated the present GxE in the model, resulting in further improvement.

The prediction accuracies are high for HT12, HT35, and WD_{res} GEBVs (0.87–0.92, 0.79–0.92, and 0.78–0.88, respectively) (Fig. 2.2a and b). They are moderately higher than those in previous studies including other forest tree species (Beaulieu et al., 2014b; Isik et al., 2016; Resende et al., 2012a, 2012b, 2012c), largely due to the inclusion of the pedigree structure. In this case the GS methods are not giving much advantage over ABLUP (ABLUP multi-site cross-validation accuracies: 0.88 ± 0.002 , 0.86 ± 0.003 , and 0.84 ± 0.003 , for HT12, HT35, and WD_{res} respectively), thus both are predicting only family means. The prediction accuracies for both the GS models and the ABLUP model are much higher than theoretical

accuracy of the EBVs (0.61, 0.68, and 0.76; for HT12, HT35, and WD_{res} , respectively), which indicates that both the EBVs and GEBVs are converging on the family means, and are far from the true breeding values.

The high prediction accuracies for the GEBVs may also partially be the result of using a relatively large training population, known to correlate with accuracy. In addition, there is an interacting effect of the relatively low effective population size ($N_e = 21$), and both these data characteristics increase the accuracy of predictions (Lorenz et al., 2011). Although in this case the low effective population/family size also meant that the Mendelian sampling term could not be defined. Indeed, Gamal El-Dien et al. (2015) using an interior spruce population with $N_e = 93.8$ (estimated assuming the OP families are unrelated and contributed equally to the experiment) had lower prediction accuracies than obtained here. But these in turn were vastly higher than results from a study with 214 open-pollinated white spruce (*Picea glauca*) families in Quebec with $N_e = 622.5$ (Beaulieu et al., 2014a). Another component responsible for the increased accuracy of predictions, was the training of the models on EBVs rather than raw phenotypes (Gamal El-Dien et al., 2015).

Prediction accuracies fell dramatically for models trained on DEBVs, as marker-QTL LD could not be recovered using the available SNP data set, indicating that the SNP markers effectively tracked the pedigree.

2.4.5 Cross-site validation

Similar prediction accuracies were observed in the cross-site compared to the single-site and combined-site GS analyses (with models trained on EBVs). Genotype x Environment interaction is an important consideration of forest tree selective breeding, more so than in animal breeding where individuals are considered to share a common environment (Burdon, 1977; Matheson and Cotterill, 1990; Matheson and Raymond, 1984; Owino, 1977). Prediction accuracy for HT12 across sites was relatively high. An unexpected result given the fact that the heritability was only 0.17 (combined sites), and there is expected to be a competition effect at early stand development (i.e., strong environmental component). This is possibly an indication of the partially controlled experimental environment (not natural stands) in which

the trees were growing. Given these experimental conditions, it is also conceivable that GxE effects are minimal. Which is reflected in the cross-site predictions (Fig. 2.2a and b), which are of a similar magnitude to the within-site prediction accuracies for all attributes.

2.4.6 Time- time and trait- trait correlations

It is thought that forward selection in Douglas-fir should be carried out at a minimum of 17 years (Magnussen and Yanchuk, 1993), when accurate predictions of phenotype at time of harvest can be measured. We tested the correlation of height at 12 years and height at 35 years (HT12-HT35), and wood density (38 years) (HT12-WD_{res}) and positive (0.71 for height) and negative (-0.46 for wood density) correlations were detected by the GS models trained on EBV data. Although they did not offer accuracy above that of ABLUP, indicative of a strong reliance on pedigree information rather than marker-QTL LD. The results give an indication of how useful early selection could be. Correlations for height are moderately strong and low-moderate for wood density. Marker-trait associations are known to vary according to the tree age, limiting any correlation. This would certainly hamper efforts to create a highly correlated age-age model in trees (Beaulieu et al., 2014a; Lerceteau et al., 2001). At the moment providing such predictions based on EBVs obtained for an age younger than 12 years would not be recommended (note that height at age 12 was the earliest measurement available in the present study). Since larger age differences have been shown to produce less accurate models (Ratcliffe et al., 2015). The discrepancy in prediction accuracy between the time-time correlations and the cross-validations suggest that there is some inconsistency between EBVs and marker effects at these two ages (12 and 35 years). Although the discrepancy is relatively small, and so meaningful results may still be obtained through age-age and trait-trait correlations. This is, in addition to the varied environmental conditions the trees endure over their long lifespans, lessening time- time correlations. To this effect and as White et al. (2007) suggested, training models will need to be updated with mid-rotation phenotypes in order to accurately predict mature trait values.

2.4.7 Genomic selection and forest tree breeding

Several genomic selection “proof-of-concept” studies have been conducted on few coniferous tree species (e.g., loblolly pine, maritime pine, Sitka spruce, white spruce, and white-Engelmann spruce hybrid), all concluded that GS has the potential to increase the genetic gain through speeding breeding generation turnover and increasing the selection differential. It should be stated that, with the exception of the maritime pine study (Bartholomé et al., 2016), none of the derived GS predictive models have been validated on independent “validation” populations. The success of GS is dependent on the linkage disequilibrium between the markers used (i.e., SNP panels) and causal genes underpinning the traits of interest, and the degree of relatedness between the training and validation populations. Therefore, caution is required during GS implementation as LD changes after every round of breeding (i.e., recombination); the fact that it does rapidly decay called for using dense marker coverage to overcome this caveat. Still we have found that by only using sequence capture data, we were unable to successfully resolve this marker-QTL LD. We only had success in capturing among family effects (i.e., pedigree). Even with a SNP panel designed appropriately based on an informative SNP library, and large enough to handle a conifer genome, there are additional hurdles. LD only survives over relatively short distances in conifers compared to livestock species due to their relatively large N_e (Neale and Savolainen, 2004). This led Fuentes-Utrilla et al. (2017) to conclude that GS may only be useful in tree populations with reduced N_e , for example seed orchards, or lines/ sub-groups which have been produced through selective breeding. Though as they demonstrate with their analysis, it is possible to generate very large full-sib families in trees, by controlled crossing. In this type of population, LD extends over longer distances than in open-pollinated populations. As a result, they suggest that GS could be employed to make selections within families.

Based on comparable studies: 1) a greater number of markers, and 2) wider coverage throughout the whole genome or dense unordered SNP genotyping platform (e.g., GBS), would be needed to capture this LD and additional intergenic variation (Resende et al., 2017). Or indeed a shift in the level at which GS is applied, i.e. to within families rather than across families (Fuentes-Utrilla et al., 2017). Whilst GS

still has the potential to deliver unprecedented gains, it does not seem likely that was achieved in the present study as the prediction driving force was the pedigree rather than LD. Despite the low N_e , the small family size prevented the accurate assessment of the Mendelian sampling term in this population. Therefore, the EBVs were heavily shrunk toward the family mean and all within family deviations were not estimated precisely.

Finally, it should be emphasized that any gains captured through GS require unique tree improvement delivery methods as the traditional seed orchards' production mode requires time for reaching sexual maturity, even under intensive management such as top grafting or hormonal applications, and its sexual-production effectively breaks the LD between selected traits and markers.

2.5 Summary

Genomic selection (GS) can offer unprecedented gains, in terms of cost efficiency and generation turnover, to forest tree selective breeding; especially for late expressing and low heritability traits. Here, we used: 1) exome capture as a genotyping platform for 1372 Douglas-fir trees representing 37 full-sib families growing on three sites in British Columbia, Canada and 2) height growth and wood density (EBVs), and deregressed estimated breeding values (DEBVs) as phenotypes. These are representative of models with (EBVs) and without (DEBVs) pedigree structure. Ridge regression best linear unbiased predictor (RR-BLUP) and generalized ridge regression (GRR) were used to assess their predictive accuracies over space (within site, cross-sites, multi-site, and multi-site to single site) and time (age-age/ trait-trait).

The RR-BLUP and GRR models produced similar predictive accuracies across the studied traits. Within-site GS prediction accuracies with models trained on EBVs were high (RR-BLUP: 0.79–0.91 and GRR: 0.80–0.91), and were generally similar to the multi-site (RR-BLUP: 0.83–0.91, GRR: 0.83–0.91) and multi-site to single-site predictive accuracies (RR-BLUP: 0.79–0.92, GRR: 0.79–0.92). Cross-site predictions were surprisingly high, with predictive accuracies within a similar range (RR-BLUP: 0.79–0.92, GRR: 0.78–0.91). Height at 12 years was deemed an acceptable age at which moderately accurate predictions can be made concerning future height (age-age) and wood density (trait-trait) based on genomic

information. Using DEBVs reduced the accuracies of all cross-validation procedures dramatically, indicating that the models were tracking pedigree (family means), rather than marker-QTL LD.

While GS models' prediction accuracies were high, the main driving force was the pedigree tracking rather than LD. It is likely that many more markers are needed to increase the chance of capturing the LD between causal genes and markers.

Table 2.1 Within single-site prediction accuracies and their standard errors for ABLUP and genomic selection (GS) models (ridge regression (RR-BLUP) and generalized ridge regression (GRR)) for height (HT12 and HT35) and wood density (WD_{res}). GS prediction accuracy is calculated as the mean of the replications of the Pearson product-moment correlation between EBVs of the validation set and their predicted GEBVs. ABLUP prediction accuracy is the mean of the replications of the Pearson product-moment correlation between original EBVs and predicted EBVs.

Trait	Model	Site		
		Adam	Fleet River	Lost Creek
HT12	ABLUP	0.81 ± 0.002	0.77 ± 0.002	0.88 ± 0.001
	RR-BLUP	0.91 ± 0.010	0.89 ± 0.012	0.87 ± 0.013
	GRR	0.91 ± 0.010	0.89 ± 0.012	0.87 ± 0.012
HT35	ABLUP	0.85 ± 0.002	0.83 ± 0.002	0.88 ± 0.002
	RR-BLUP	0.89 ± 0.007	0.86 ± 0.020	0.82 ± 0.010
	GRR	0.89 ± 0.008	0.84 ± 0.020	0.80 ± 0.013
WD_{res}	ABLUP	0.94 ± 0.001	0.96 ± 0.001	0.94 ± 0.001
	RR-BLUP	0.87 ± 0.007	0.83 ± 0.026	0.79 ± 0.015
	GRR	0.86 ± 0.008	0.83 ± 0.026	0.84 ± 0.016

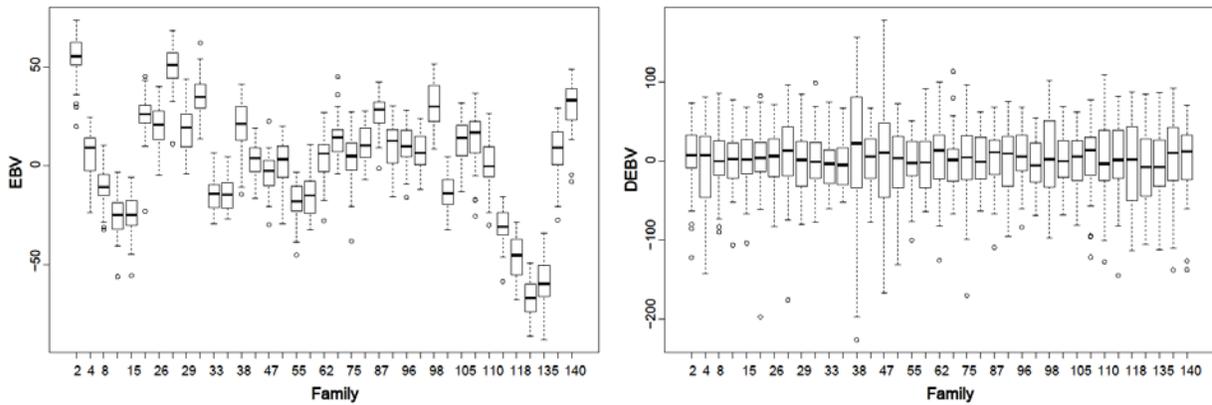
Table 2.2 Within multi-site genomic selection prediction accuracies and their standard errors for ridge regression (RR-BLUP) and generalized ridge regression (GRR) models, estimating GEBVs and GDEBVs for heights (HT12 and HT35) and wood density (WD_{res}).

Trait	GS Model	GEBVs	GDEBVs
HT12	RR-BLUP	0.91 ± 0.004	-0.09 ± 0.019
	GRR	0.91 ± 0.003	-0.04 ± 0.017
HT35	RR-BLUP	0.88 ± 0.006	-0.02 ± 0.021
	GRR	0.88 ± 0.006	-0.01 ± 0.029
WD_{res}	RR-BLUP	0.83 ± 0.009	0.00 ± 0.032
	GRR	0.83 ± 0.010	-0.01 ± 0.025

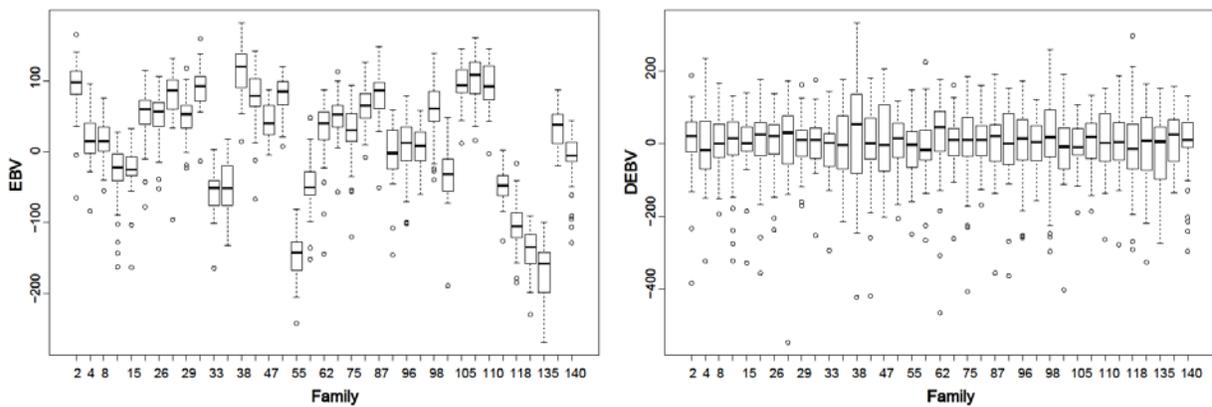
Table 2.3 Genomic selection prediction accuracies for time- time correlations for HT12 for RR-BLUP and GRR models (standard errors). Prediction accuracies based on correlations between GEBVs at age 12 and trait EBVs at age 38; and correlations between GDEBVs at age 12 and trait DEBVs at age 38. Traits key: HT35 = height at 35 years (cm); WD_{res} = wood density calculated from resistance drilling to drilling (g/cm^3); HT12 = height at 12 years (cm)

GS Model	EBVs		DEBVs	
	WD_{res}	HT35	WD_{res}	HT35
RR-BLUP	-0.46 ± 0.0005	0.71 ± 0.0004	0.02 ± 0.0038	-0.03 ± 0.0039
GRR	-0.46 ± 0.0006	0.71 ± 0.0004	0.002 ± 0.0040	-0.004 ± 0.0047

a)



b)



c)

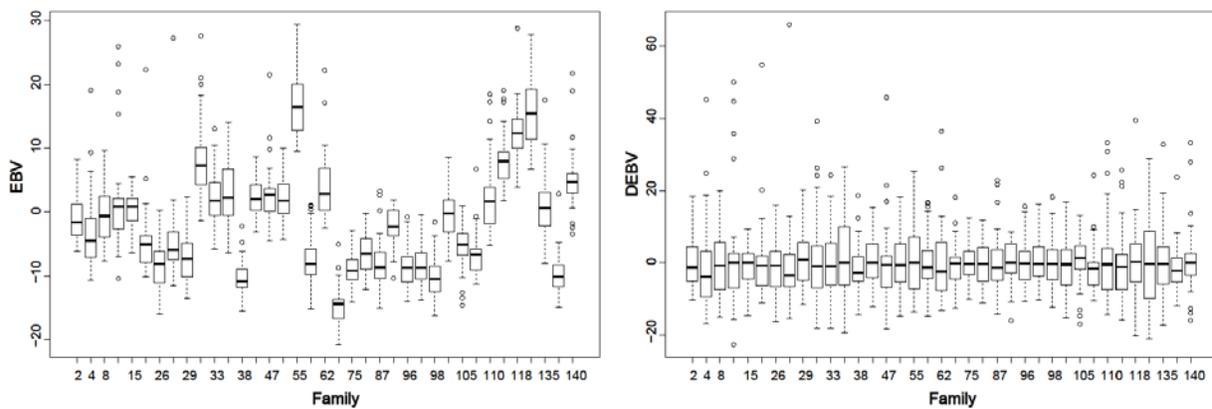


Figure 2.1 Distribution of estimated breeding values (EBVs) and deregressed estimated breeding values (DEBVs) for (a) height at 12 years (cm), (b) height at 35 years (cm), and (c) wood density (g/cm^3) calculated from resistance to drilling.

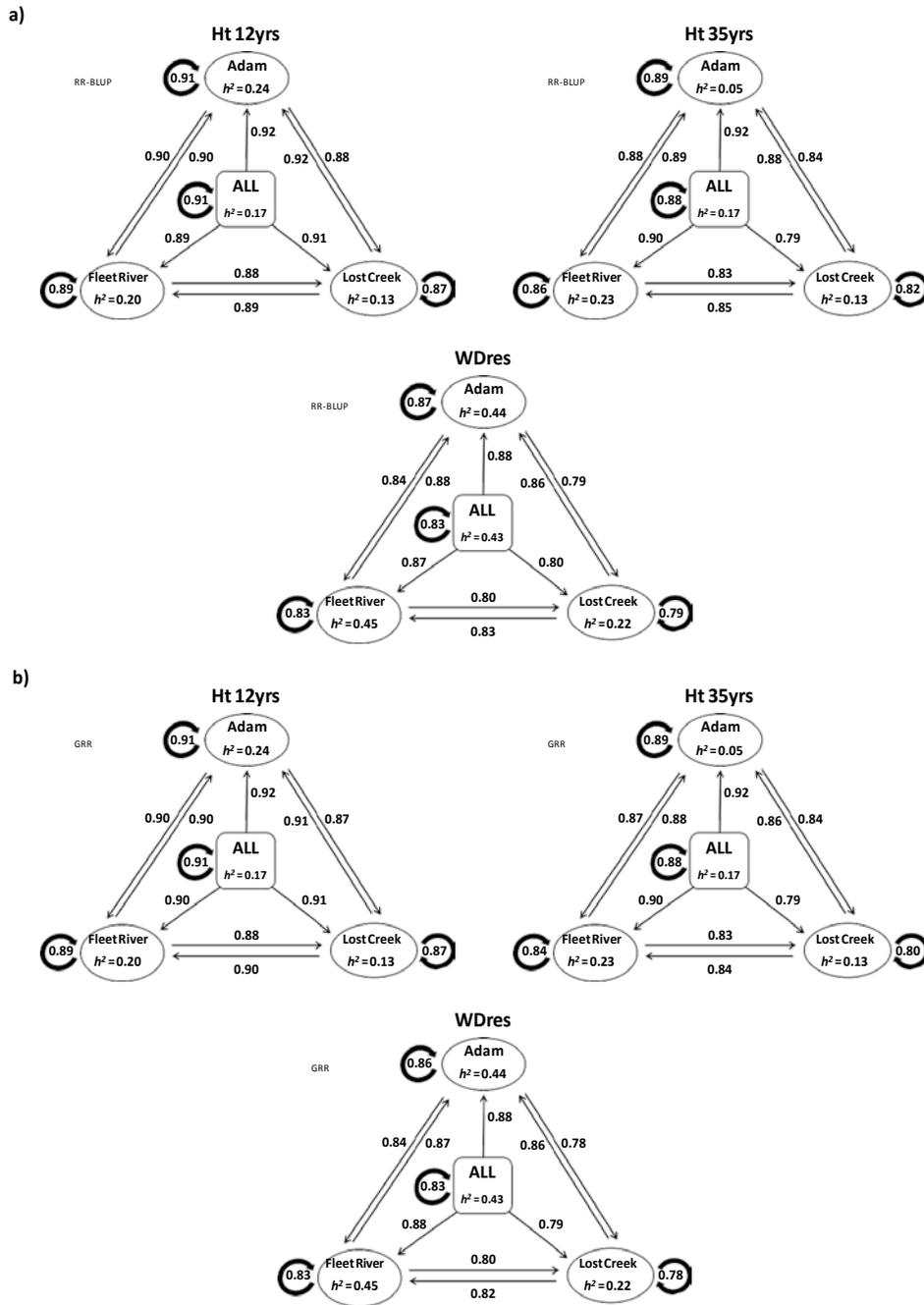


Figure 2.2 Heritabilities and GS prediction accuracies for models trained on EBVs and predicting GEBVs for each of the traits. Showing the results of within-site, cross-site, combined-sites, and multi-site to single-site cross-validation (Top (a), using RR-BLUP, Bottom (b), using GRR). The direction of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance drilling to drilling (g/cm^3). Sites are Adam, Fleet River, Lost Creek, and multi/combined-site (ALL).

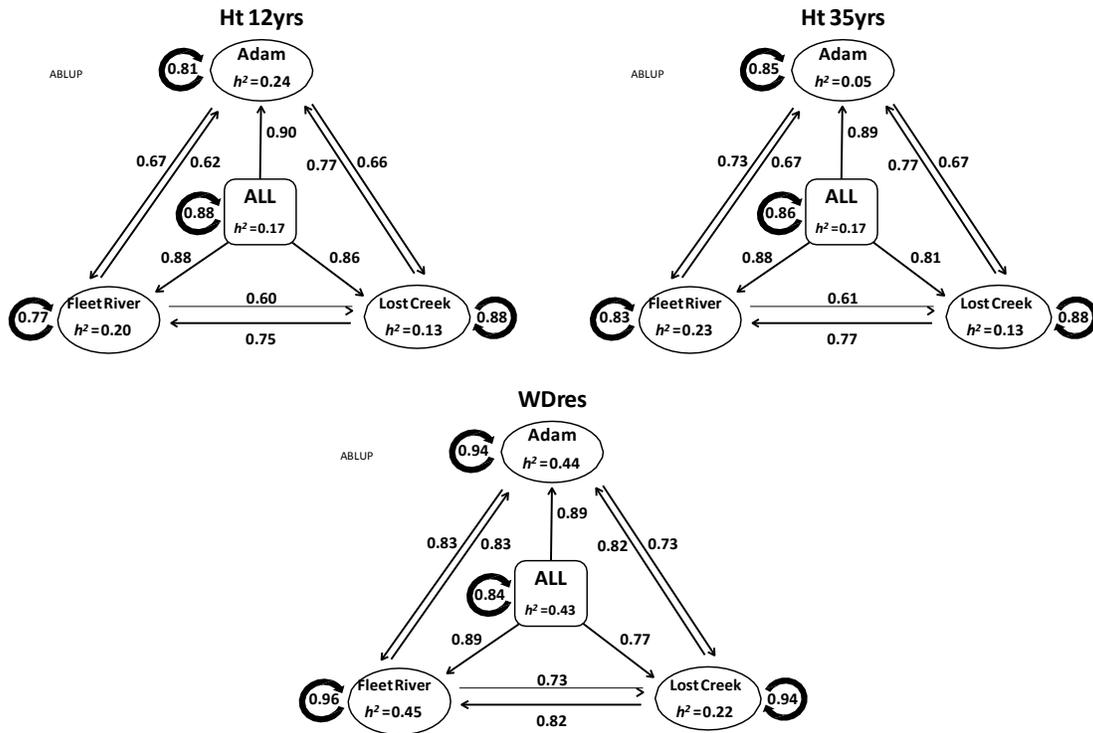


Figure 2.3 Heritabilities and prediction accuracies of the ABLUP model for each of the traits. Showing the results of within-site, cross-site, combined-sites, and multi-site to single-site cross-validation. The direction of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance drilling to drilling (g/cm^3). Sites are Adam, Fleet River, Lost Creek, and multi/combined-site (ALL).

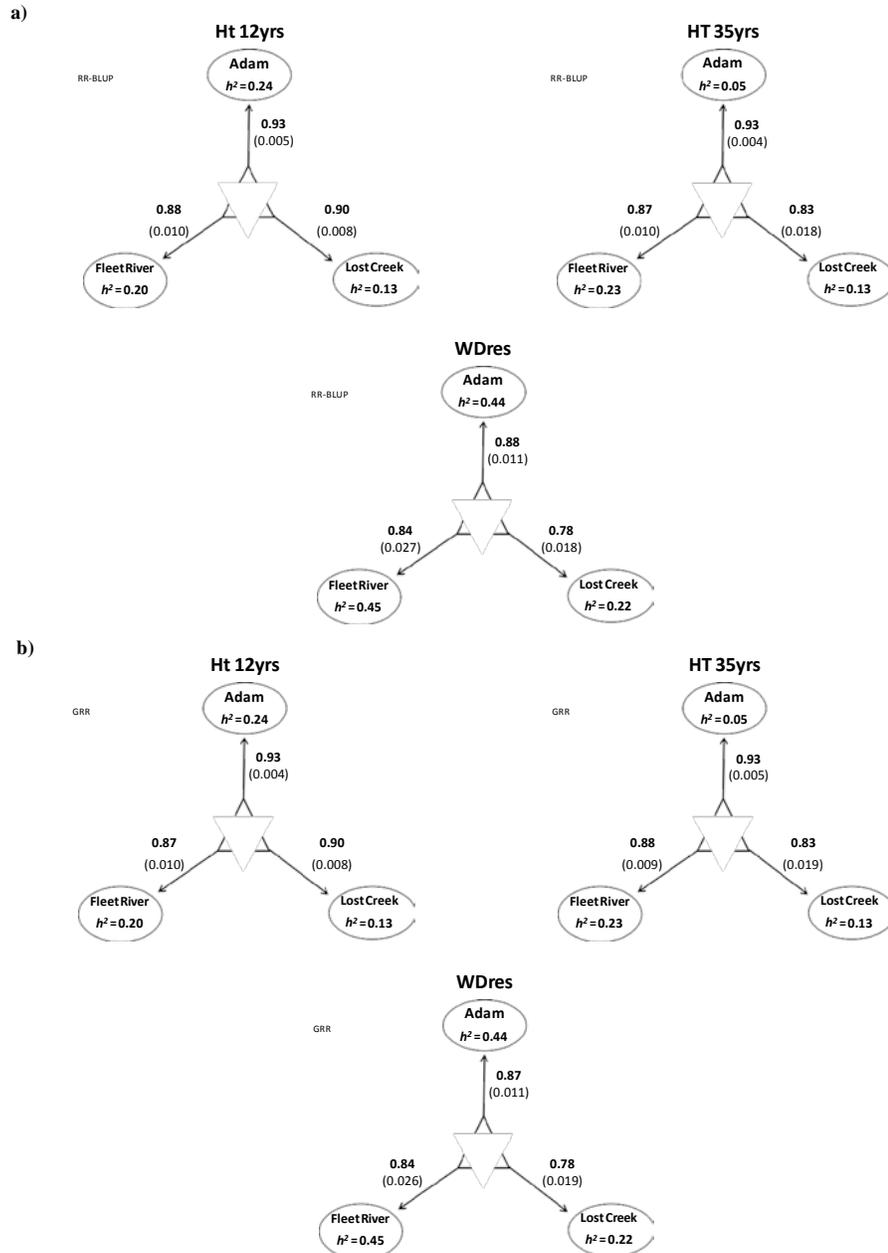


Figure 2.4 Heritabilities and GS prediction accuracies for models trained on EBVs and predicting GEBVs for each of the traits. Showing the results of two sites predicting one site cross-validation (Top (a) using RR-BLUP, Bottom (b), using GRR). The direction of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance to drilling (g/cm^3). Sites are Adam, Fleet River, Lost Creek.



Figure 2.5 Heritabilities and prediction accuracies of the ABLUP model for each of the traits. Showing the results of two sites predicting one site cross-validation. The direction of the arrows depicts what information (site) is used to train the model (shaft end), and which is being predicted (head). Traits key: Ht 35yrs = height at 35 years (cm); Ht 12yrs = height at 12 years (cm); WD_{res} = wood density calculated from resistance to drilling (g/cm^3). Sites are Adam, Fleet River, and Lost Creek.

3 Genomic selection of juvenile height across a single generational gap in Douglas-fir

3.1 Introduction

There is a strong drive to incorporate genomic selection (GS) methodologies, as first proposed by Meuwissen et al. (2001), into forest tree selective breeding. With a proliferation of genomic technologies and steady decline in genotyping costs (Heffner et al., 2010a; Thomson, 2014), breeders are taking full advantage of the availability of large SNP data sets. With these large SNP sets, it is envisaged that linkage disequilibrium (LD) between the markers and most sources of variation for valued complex phenotypes can be tracked. In doing so, capturing more variance than the well-established marker assisted selection (MAS), which relies on fewer, large effect QTLs (El-Kassaby, 1982). Genetically complex traits (such as height, growth, wood quality) are now amenable to selection with the use of dense marker data. Given this statistical advantage, it is anticipated that GS may be implemented into tree selective breeding, as it has been done in livestock breeding (Van Eenennaam et al., 2014), resulting in higher genetic gain per unit time for such traits of interest. This will largely be achieved through the reduction of trait evaluation time for such late expressing traits, leading to a faster turn-over in breeding generations, a significant time-sink in current breeding programs (Hayes et al., 2009b; Heffner et al., 2010a). Furthermore, breeding programs will become more dynamic as they will be able to ensure adaptation to capricious influences such as climate change and biotic disturbance in less time (Grattapaglia, 2014).

Early deterministic simulations by Grattapaglia and Resende (2011) of GS that modeled forest tree species reported promising results. Following this, several experimental investigations have built upon this concept with varying success (Bartholomé et al., 2016; Beaulieu et al., 2014a, 2014b; Fuentes-Utrilla et al., 2017; Gamal El-Dien et al., 2015; Grattapaglia, 2014; Isik et al., 2016; Müller et al., 2017a; Ratcliffe et al., 2015; Resende et al., 2012a, 2012b, 2012c, 2017; Tan et al., 2017; Thistlethwaite et al., 2017). In general, the following phases are involved in the GS process: 1) the genetic and phenotypic

evaluation of a random subset of samples from within a selected population forming the training set from the tree breeding population under investigation; 2) creating a predictive model using this data, in which alleles at all marker loci have their effects simultaneously estimated; 3) implementing a validation or cross-validation process to test the developed models' robustness; and 4) genomic prediction on a different subset of individuals from the same breeding population and selection of candidates from this population for next-generation breeding based on their genomic estimated breeding values (GEBVs) (Grattapaglia, 2014; Meuwissen et al., 2001). The repercussion of this is a paradigm shift, in which the model unit of these breeding analyses shifts from being the line of descent to the allele.

Factors influencing the success of GS are varied, but one, which is entirely at the discretion of the investigator, is the statistical method. Many methods have been proposed for GS and can be differentiated largely by their *a priori* assumptions of variance distribution. Ridge regression best linear unbiased prediction (RR-BLUP) and generalized ridge regression (GRR) have been selected for their computational efficiency. RR-BLUP assumes marker effects to be normally distributed (mean = 0), and to have equal variance. Conversely GRR allows for heterogeneous variances and so employs a step which sets marker-specific shrinkage parameters on BLUP. In addition to these two methods, Bayes-B will also be implemented. The former have been shown to be highly sensitive to genetic relationships, leading to accuracy of GEBV predictions based on markers tracking this relationship rather than LD (Habier et al., 2007; Thistlethwaite et al., 2017). This conclusion led Habier et al. (2007) to recommend the use of Bayes-B as an alternative.

Other features which independently and collectively alter the success of GS include: 1) The extent of LD between markers and QTL, 2) density of marker coverage; 3) training population size; 4) relatedness of samples (Habier et al., 2007), 5) trait heritability; 6) genetic architecture of trait (number of loci and effect size) (Grattapaglia, 2014); and 7) effective population size (N_e) (Lorenz et al., 2011). Again features 2, 3, and 7 are (to a certain extent) under the control of the experimenter.

Regarding relatedness between individuals (feature 4), the parent average effect on EBVs was removed in this study following Garrick et al. (2009). The deregressed information is used as an

alternative to EBVs in genomic selection to reduce the bias and increase reliability. Using the BLUP procedure shrinks both individual and progeny information towards the parent average EBV, and so there is motivation to remove this effect for the following rationale: 1) those records with an EBV but no individual or progeny information do not contribute to genomic estimation. Their EBV is calculated purely on the parental average and therefore do not afford any additional information besides that of the parental EBVs and genotypes; 2) to avoid the shrinking of major effects that are potentially segregating in the parents. Without deregression the EBVs of the offspring will all be shrunk towards the parent average, regardless of whether they inherit the favourable or unfavourable allele (Garrick et al., 2009).

The extent of marker-QTL LD is dictated by the relationship between N_e and the number of markers used (Grattapaglia and Resende, 2011). In accordance with established theory, in a population with low N_e , genetic drift has a stronger effect, resulting in an increase in non-random association of markers and QTL (LD). Thus in this situation, fewer markers are required to capture the variation of the trait of interest. In outbreeding populations, recombination negates LD and so, long-ranging LD is lost over generations. Because of this, models must be continually updated according to the distance between generations. Wu et al. (2015) found that increasing the generational gap between training and testing populations, reduced predictive accuracy of GS. Two factors can be garnered from this information, and their interaction is thought to have the highest impact on the success of GS in trees: 1) N_e must dictate the scale of marker density and 2) LD can be controlled through choosing the N_e (Grattapaglia, 2014). Larger training population sizes were shown to increase predictive accuracy up to a threshold of around 1,000 individuals, beyond which there were negligible gains (Grattapaglia and Resende, 2011). Although Meuwissen et al. (2001) reported findings that suggested larger training populations negated somewhat the effects of low trait heritability, this had little effect on GS predictive accuracy of more complex traits (50 - 100+ QTL) (Grattapaglia and Resende, 2011).

Substantial work has been carried out on testing the efficacy of GS in forest tree selective breeding. Validation populations (or cross-validation) have largely been comprised of individuals from within the same generation as the training population, aside from work performed by Isik et al. (2016)

and Bartholomé et al. (2016). This practice has resulted in effectively side-stepping the issue of the generational gap. To address this issue, we performed a cross-generational GS analysis on coastal Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)). The training population on which the predictive models were trained, was collected from the parental generation (F_1). The set was composed of 1,321 randomly selected, 38-year-old trees, representing 37 full-sib families with replications over three sites in British Columbia (Canada) (Thistlethwaite et al., 2017). The validation population was collected from the progeny of the parental generation (F_2 , $n = 136$) with shared pedigree to the studied 37 full-sib families. A total of 69,551 SNPs were used in the GS analyses to produce a predictive model for juvenile height, which was compared to pedigree based (ABLUP) predictions.

3.2 Materials and Methods

3.2.1 Experimental population

Samples from a 38-year-old replicated coastal Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)) progeny testing population (F_1) was used as the training set to develop the GS predictive models. The breeding program was established by the Ministry of Forests, Lands and Natural Resource Operations of British Columbia (BC), Canada in 1975. A total of 37 of 165 full-sib families were randomly selected for sampling from 3 sites in British Columbia, Canada (Adams (Lat. 50° 24' 42" N, Long. 126° 09' 37" W, Elev. 576 mas), Fleet River (Lat. 48° 39' 25" N, Long. 128° 05' 05" W, Elev. 561 mas), and Lost Creek (Lat. 49° 22' 15" N, Long. 122° 14' 07" W, Elev. 424 mas)). A total of 1,321 individuals ($N_e \approx 21$) from these 3 sites (Adams: $N = 449$, Fleet River: $N = 441$, Lost Creek: $N = 431$) were selected for genotyping and to train the GS models and ABLUP validation model. The N_e was estimated using a program developed by Dr. Milan Lstiburek (Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic) based on the status number concept of Lindgren et al. (1997).

The validation population (F_2) is represented by 247 samples from control pollinated offspring derived from the 37 full-sib families described above with offspring from an additional 5 full-sib families

selected from the same progeny testing population (F₁) (42 F₂ families total). ABLUP accuracies were derived using these 247 samples. Due to missing genotype information some F₂ samples were discarded for GS analysis. The remaining 136 samples were used as the GS validation population, representing 17 parents located at Jordan River, BC (Lat. 48° 25' 52.6 N, Long. 124° 02' 46.2 W, Elev. 150 mas) and established in 2003.

In order to derive best possible estimates for the EBVs, an increased number of individuals (total N = 36,311) was used to provide as much information as possible when fitting the ABLUP model in ASReml-R v4.0 (Butler et al., 2017). Information was used from 11 sites of the 38-year-old progeny testing population (F₁) (N = 33,931), plus their wild progenitors (N = 108) as described previously Yanchuk (1996), an ungenotyped replicate of the F₂ Jordan River validation population (N = 2,025) (North Arm, Lat. 48° 50' 41.7" N, Long. 124° 06' 34.8" W), and the Jordan River site itself (N = 247) were also incorporated into the analysis (total N = 36,311). The EBVs of a genotyped subset were used in the training of all models (Table 3.1). The ‘original’ EBVs of the genotyped subset from Jordan River were used to validate each model.

3.2.2 Phenotyping, deregression, tissue sampling, DNA extraction and genotyping

Early-rotation (juvenile height) (1988 for the training population and 2010 for the validation population) height measurements of the studied trees were recorded (HTJ: in cm). EBVs for HTJ were obtained in ASReml-R v4.0 (Butler et al., 2017) and used as phenotypes for the genomic prediction analysis. In addition, the EBVs were deregressed and parental averages removed, using the method “Removing parent average effects” proposed by Garrick et al. (2009). The resulting deregressed estimated breeding values (DEBVs) were used as alternative phenotypes for GS analysis. This type of deregressed data can be obtained by approximating and back-solving the evaluation equations. The following equations were solved for each individual tree:

$$\begin{bmatrix} \mathbf{Z}'_{PA}\mathbf{Z}_{PA} + 4\lambda & -2\lambda \\ -2\lambda & \mathbf{Z}'_{i-PA}\mathbf{Z}_{i-PA} + 2\lambda \end{bmatrix} \begin{bmatrix} \mathbf{PA} \\ \mathbf{EBV} \end{bmatrix} = \begin{bmatrix} y_{PA} \\ y_{i-PA} \end{bmatrix} \quad (1)$$

Where **PA** and **EBV** represent the parental average and estimated breeding value vectors respectively; y_{PA} and y_{i-PA} represent information equivalent to the right-hand-side elements referring to the PA and individual respectively; $\lambda = (1-h^2)/h^2$; $\mathbf{Z}'_{PA}\mathbf{Z}_{PA}$ and $\mathbf{Z}'_{i-PA}\mathbf{Z}_{i-PA}$ express the unknown information content of the parental average, and individual effect without parental average, respectively. These latter terms can be equated by solving firstly equation 2 and using the result to solve for equation 3:

$$\mathbf{Z}'_{PA}\mathbf{Z}_{PA} = \lambda(0.5\alpha - 4) + 0.5\lambda\sqrt{\left(\alpha^2 + \frac{16}{\delta}\right)} \quad (2)$$

$$\mathbf{Z}'_{i-PA}\mathbf{Z}_{i-PA} = \delta\mathbf{Z}'_{PA}\mathbf{Z}_{PA} + 2\lambda(2\delta - 1) \quad (3)$$

Where $\alpha = 1/(0.5 - r_{PA}^2)$; and $\delta = (0.5 - r_{PA}^2)/(1 - r_i^2)$. r_{PA}^2 is defined as the reliability of the PA for individual i with parents 'sire' and 'dam', and can be calculated by: $r_{PA}^2 = \frac{r_{sire}^2 + r_{dam}^2}{4}$. Whilst r_i^2 the reliability of the EBV, was calculated as the square of the correlation between the true and predicted breeding values (r_i) according to (Butler et al., 2017):

$$r_i = \sqrt{1 - \frac{s_i^2}{(1+f_i)\sigma_A^2}} \quad (4)$$

Where s_i^2 is the prediction error variance for individual i ; f_i is the inbreeding coefficient for individual i calculated in ASReml-R v4.0 (Butler et al., 2017); and σ_A^2 is the additive genetic variance.

We can now complete and solve the coefficient matrix (Eq. 1) using equations 2 and 3, and multiply this by the vectors PA and EBV. The deregressed information regarding the individual without PA effects is obtained using this simplified formula:

$$\frac{y_{i-PA}}{\mathbf{Z}'_{i-PA}\mathbf{Z}_{i-PA}} \quad (5)$$

Cambial tissue was collected from the mature trees of the training population; this was an elegant solution to overcome the difficulty of obtaining foliage tissue from older/taller trees. Using a hammer and punch tool (approx. 2 cm diameter) two small circular disks of bark, cambium and developing tissue were removed from each tree. Once separated, the cambial tissue was immediately stored in a 2 ml

collection tube with 1 ml of storage buffer (10 mM EDTA pH 8.0, 10 mM Na₂SO₃), these were kept at 4 °C until DNA extraction. Foliage DNA extraction is easier using a standard protocol, therefore leaf bud tissue was collected from the juvenile trees. Two samples from each juvenile tree were taken and stored in the same way as the cambial tissue (in 1 ml of storage buffer and kept at 4 °C). The same DNA extraction protocol was used on both forms of tissue sample. This was a modified protocol developed by Ivanova et al. (2008) (R. Whetten, unpublished, North Carolina State University, personal communications). Whole exome capture genotyping was carried out in a commercial facility (RAPiD Genomics[®], Florida, US), with probes designed using the available Douglas-fir transcriptome assembly (Howe et al., 2013). For further details on the genotyping process see Thistlethwaite et al. (2017), both the training and validation populations were genotyped at the same time using the same procedure. For more information regarding exome capture, see Neves et al. (2013).

3.2.3 Estimated breeding value (EBV) prediction and accuracy

ASReml-R v4.0 (Butler et al., 2017) was used to fit EBVs, using information from the 11 F₁ parental sites, their parents (P₀) and the 2 F₂ juvenile sites (total N = 36,311) (Table 3.1). As environmental effects are an important consideration in forestry (Cappa et al., 2016), and to account for environmental (site) and age differences, a linear mixed model analysis was carried out,

$$y = X\beta + Z_1a + Z_2sa + Z_3s(rep) + Z_4sf + Z_5f + e \quad (6)$$

Where y is the phenotypic trait measurement; β is a vector of fixed effects (i.e., mean, site, age effects); a is a vector of individual random additive effects following $\sim N(0, A\sigma_a^2)$; sa is a site x additive genetic interaction following $\sim N(0, I\sigma_{sa}^2)$; $s(rep)$ is a vector of the block effect nested within site following $\sim N(0, I\sigma_{s(rep)}^2)$; sf is a random effect site x family interaction following $\sim N(0, I\sigma_{sf}^2)$; f is the effect of family representing 25% of dominance variance (Falconer and Mackay, 2009) and following $\sim N(0, I\sigma_f^2)$; and e is the random residual effect following $\sim N(0, I\sigma_e^2)$; X and Z_{1-5} are incidence matrices assigning fixed and random effects to each observation at vector y ; lastly I is the identity matrix and A

the average numerator relationship matrix (Wright, 1922). We chose to use a common variance for all environments since this is the most parsimonious model when using a large number of environments. This avoids over-fitting the model, since the number of parameters increases much faster than the number of environments (Isik et al., 2017). Theoretical accuracy of the EBVs (\hat{r}) was calculated following Dutkowski et al. (2002).

$$\hat{r} = \sqrt{1 - \frac{SE_i^2}{(1+F_i)\hat{\sigma}_a^2}} \quad (7)$$

Where SE_i is the standard error of breeding value, and F_i is the inbreeding coefficient of the i^{th} individual. Narrow-sense heritability was calculated as $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_{sa}^2 + \sigma_{sf}^2 + \sigma_f^2 + \sigma_e^2)$, where σ_a^2 , σ_{sa}^2 , σ_{sf}^2 , σ_f^2 and σ_e^2 are the variances of additive genetic, site x additive genetic, site x family, family, and residual effects, respectively.

ABLUP validation was carried out in ASReml-R v4.0, predicting the breeding values of the validation population using an expected relationship matrix (A) based on pedigree information. A 10-fold validation approach was used. Briefly, samples from the F1 parental generation at Adams, Fleet River and Lost Creek (N = 1,321) were randomly partitioned into 10 training subsets. 9 of these subsets (approximately 90% of the F1 samples) were used as the training set, on which the ABLUP model would be trained to estimate breeding values. On the basis of this model training, EBVs of the validation set were predicted. The validation set was composed of the 136 individuals from the F2 Jordan River site. This was repeated 10 times, until all F1 subsets had been included in the training set. Then the whole process was repeated 10 times, randomly assigning the training subsets. Prediction accuracy was measured as the correlation between the predicted EBVs from the cross-validation, and their original EBVs calculated using all possible pedigree information. In addition the predictive ability was calculated as the correlation between predicted EBVs and actual height phenotypes.

3.2.4 Genomic selection analysis

Three statistical methods were used to perform genomic selection: ridge regression (RR-BLUP), generalized ridge regression (GRR), and Bayes-B (Lorenz et al., 2011). Four GS analyses were performed: 1) models were trained on EBVs for juvenile height of the F₁ trees, genomic estimated breeding values (GEBVs) for height of the F₂ validation set were predicted (HTJ EBVs → HTJ GEBVs); 2) models were trained on DEBVs for juvenile height of the F₁ trees, genomic estimates of deregressed breeding values (GDEBVs) for height of the F₂ validation set were predicted (HTJ DEBVs → HTJ GDEBVs); 3) GS models were trained on EBVs of mature height (age 35) of the F₁ samples and GEBVs of the F₂ validation set were predicted and correlated with their juvenile EBVs to ascertain any relationship (HT35 EBVs → HTJ GEBVs); 4) models were trained on DEBVs for mature height of the F₁ samples, GDEBVs of the F₂ validation set were predicted and correlated with their juvenile DEBVs (HT35 DEBVs → HTJ GDEBVs). A 10-fold validation process repeated 10 times, was again used to randomly select individuals from the F₁ generation to construct the training set of the models. Prediction accuracy for each model in all 4 analyses was calculated as the mean of the replications of the Pearson product-moment correlation between the original EBVs (as calculated with all pedigree information N = 36,311) for HTJ of the 136 F₂ validation trees from Jordan River and their predicted GEBVs. Alternatively, using DEBVs to train the models, the prediction accuracy is the Pearson product-moment correlation between DEBVs of the validation set and their predicted GDEBVs. Similarly to the ABLUP investigation, the predictive ability was calculated as the correlation between predicted GEBVs and actual height phenotypes.

RR-BLUP

Ridge regression (RR-BLUP: (Whittaker et al., 2000)) was implemented using the R package ‘bigRR’ (Shen et al., 2014). The predicted heights (or GEBVs) are obtained by the summing of all the marker effects of an individual tree. Marker effects were estimated as in Henderson (1976), under the following mixed model:

$$y_{D(EBV)} = \vec{1}\mu + \mathbf{Z}g + e \quad (8)$$

where $y_{D(EBV)}$ is the vector of n tree height records (EBVs or DEBVs in this case), $\vec{1}$ is a vector of 1, μ is an intercept, g is the vector of random marker effects, \mathbf{Z} is the design matrix for the random marker effects, and e is the residual vector for random effects. In RR-BLUP the residuals and marker effects follow normal distributions with constant variance, i.e. $e \sim N(0, \mathbf{I}\sigma_e^2)$ and $g \sim N(0, \mathbf{I}\sigma_g^2)$, where \mathbf{I} is an identity matrix. The solution for the marker effects is given by the following equation:

$$\hat{g} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1} \mathbf{Z}'y \quad (9)$$

Where $\lambda = \sigma_e^2/\sigma_g^2$ is the ridge penalization parameter. An assumption of this method is that all marker effects are distributed equally, and therefore all effects are equally shrunk towards zero.

GRR

Generalized ridge regression (GRR) was implemented in the R package ‘bigRR’ (Shen et al., 2014). The first step, in this two-step variable selection method, is to use linear mixed models optimising λ , to estimate marker effects (the same as RR-BLUP). Where it differs from RR-BLUP is in a second step. In which an alternative, marker specific, shrinkage parameter is imposed on the BLUP for \hat{g} . In this heterogeneous error model, $\lambda\mathbf{I}$ becomes $\text{diag}(\lambda)$ in equation (4):

$$\hat{g} = (\mathbf{Z}'\mathbf{Z} + \text{diag}(\lambda))^{-1} \mathbf{Z}'y \quad (10)$$

Here λ is a vector of p shrinkage parameters. For the k^{th} element: $\lambda_k = \widehat{\sigma}_e^2/\widehat{\sigma}_{gk}^2$, is the parameter, where $\widehat{\sigma}_{gk}^2$ is the variance of marker effect k ($\widehat{\sigma}_{gk}^2 = \hat{g}_k^2/(1 - h_{kk})$). Where \hat{g} is the BLUP marker effect (from step 1), and h_{kk} is the effect of the dependant variable on the fitted value for observation k . To wit, h_{kk} represents the diagonal element ($n+k$) of the influence matrix $\mathbf{H} = \mathbf{T}(\mathbf{T}'\mathbf{T})^{-1} \mathbf{T}'$, and:

$$\mathbf{T} = \begin{pmatrix} \vec{1} & \mathbf{Z} \\ 0 & \text{diag}(\lambda) \end{pmatrix} \quad (11)$$

Bayes-B

Bayesian methods, as first proposed by Meuwissen et al. (2001), seek to relax the assumption that genetic effects are evenly distributed across the genome (as in RR-BLUP). In this analysis, we use Bayes-B, another variable selection method, in which there exists a probability that a marker has no effect (π). This would correspond to a situation where the genetic architecture of the trait was such that genetic variance was present at few, major effect loci only (Heffner et al., 2009; Lorenz et al., 2011). Bayes-B is thought to be a more realistic prior, since some genomic regions will be absent of QTL (Heffner et al., 2009). An assumption of this model is that marker effects are normally distributed with zero mean and finite variance. The prior distribution of the marker effect variance, is a mixture of two finite prior densities: $\text{var}(g) = 0$, with probability π ; and $\text{var}(g) \sim \chi^2(v, S)$, with probability $(1-\pi)$ (Gezan et al., 2017; Lorenz et al., 2011). π is assumed known and specified arbitrarily, the default value of 0.5 was used. Bayes-B analysis was carried out in the R package BGLR v1.0.4 (Perez and de los Campos, 2014), with the Gibbs sampler run for 100,000 iterations and a burn-in of 20,000, with a thinning rate of 100. Default rules of the BGLR R package (Perez and de los Campos, 2014) were used for the initial hyperparameter values.

3.3 Results

3.3.1 Heritability and EBV accuracy

Juvenile height (HTJ) heritability was estimated using a pedigree-based relationship matrix (ABLUP), including individuals from the 11 parental (F_1) sites, their parents (P_0), and the 2 progeny (F_2) sites. A summary of the contribution of each site to the both the ABLUP and GS models is in Table 3.1, along with site heritabilities for HTJ which ranged from 0.09 to 0.39. The overall HTJ heritability estimate was 0.14 (SE 0.025). The average theoretical accuracy for the EBVs of the sampled, genotyped individuals (from 3 parental sites: Adams, Fleet River and Lost Creek; and 1 progeny site: Jordan River) was 0.68, and 0.61 for the validation site (Jordan River) alone.

3.3.2 Validation in the progeny generation

3.3.2.1 ABLUP

The average prediction accuracy for Jordan River EBVs derived from ABLUP was 0.92 (SE 0.001) (Table 3.2), using a pedigree including genotyped samples only, i.e. the validation set and the F₁ samples from the 3 sites (Adams, Fleet River and Lost Creek). Using a larger pedigree based on the full 11 F₁ sites (plus their parents), and 2 progeny sites, the prediction accuracy becomes 0.95 (SE 0.0005). Whilst the pedigree for the ABLUP analyses included multiple generations, phenotypic information from the F₁ generation only was used to predict the validation (F₂) EBVs in the cross-validation. For comparison with the GS analyses, it is these results that we shall concentrate on. The average predictive ability for HTJ using ABLUP was calculated as the correlation between EBVs and juvenile height measurements. The predictive ability of ABLUP in the validation set was calculated as 0.71 (SE 0.003) using a pedigree including genotyped samples only, i.e. the validation set from Jordan River (F₂) and the F₁ samples from the 3 sites: Adams, Fleet River and Lost Creek (Table 3.3). Higher than the theoretical accuracy (0.61) calculated for the validation site (Jordan River) alone.

3.3.2.2 GS Scenario 1: HTJ EBVs → HTJ GEBVs

Prediction accuracies for GEBVs derived from RR-BLUP, GRR, and Bayes-B are shown in Fig. 3.1. In the first cross-generational GS analysis, GS prediction accuracies for validation were determined by the correlation between EBVs and GEBVs for the F₂ validation (progeny generation) set at Jordan River. GS prediction accuracies with models trained on F₁ EBVs were very similar over all GS methods used (Table 3.2), the average was 0.91. Their corresponding predictive abilities were also similar to each other with an average of 0.43, somewhat lower than that of ABLUP despite similar prediction accuracies (Table 3.3).

As is evident from the figures 3.1a-c, there is a pattern of distinct grouping characterised by those individuals with EBVs over 20 and those below. We suspected this was the major factor in causing such high prediction accuracies. To this end we further analysed these groups, or clusters, separately,

each with 3 different training set combinations: $EBV < 20^1$ represents the analysis of validation individuals within the lower cluster, using all genotyped F_1 individuals (1321 trees from Adams, Fleet River, and Lost Creek) as the training set; $EBV < 20^2$ is the analysis of validation individuals within the low cluster, using all genotyped F_1 individuals minus the parents of the high cluster as the training set (total= 1104); $EBV < 20^3$ is the analysis of validation individuals within the low cluster, using only information from their parents (in all crosses) from the F_1 generation as the training set (total= 132). Similarly: $EBV > 20^1$ represents the analysis of validation individuals within the higher cluster, using all genotyped F_1 individuals as the training set (total= 1321); $EBV > 20^2$ is the analysis of validation individuals within the high cluster, using all genotyped F_1 individuals minus the parents of the low cluster as the training set (total= 1189); and $EBV > 20^3$ is the analysis of validation individuals within the high cluster, using only information from their parents (in all crosses) from the F_1 generation as the training set (total= 217).

We subsequently found more limited correlations between EBVs and GEBVs within these clusters. When analysed separately the lower cluster, with individuals with EBVs lower than 20 (hereafter referred to as $EBV < 20$), observed similar yet moderate predictive accuracies across all training set combinations and GS statistical methods (Table 3.2), obtaining an average of 0.42. GS method RR-BLUP and training set $EBV < 20^3$ performed only slightly better than the other combinations. The higher cluster, the group with individuals with EBVs over 20 (hereafter referred to as $EBV > 20$), observed markedly low prediction accuracies for $EBV > 20^1$ and $EBV > 20^2$ (average: -0.05). Yet with a training set comprised only of the parents of this cluster ($EBV > 20^3$), that rose to an average of 0.68 over all GS methods (Table 3.2). Bayes-B in this case outperformed the other GS methods significantly, and indeed provided the best prediction accuracy in this investigation (0.99), greater than that of the ABLUP model.

3.3.2.3 GS Scenario 2: *HTJ DEBVs* → *HTJ GDEBVs*

The second GS analysis used the correlation between DEBVs and GDEBVs, of the F_2 validation set at Jordan River, as an indication of predictive accuracy. Deregression was carried out using the procedure

of Garrick et al. (2009), to account for family means and their effect on EBVs. The resulting deregressed EBVs (DEBVs) contain information regarding individuals only, without influence from parental BVs. Model accuracy fell dramatically when trained on DEBVs of the F₁ training set rather than on EBVs, 0.10 (SE 0.008) for RR-BLUP (Fig. 3.1d), 0.05 (SE 0.007) for GRR (Fig. 3.1e), and 0.48 (SE 0.002) for Bayes-B (Fig. 3.1f). The Bayes-B analysis being noticeably higher than the other two analyses due to the aforementioned clustering structure of the validation set. Similarly the predictive abilities of these three models were much lower for RR-BLUP and GRR, and moderate for Bayes-B: 0.06 (SE 0.008), 0.009 (SE 0.007), and 0.40 (SE 0.003) respectively (Table 3.3). In this scenario Bayes-B seemed to outperform the other two GS models, however the appearance of clustering in Fig. 3.1f called for further investigation.

Again, separate analyses were carried out according to the grouping of individuals mentioned previously. Although it is not immediately obvious in figures 3.1d-e, there is the emergence of a grouping pattern in figure 3.1f. The average predictive accuracy for EBV<20 across all training set combinations and GS methods was -0.09 and average predictive ability for EBV<20 was -0.19. For EBV>20 the average predictive accuracy of the training set combinations EBV>20¹ and EBV>20² was 0.17. This contrasts the results using only the parents of this high cluster as the training set, which gives an average predictive accuracy of -0.12 (Table 3.2). The average predictive ability for EBV>20 was 0.18 (Table 3.3).

3.3.2.4 GS Scenario 3: HT35 EBVs → HTJ GEBVs

The third GS analysis, using EBVs for mature height (age 35) of the F₁ samples as the training set for the models, and juvenile height GEBVs of the F₂ generation at Jordan River as the validation set. The prediction accuracies achieved were similar between all three GS models, with an average of 0.57 (Fig. 3.1g-i). They were comparable but lower than the result of the ABLUP analysis in which HT35 was used to predict HTJ EBVs (0.60, SE 0.010) (Table 3.2). Likewise their predictive abilities were generally

quite similar to that of the ABLUP analysis (0.24, SE 0.005) with an average of 0.21 across all GS methods (Table 3.3).

The appearance of groups again appears in figures 3.1g-i as a result of individual EBV in the validation set. For the group EBV<20 the prediction accuracy was on average 0.12 for EBV<20¹ and EBV<20². However this increased to an average of 0.37 using a training set made up of the parents of the low cluster only. Prediction accuracy for the EBV>20 cluster was more consistent across all training sets and GS methods (Table 3.2), with an average of -0.25. Although in this latter case the predictive accuracy correlation is driven by the presence of two smaller sub-groups within the EBV>20 group. These sub-groups have limited structure on their own.

3.3.2.5 GS Scenario 4: HT35 DEBVs → HTJ GDEBVs

Finally in the fourth analysis, using DEBVs for mature height of the F₁ samples as the training set for the models, and GDEBVs for juvenile height of the F₂ generation at Jordan River as the validation set; the average prediction accuracy or correlation between DEBVs and GDEBVs was -0.05 across all GS methods (Fig. 3.1j-l). The average predictive ability for this scenario was -0.09. All results were much lower than those for the ABLUP analysis for HT35 predicting HTJ EBVs (Tables 3.2 and 3.3).

The same groupings as before were tested for their prediction accuracies within the fourth analysis. The predictive accuracies for EBV<20 were generally very low across all training set combinations and GS methods. The average for EBV<20¹ and EBV<20² was -0.06. With the direction of the linear relationship changing when only the EBV<20 parents made up the training set (Table 3.2). The average for this training set combination over all GS methods was 0.06. Their average predictive abilities were: -0.16 for EBV<20¹ and EBV<20²; with a slight drop in predictive ability in EBV<20³ to -0.04. For EBV>20 the predictive accuracies were similar in magnitude for all training set combinations and GS methods, with an average of 0.02. However, whilst RR-BLUP and GRR gave slight positive correlations, Bayes-B results were all negative in their direction. All within cluster results were close to zero for this scenario, although the variation in the sign of the linear relationship causes some doubt as

to their reliability. The predictive abilities for EBV>20 were on average 0.08, predictably with the limited training set of EBV>20³ having slightly higher predictive abilities than the other training sets (for RR-BLUP and GRR only).

3.4 Discussion

3.4.1 Heritability of juvenile height

The estimated heritability of juvenile height was substantially low (0.14), but only slightly lower than previous studies suggest for Douglas-fir given the age of the trees used in the estimation (Dean and Stonecypher, 2006; Thistlethwaite et al., 2017; Ukrainetz et al., 2008; Yeh and Heaman, 1982). However, it should be noted that information from many more individuals was used to estimate this heritability ($N = 36,311$), increasing the environmental variance (error term), effectively shrinking the heritability. Using only the 1,321 F_1 trees, the heritability estimate is 0.17. The effect of heritability appears to be minimal in this case, since high predictive accuracies were obtained for validation models for HTJ predicting HTJ: 0.92 for ABLUP, 0.92, 0.91 and 0.91 for RR-BLUP, GRR, and Bayes-B respectively (Figs. 3.1a,b and c), using EBVs as the GS model input; and for HT35 predicting HTJ: 0.60 for ABLUP, 0.57, 0.56, and 0.58 for RR-BLUP, GRR, and Bayes-B respectively (Figs. 3.1g,h and i). The negligible effect of heritability in this instance is likely a consequence of the large sample size and low N_e used in the present investigation (Meuwissen et al., 2001).

3.4.2 Pedigree vs. marker based models

Although some results were similar, most predictive accuracies and especially predictive abilities for GS models trained on EBVs were still lower than those of ABLUP (Tables 3.2 and 3.3), a common issue in forestry (Bartholomé et al., 2016). It is thought that since the Douglas-fir genome is so exceedingly large and complex (Neale et al., 2017), many more markers will be needed in order to track LD between sources of variation and markers (Thistlethwaite et al., 2017). Simulations have yielded evidence that GS prediction accuracy increases markedly with marker density at least up to $8N_e$ /Morgan (Solberg et

al., 2008). In this study we have been moderately successful in capturing both contemporary and historical pedigree information. Although this is somewhat driven by groupings in the validation generation (F_2) characterised by high and low juvenile height values. Further investigation led to the discovery that the ultimate cause of this clustering was familial grouping. Each cluster contains whole families which can be traced back to distinct parents in the previous generation (F_1). There are no parents who are represented in both groups. In addition to this, the average heights and EBVs of the parents of the 'high' cluster ($EBV > 20$) were 765.04 cm and 60.70 respectively; are indeed higher than those parents of the 'low' cluster ($EBV < 20$), 709.88 cm and 0.81 respectively.

Analysis within each cluster provided mixed results. For those scenarios in which the GS models were trained on EBV data (scenarios 1 and 3), the within-cluster results were always lower than the overall predictive accuracy (and ABLUP predictive accuracy) with one exception. Convincing evidence that relationship tracking was the major driving force behind the strong overall correlations. Again within these two scenarios, training set composition had an effect. Unsurprisingly the analyses with training sets comprised of parents of the cluster in question only, generally gave higher prediction accuracies due to the high relationship between the two sets. Indeed the only scenario in which the GS prediction accuracy surpassed the ABLUP prediction accuracy was for the correlation between HTJ EBVs and HTJ GEBVs with the limited training set $EBV > 20^3$ and using Bayes-B as the GS method (0.99 vs. 0.92 for ABLUP). Although in some cases, the effect of training set composition was minimal (scenario 1: $EBV < 20$ and, Scenario 3: $EBV > 20$) (Table 3.2). Taking into account their clustering, predictive abilities for scenarios 1 and 3 fell far short of their ABLUP predictive abilities (Table 3.3), this could be an effect of a reduction in training population size. With the exception of the correlation between HTJ and HTJ GEBVs, which showed moderate success with a value of 0.59. This yet still falling short of the ABLUP value of 0.71. Within cluster correlations between HTJ and HT35 GEBVs (scenario 3) were all negative, whilst the overall correlations were marginally positive due to the structure of the data as a whole.

Following the removal of family means by deregression (scenarios 2 and 4), the lack of available marker-QTL LD failed to raise the GS prediction accuracies enough to even approximate those of ABLUP. However, Bayes-B notably performed better than both RR-BLUP and GRR in scenario 2 using the correlation between HTJ DEBVs and HTJ GDEBVs of the whole F₂ validation set as an indication of prediction accuracy (Table 3.2). A fundamental difference between the GS methods used is that, as opposed to RR-BLUP and GRR, Bayes-B gives different variances to each locus (including zero), therefore it allows for more weight to be put on the causative SNPs. As is evident from the results here, this drives up the prediction accuracy (Meuwissen et al., 2001). No such differences were seen between the GS methods subsequent to cluster analysis in these scenarios. When analysed to their full extent given the aforementioned data groupings, these deregressed analyses showed a dramatic drop in prediction accuracy. Suggesting the strong correlation was merely driven by tracking family means. Meanwhile, their predictive abilities show some striking trends. In both deregression scenarios (2 and 4) the EBV<20 cluster has only negative prediction abilities, averaging -0.19 and -0.12 for scenarios 2 and 4 respectively. Conversely, the EBV>20 cluster has only positive prediction abilities (with one exception), averaging 0.18 and 0.08 for scenarios 2 and 4 respectively. Discrepancy in the direction of the linear relationship casts some doubt on the reliability of the within cluster results. A likely symptom of the reduced sample size necessitated for these analyses.

Although it is disappointing to not have captured enough LD to raise the prediction accuracy above ABLUP in a diverse validation population, in terms of real-world application there is a positive finding. This use of marker-based selection nevertheless reduces the need for time-intensive practices such as performing specific crosses and building a structured pedigree (El-Kassaby and Lstibůrek, 2009). Potentially quickening the breeding process, and perhaps off-setting costs involved in genotyping.

3.4.3 GS across generations

Once GS becomes a viable option for tree breeders it will likely, in most cases, be deployed to select progeny of the training population without the need for explicit crosses and a lengthy testing phase.

Bearing this in mind, it is important to validate models across generations rather than cross-validate within the same generation, as many have done before.

GS relies on LD between markers and causal gene variants. LD breaks down after every breeding cycle due to recombination. This is especially pertinent to breeding within the forestry sector, for forest tree species have lower observed levels of LD (Neale and Kremer, 2011). Given these circumstances, large SNP sets will be required to provide dense coverage of the genome to find LD between SNPs and QTL (Jaramillo-Correa et al., 2015). In a simulation study it was shown that higher marker densities allowed prediction accuracies to be maintained over longer generational gaps (Müller et al., 2017b). Thus, without addressing this issue, predictive accuracy is expected to fall dramatically (Atefi et al., 2016; Habier et al., 2007).

There are a reported 54,830 gene models in the Douglas-fir genome (Neale et al., 2017). With our 69,551 SNPs, using juvenile height of an F_1 generation as the training set and an F_2 generation as the validation set, we obtained an overall predictive accuracy of 0.92, 0.91, and 0.91 for RR-BLUP, GRR, and Bayes-B respectively for juvenile height. This is in line with results from a previous study of GS in Douglas-fir (0.79 - 0.92) using cross-validation within the same generation (Thistlethwaite et al., 2017).

The results we present here did not show a drop in predictive accuracy as one would expect (with a caveat described later). As has been shown before, prediction accuracy increases when the genetic relationship between training and validation sets is closer (Habier et al., 2007; Lorenz et al., 2012; Sallam et al., 2015). Indeed, it is the case here that our validation set is closely related to the training set by virtue of being their (the training set's) offspring. Although our results are not as disparate, Bartholomé et al. (2016) observed a similar trend. When progeny validation was used, prediction accuracies of GS methods were higher (0.70 using genomic BLUP (GBLUP), and 0.71 using Bayesian LASSO (B-LASSO)) than when within generation cross-validation was carried out (0.66, for both GBLUP and B-LASSO). With that in mind, there may be some benefit in combining phenotypes across generations which helps to infer the Mendelian sampling term.

However, as is evident from Figures 3.1a-c g-i, two distinct data groups had arisen which had caused the high correlations. These two groups were characterised by a) those individuals with low juvenile height values (mean= 641.81 cm), and b) those with high juvenile height values (mean= 696.79 cm). They were later found to be the result of familial clusters, with no overlapping parents from group to group. The estimation of breeding values exacerbated this trend and thus two non-overlapping groups can be seen defined as $EBV < 20$ (individuals with EBVs less than 20) and $EBV > 20$ (individuals with EBVs over 20).

Given this data structure further analysis showed that prediction accuracies for scenario 1 $EBV < 20$ were in fact only moderate, with an average of 0.42 for all GS methods compared to ABLUP (0.92). In the case of $EBV > 20$ scenario 1 results were close to zero with the exception of $EBV > 20^3$. High prediction accuracies in this case were driven by close relationships between the training and validation sets and small sample size.

These results might possibly be improved by re-estimation of EBVs using a realised relationship matrix (**G**-matrix) rather than the estimated relationship **A**-matrix used here (Munoz et al., 2014).

Prediction accuracies for deregressed values reported here, align with those reported in an earlier investigation also carried out on Douglas-fir. Thistlethwaite et al. (2017) describe obtaining prediction accuracies for GS models, trained on DEBVs of height at age 12, that were also approximately 0. These models were cross-validated with individuals from within the same generation. Here, using similar parameters but using a cross-generational validation process, we have obtained similar results with a few exceptions. Once data grouping was taken into account, correlations between HTJ DEBVs and HTJ GDEBVs (scenario 2) were low-moderately negative, -0.12 - -0.02 for $EBV < 20$ and -0.23 - 0.26 for $EBV > 20$.

Although a lower predictive accuracy for F_2 HTJ was obtained when the GS models were trained on F_1 mature height (age 35) (scenario 3), there was still a significant positive correlation between the two (0.57, 0.56, and 0.58 for RR-BLUP, GRR, and Bayes-B respectively). These results are lower but similar to findings in Thistlethwaite et al. (2017) where a positive time-time correlation between juvenile

(age 12) and mature (age 35) height within the same generation was found (0.71 (SE 0.0004) for both RR-BLUP and GRR). However here again, the groups EBV<20 and EBV>20 were the driving force behind these correlations. For scenario 3 EBV<20 results were likewise reduced, averaging 0.12 for EBV<20¹ and EBV<20², and 0.37 for EBV<20³ across all GS methods. The stronger correlation coming from the increased relationship between the training and validation sets, yet falling quite short compared to ABLUP (0.60). In scenario 3 the EBV>20 accuracy results show a small-moderate but negative correlation averaging -0.25 across all GS methods. Marker-trait associations vary with tree age, and predictably the ABLUP correlation here was found to be lower than in scenario 1 (0.92) at 0.60. The high positive EBV>20³ juvenile-juvenile correlation and small-moderate negative EBV>20 mature-juvenile correlation, may also be symptoms of the effect of tree age on marker-trait associations (Lerceteau et al., 2001) and as seen here, recombination. These results do not provide a sound basis on which to perform selection decisions. However should there be an improvement in GS accuracy with the re-estimation of EBVs, useful information may still be procured from moderate correlations, for input into early selection decisions.

In the final scenario, deregressed breeding values for mature height in the F₁ generation were used as training data for the GS models. Predictive accuracy fell significantly both for the validation set as a whole (-0.15 – 0.11), and for each of EBV<20 (-0.08 – 0.07) and EBV>20 (-0.07 – 0.08).

In the absence of average parental information (i.e. deregressed phenotypes), the ability of the markers to predict phenotypes in the next generation was consistently poor (Table 3.2). None of the GS analyses in these deregressed scenarios (2 and 4) surpassed or even matched the predictive ability of the ABLUP models, despite having very similar overall prediction accuracies in scenarios 1 and 3 (although this itself can be attributed to the effects of clustering in the validation set). Predictive ability similarly dropped dramatically after deregression and re-adjusting for clustering. The opposing signs of coefficients for the high and low clusters after deregression suggest some unreliability of these values. This trend is reflected in throughout, thus any of these within cluster GS ‘accuracies’ should be treated with caution.

3.4.4 Main factors that affect GS: relatedness and LD

Whilst some of our observations during this study align with previously published work, the high predictive accuracies using EBVs as the model input are a result of tracking family means as opposed to true LD being captured. GS BLUP methods are robust in most circumstances and perform well; however, it is now known that these methods primarily capture marker derived relatedness more readily than actual LD (Habier et al., 2007; Zhong et al., 2009). We base this viewpoint on trends seen in Thistlethwaite et al. (2017) whereby removing family derived means from trait breeding values, caused virtually no prediction accuracies of GS methods. Simulation evidence suggests that pedigree based relationships contribute to predictive accuracy for a few generations, especially given low N_e (Müller et al., 2017b), which the progeny validation set presented here certainly does have. In empirical studies, there is also evidence that the relatedness of the training and validation sets has an impact on the accuracy of GS (Bartholomé et al., 2016; Beaulieu et al., 2014a; Märtens et al., 2016; Resende et al., 2012a; Varshney et al., 2017). With higher relatedness between sets producing more accurate predictions. Our training and validation set are only a generation apart; it is therefore likely that their relatedness is the main force behind the GS prediction accuracies. Even though the alternative, more computationally intense GS method of Bayes-B was employed to help overcome this (Habier et al., 2007), we have not been able to resolve any additional LD using this method. In order to ‘reveal’ the effect of LD on GS predictive accuracy, empirical investigations on multiple generations further apart will have to be conducted using much denser SNP genotyping.

Apropos of marker density, we carried out some preliminary analyses concerning the effect of marker density on the prediction accuracy of GS. We carried out a 10-fold cross-validation on HT35 EBVs within the genotyped parental (F_1) generation samples using RR-BLUP. Randomly selected marker sets were chosen that had progressive totals from 200-50,000 SNPs. These sets were tested and replicated 10 times, with random SNP sampling for each repetition. Our investigations led us to the conclusion that an increase in marker density leads to an increase in GS prediction accuracy. A caveat

being that the magnitude of prediction accuracy gains falls to a plateau as the number of markers approaches approximately 15,000 SNPs. This also happens to be the point at which the prediction accuracy is similar to the ABLUP prediction accuracy for HT35 EBVs for these samples. We believe that relatedness is the driving-force behind these additional GS predictive accuracy results for the HT35 EBVs for a couple of reasons. Firstly the variance within each SNP set total was modest even though the SNPs were randomly selected. Secondly the prediction accuracies fell dramatically when the EBVs were deregressed to remove the parental average effects, for the purposes of extricating LD from pedigree in a previous study (Thistlethwaite et al., 2017). Given this insight, in order to capture short-range marker-QTL LD in conifer species, we recommend the use of deregression to remove parental average effects and high density marker sets.

The effect of capturing marker derived relatedness may lead to sibling co-selection due to an increased correlation between EBVs within families (Wray and Thompson, 1990). In the long-run this will lead to a reduction in genetic variation in subsequent generations, and a loss of potential genetic gain (Hallander, 2009). A reduction in Mendelian segregation variance due to inbreeding (sibling co-selection) will eventually lead to a decrease in the Mendelian sampling term of each individual, and population-wide additive variance. Presenting a problem for breeders in that the expectation of the rate of genetic gain in a population is proportional to the Mendelian sampling term of selection targets (Avendano et al., 2004; Woolliams and Thompson, 1994).

Breeding programmes often implement selection methods that optimise the selection differential whilst constrained by a limit on increasing coancestry. Hallander and Waldmann (2009) tested such methods on a diallel progeny trial of Scots pine (*Pinus sylvestris* L.) investigating height and stem diameter, and found optimum contribution (OC) dynamic selection resulted in the highest genetic gain over other methods (standard restricted selection). One of the main drivers for the success is that the selective advantage in the OC method is the Mendelian sampling term. The OC method maximises the selection differential between families by using best estimates of the Mendelian sampling term for each

tree when calculating the mating contributions (Avendanõ et al., 2004). Selection methods must be considered when using GS results as a basis for selection, to avoid inbreeding and to maintain Mendelian segregation variance.

3.5 Summary

Here we perform cross-generational GS analysis on coastal Douglas-fir (*Pseudotsuga menziesii*), reflecting trans-generational selective breeding application. 1,321 trees, representing 37 full-sib F_1 families from 3 sites in British Columbia, Canada, were used as the training population for: 1) EBVs (estimated breeding values) of juvenile height (HTJ) in the F_1 generation predicting genomic EBVs of HTJ of 136 individuals in the F_2 generation, 2) deregressed EBVs of F_1 HTJ predicting deregressed genomic EBVs of F_2 HTJ, 3) F_1 mature height (HT35) predicting HTJ EBVs in F_2 , and 4) deregressed F_1 HT35 predicting genomic deregressed HTJ EBVs in F_2 . Ridge regression best linear unbiased predictor (RR-BLUP), generalized ridge regression (GRR), and Bayes-B GS methods were used and compared to pedigree-based (ABLUP) predictions. GS accuracies for scenarios 1 (0.92, 0.91, and 0.91) and 3 (0.57, 0.56, and 0.58) were similar to their ABLUP counterparts (0.92 and 0.60 respectively) (using RR-BLUP, GRR, and Bayes-B). Results using deregressed values fell dramatically for both scenarios 2 and 4 which approached zero in many cases. Cross-generational GS validation of juvenile height in Douglas-fir produced predictive accuracies almost as high as that of ABLUP. But without capturing LD, we cannot surpass the ABLUP prediction. Here we tracked pedigree relatedness between training and validation sets. More markers or improved distribution of markers, are required to capture LD in Douglas-fir. This is essential for accurate forward selection among siblings as makers that track pedigree are of little use for forward selection of individuals within controlled pollinated families.

Table 3.1 Summary of the number of individuals per site, how many were genotyped, to which generation they belong, and to which models they contributed; as well as the within-site heritability estimates for juvenile height (HTJ).

Site	n	Number genotyped	Generation	Model contribution	Heritability (h^2)
Wild progenitors	108		P ₀	ABLUP ¹	NA
Adams	3,478	449	F ₁	ABLUP, GS	0.19
Fleet River	2,944	441	F ₁	ABLUP, GS	0.22
Lost Creek	3,244	431	F ₁	ABLUP, GS	0.13
Sechelt	2,909		F ₁	ABLUP	0.20
Squamish River	3,153		F ₁	ABLUP	0.25
Eldred River	3,395		F ₁	ABLUP	0.13
Tansky Creek	2,974		F ₁	ABLUP	0.27
Sproat Lake	2,881		F ₁	ABLUP	0.20
White River	3,010		F ₁	ABLUP	0.17
Gold River	3,067		F ₁	ABLUP	0.09
Menzies	2,876		F ₁	ABLUP	0.17
Jordan River	247	136	F ₂	ABLUP, GS	0.39
North Arm	2,025		F ₂	ABLUP	0.31

¹Pedigree only, “GS” includes all three genomic selection methods (RR-BLUP, GRR, and Bayes-B), and only genotyped individuals were used in the construction and validation of these models. Heritability was calculated within sites (with n individuals), using a full pedigree containing all generations.

Table 3.2 Prediction accuracies of four models using ABLUP and three genomic selection (GS) statistical methods (RR-BLUP, GRR, and Bayes-B).

Analysis	Accuracy (SE)			
	ABLUP	RR-BLUP	GRR	Bayes-B
HTJ EBVs → HTJ GEBVs	0.92 (0.001)	0.92 (0.0002)	0.91 (0.0003)	0.91 (0.0007)
EBV<20 ¹		0.43 (0.003)	0.42 (0.003)	0.42 (0.004)
EBV<20 ²		0.43 (0.004)	0.38 (0.005)	0.38 (0.005)
EBV<20 ³		0.45 (0.006)	0.43 (0.006)	0.44 (0.005)
EBV>20 ¹		-0.005 (0.004)	-0.05 (0.004)	-0.02 (0.006)
EBV>20 ²		-0.04 (0.004)	-0.11 (0.004)	-0.10 (0.005)
EBV>20 ³		0.54 (0.009)	0.52 (0.010)	0.99 (0.0001)
HTJ DEBVs → HTJ GDEBVs		0.10 (0.008)	0.05 (0.007)	0.48 (0.002)
EBV<20 ¹		-0.12 (0.004)	-0.11 (0.004)	-0.02 (0.005)
EBV<20 ²		-0.12 (0.004)	-0.14 (0.004)	-0.09 (0.005)
EBV<20 ³		-0.08 (0.004)	-0.07 (0.004)	-0.07 (0.004)
EBV>20 ¹		0.17 (0.004)	0.26 (0.004)	0.16 (0.006)
EBV>20 ²		0.14 (0.004)	0.17 (0.004)	0.13 (0.006)
EBV>20 ³		-0.23 (0.003)	-0.22 (0.003)	0.09 (0.009)
HT35 EBVs → HTJ GEBVs	0.60 (0.010)	0.57 (0.002)	0.56 (0.002)	0.58 (0.003)
EBV<20 ¹		0.15 (0.004)	0.20 (0.004)	0.04 (0.007)
EBV<20 ²		0.13 (0.004)	0.18 (0.004)	0.04 (0.007)
EBV<20 ³		0.35 (0.005)	0.41 (0.005)	0.36 (0.007)
EBV>20 ¹		-0.21 (0.002)	-0.24 (0.002)	-0.28 (0.004)
EBV>20 ²		-0.22 (0.002)	-0.25 (0.002)	-0.30 (0.003)
EBV>20 ³		-0.23 (0.003)	-0.22 (0.003)	-0.27 (0.003)
HT35 DEBVs → HTJ GDEBVs		-0.15 (0.005)	-0.11 (0.005)	0.11 (0.007)
EBV<20 ¹		-0.08 (0.003)	-0.02 (0.003)	-0.02 (0.003)
EBV<20 ²		-0.08 (0.003)	-0.07 (0.003)	-0.06 (0.003)

Analysis	Accuracy (SE)			
	ABLUP	RR-BLUP	GRR	Bayes-B
EBV<20 ³		0.06 (0.002)	0.07 (0.003)	0.06 (0.003)
EBV>20 ¹		0.02 (0.004)	0.07 (0.004)	-0.06 (0.004)
EBV>20 ²		0.02 (0.005)	0.03 (0.006)	-0.07 (0.003)
EBV>20 ³		0.07 (0.004)	0.08 (0.004)	-0.01 (0.006)

ABLUP is a pedigree only model with no marker information used. Results are from the validation procedure replicated 10 times, in which a random 90% of the genotyped F₁ generation (1321 trees from Adams, Fleet River, and Lost Creek) was used as the training set and the validation set was comprised of the 136 genotyped F₂ trees from Jordan River. Accuracy was calculated as the mean of the replications of the Pearson product-moment correlation between the original EBVs for HTJ of the 136 F₂ validation trees from Jordan River and their predicted GEBVs or GDEBVs. The four analyses are: F₁ juvenile height EBVs predicting F₂ juvenile height EBVs (HTJ EBVs → HTJ GEBVs); F₁ juvenile height DEBVs predicting F₂ juvenile height DEBVs (HTJ DEBVs → HTJ GDEBVs); F₁ mature (age 35) height EBVs predicting F₂ juvenile height EBV (HT35 EBVs → HTJ GEBVs); and F₁ mature height DEBVs predicting F₂ juvenile height GDEBVs (HT35 DEBVs → HTJ GDEBVs). Results for the validation set as a whole are in bold (N = 136), following these are the results for each of the two clusters EBV<20 (N = 83) and EBV>20 (N = 53), with indices representing different training set composition: ¹ all genotyped F₁ individuals; ² all genotyped F₁ individuals minus the parents of the opposing cluster; ³ only the F₁ parents of the cluster in question (including all crosses to which they contribute).

Table 3.3 Corresponding predictive abilities for ABLUP and GS prediction accuracies in table 3.2, calculated as the correlation between the raw phenotype (juvenile height: HTJ) and their genomic estimated breeding values (GEBVs) or deregressed genomic estimated breeding values (GDEBVs).

Analysis	Predictive ability (SE)			
	ABLUP	RR-BLUP	GRR	Bayes-B
r (HTJ, HTJ GEBVs)	0.71 (0.003)	0.43 (0.0004)	0.42 (0.0006)	0.43 (0.0007)
EBV<20 ¹		0.10 (0.002)	0.07 (0.002)	0.07 (0.003)
EBV<20 ²		0.04 (0.002)	-0.01 (0.003)	-0.01 (0.003)
EBV<20 ³		-0.10 (0.003)	-0.12 (0.004)	-0.09 (0.003)
EBV>20 ¹		0.05 (0.002)	0.03 (0.002)	0.07 (0.004)
EBV>20 ²		0.04 (0.002)	0.01 (0.002)	0.07 (0.003)
EBV>20 ³		0.15 (0.004)	0.12 (0.005)	0.59 (0.0004)
r (HTJ, HTJ GDEBVs)		0.06 (0.008)	0.009 (0.007)	0.40 (0.003)
EBV<20 ¹		-0.20 (0.006)	-0.21 (0.005)	-0.04 (0.006)
EBV<20 ²		-0.22 (0.005)	-0.25 (0.005)	-0.15 (0.007)
EBV<20 ³		-0.21 (0.006)	-0.20 (0.006)	-0.21 (0.006)
EBV>20 ¹		0.23 (0.004)	0.31 (0.005)	0.20 (0.005)
EBV>20 ²		0.19 (0.005)	0.22 (0.005)	0.16 (0.006)
EBV>20 ³		0.15 (0.007)	0.11 (0.007)	0.07 (0.010)
r (HTJ, HT35 GEBVs)	0.24 (0.005)	0.21 (0.001)	0.21 (0.001)	0.22 (0.002)
EBV<20 ¹		-0.19 (0.001)	-0.13 (0.002)	-0.09 (0.005)
EBV<20 ²		-0.23 (0.001)	-0.17 (0.002)	-0.14 (0.004)
EBV<20 ³		-0.10 (0.002)	-0.07 (0.002)	-0.06 (0.004)
EBV>20 ¹		-0.03 (0.001)	-0.04 (0.001)	-0.09 (0.002)
EBV>20 ²		-0.03 (0.0009)	-0.05 (0.001)	-0.07 (0.002)
EBV>20 ³		-0.04 (0.001)	-0.04 (0.002)	-0.06 (0.001)
r (HTJ, HT35 GDEBVs)		-0.17 (0.005)	-0.13 (0.004)	0.04 (0.006)
EBV<20 ¹		-0.18 (0.003)	-0.13 (0.003)	-0.13 (0.003)

Analysis	Predictive ability (SE)			
	ABLUP	RR-BLUP	GRR	Bayes-B
EBV<20 ²		-0.18 (0.003)	-0.18 (0.003)	-0.18 (0.003)
EBV<20 ³		-0.05 (0.003)	-0.03 (0.004)	-0.05 (0.004)
EBV>20 ¹		0.08 (0.005)	0.13 (0.004)	0.003 (0.004)
EBV>20 ²		0.09 (0.005)	0.09 (0.006)	-0.01 (0.003)
EBV>20 ³		0.15 (0.005)	0.15 (0.005)	0.06 (0.006)

Results for the validation set as a whole are in bold (N = 136), below each are the results for each of the two clusters EBV<20 (N = 83) and EBV>20 (N = 53).

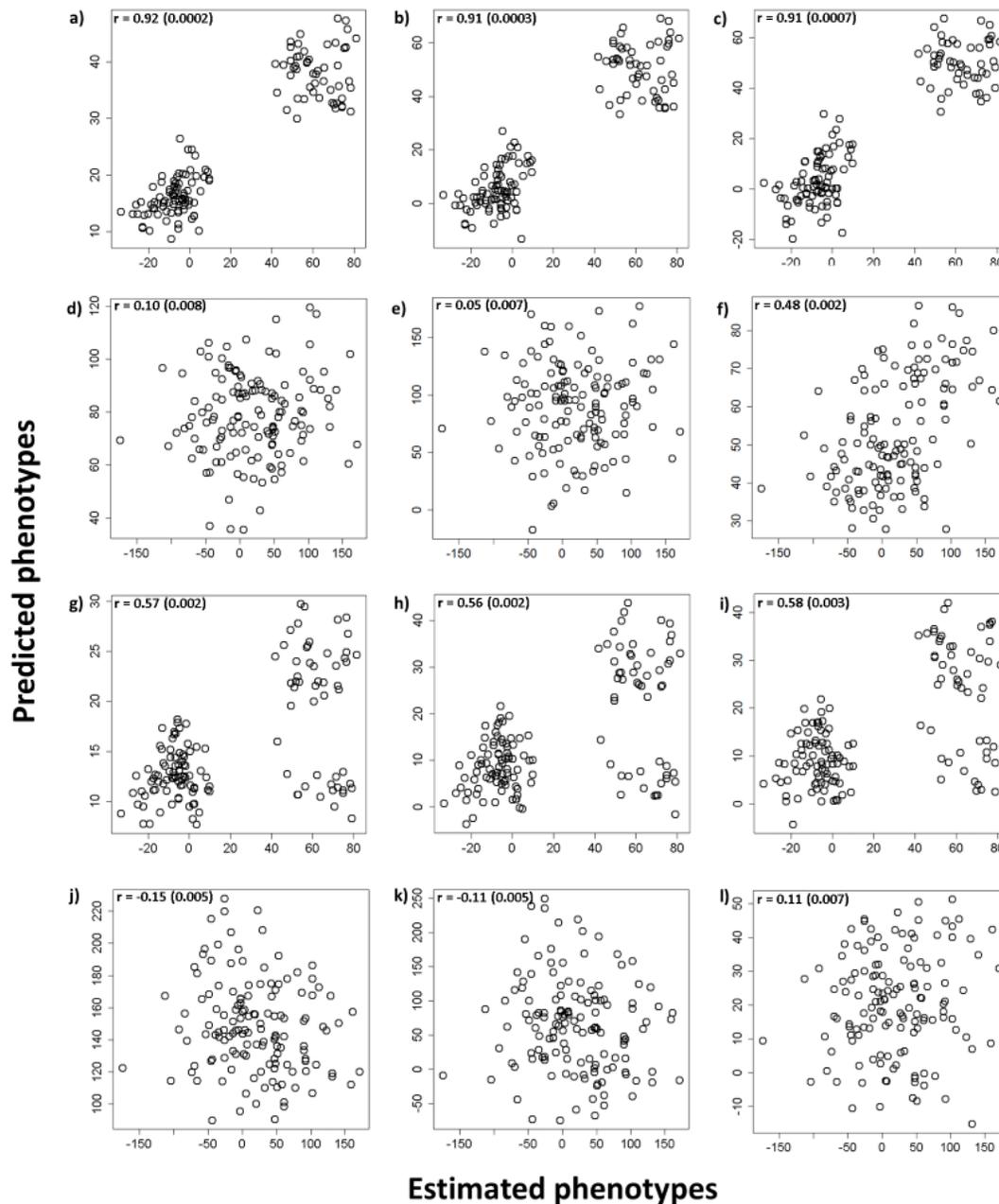


Figure 3.1 GS correlation in the validation set (Correlation in the validation set of: EBVs and genomic estimated breeding values (GEBVs) for juvenile height (HTJ) using a) RR-BLUP, b) GRR, and c) Bayes-B; deregressed estimated breeding values (DEBVs) and genomic deregressed estimated breeding values (DGEBVs) using d) RR-BLUP, e) GRR, and f) Bayes-B; EBVs for HTJ vs. GEBVs for mature height age 35 years (HT35) using g) RR-BLUP, h) GRR, and i) Bayes-B; and DEBVs for HTJ vs. DGEBVs for HT35 using j) RR-BLUP, k) GRR, and l) Bayes-B).

4 Linkage disequilibrium vs. pedigree: genomic selection prediction accuracy in two conifer species

4.1 Introduction

With genotyping costs at the lowest they've ever been (and a general decreasing trend), genomic selection (GS) becomes an ever-more viable option for forest tree breeders. This should subsequently result in beneficial gains to the industry in terms of improved wood quality, yield per unit time (generational turnover), and stress tolerance (biotic and abiotic) (Grattapaglia, 2014; Heffner et al., 2010b). In a deviation from marker assisted selection (MAS), rather than attempting to identify significant trait-loci relationships, GS employs all available marker information simultaneously to predict performance (Meuwissen et al., 2001). GS experimental results have been promising to date, with prediction accuracies higher than those of MAS (Jannink et al., 2010).

One of the major determinants of GS success is the relationship between effective population size (N_e) and marker density (Lin et al., 2014). Falconer and Mackay (2009) succinctly describe N_e as the number of randomly mating individuals that would cause the observed inbreeding rate of a population. As a result of inbreeding, which is more likely at low N_e , allele frequencies become skewed. In this situation (at low N_e) certain individuals have more of a chance to reproduce, causing genetic drift as their genes are passed more frequently onto the next generation (Lin et al., 2014). Over time and over multiple generations of these conditions those alleles may become fixed, as genetic diversity decreases (Falconer and Mackay, 2009). Low N_e populations are subject to stronger drift, which in turn is one of the major driving forces behind linkage disequilibrium (LD) between loci. Lower N_e populations have higher LD, and this LD between markers and trait QTLs is essential in the prediction of trait performance from markers (Lin et al., 2014). The success of GS is highly dependent on the marker-QTL LD and this is largely determined by the extent to which the training and validation populations are related. A certain amount of caution is also required during the implementation of GS as LD will decay subsequent to every round of breeding, due to recombination, although this can be overcome by employing dense marker arrays. It should be emphasized

that GS may require alternative delivery methods for tree improvement, as the current seed orchard production pathway and its sexual-production effectively breaks down the marker-QTL LD. Lin et al. (2014) point out that N_e can be artificially reduced by using half-sib or within-family populations. This is a good technique to use on outbred species, such as forest trees, when applying GS.

The determination of the number of markers required for GS is largely based on the occurrence of LD, which in turn is determined by population structure and N_e . Solberg et al. (2008) did some initial investigations into marker density effect, comparing two types of marker: SSR-like multiallelic markers versus SNP-like biallelic markers. They found a general increase in prediction accuracy as density increased in relation to N_e , however they failed to reach a plateau with the numbers available ($2N_e$ SSR markers per Morgan or $8N_e$ SNP markers per Morgan). Meuwissen (2009) proposed that a minimum of $10N_eL$ markers should be used to obtain accurate predictions in GS, where L is the total length of the genome in Morgans. As Lin et al. (2014) discuss, this should mean that for a population of outbreeding apples with an assumed N_e of 1000 and a genome of approximately 13 Morgans, 130,000 markers are necessary for accurate GS predictions. However Kumar et al. (2012) successfully carried out GS on apple variety *Malus domestica* Borkh with an accuracy of 0.7, with only 2,500 markers. This can be attributed to the bi-parental design used (Kumar et al., 2012). Within-family designs call for fewer markers since larger (but fewer) chromosomal segments are shared by family members, and it is these that need to be tracked (Hayes et al., 2009a). Grattapaglia and Resende (2011) using a deterministic approach found that for GS to be effective, a marker density of ~ 2 markers/cM is required when N_e is no greater than 30. Larger N_e may require up to 20 markers/cM. Yet higher densities may be required for situations in which the training and validation populations are not derived genetically from the same base population (Meuwissen, 2009). Using Grattapaglia and Resende's (2011) calculations, and an assumed map length of 2000 cM (Krutovsky et al., 2004; Pavy et al., 2008, 2017; Pelgas et al., 2011), we should aim to use 4000 markers for investigations into Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)) with an $N_e \sim 21$. By the same reasoning we should aim to use up to 40,000 markers for an Interior Spruce population (*Picea glauca* (Moench) Voss x *Picea engelmannii* Parry ex Engelm.), with $N_e \sim 93$. Indeed Howe et al. (2013) concluded that a density of 2.5

markers per cM (5000 SNPs/2000 cM), should provide effective GS results in populations no larger than $N_e \sim 30$. These numbers, for populations with low N_e , are in line Meuwissen's (2009) determination that $10N_eL$ markers should be used as a minimum, which gives us $10 \times 21 \times 20 = 4200$ markers for Douglas-fir. Yet using the same calculation from Meuwissen (2009) for Interior spruce, gives us a recommended minimum of 18,600 markers, less than half of that recommended by Grattapaglia and Resende's (2011) calculation.

Ma et al. (2016) investigated the effect of marker selection on prediction accuracy of GS in soybean (*Glycine max L.*). They tested 3 methods of marker selection: random sampling, haplotype block analysis and evenly sampling. They found that for plant height, only marginal differences in prediction accuracy were obtained with the three sampling methods. However for grain yield, the haplotype block analysis outperformed the other two methods by about 4%. This preselection method offers a comprehensive yet cost-efficient option for implementing GS, reducing the number of SNPs required to just 1 SNP per haplotype block plus those not contained within blocks. However an in-depth understanding of the structure and LD across the genome is required for this. For this reason, we have concentrated on the random sampling method only for Douglas-fir and Interior spruce.

The two species studied here are representative of full-sib and half-sib populations (Douglas-fir and Interior spruce respectively). Their differing pedigree structure should be reflected in the prediction accuracies we obtain through GS. Previously, using full-sib families, GS prediction accuracy has been shown to be moderate to high (Beaulieu et al., 2014b; Resende et al., 2012b), and in Douglas-fir specifically (Thistlethwaite et al., 2017). This is considered a result due to, primarily, long range LD arising from the increased levels of relatedness. Short range LD is not considered a strong influence in these circumstances. By comparison, using half-sib families has previously resulted in low to moderate GS prediction accuracies (Beaulieu et al., 2014a), and in interior spruce specifically (Gamal El-Dien et al., 2015; Ratcliffe et al., 2015). Higher N_e and subsequent low LD are thought to impede accuracy in these half-sib studies. Larger N_e leads to more recombination and therefore more diversity within a population. Drift associated with small populations is not significant under these conditions (open-pollinated and highly outcrossing species)

and subsequently does not significantly contribute to the build-up of LD. Nonetheless, as can be seen in Fig. 1.2, high GS prediction accuracies are not exclusive to full-sib studies and *vice versa*.

It is, however, becoming seemingly more apparent that perhaps LD is not the main driving force in all GS studies (Lenz et al., 2017; Thistlethwaite et al., 2017). Studies concerning those species with larger genomes may find that relatedness rather than LD is the compelling factor. As de los Campos et al. (2013) describe, the success of GS relies upon the similarity of the realised genomic relationships at the marker and QTL levels. The number of independently segregating segments determines the coefficient of variation of these relationships across the genome. In unrelated individuals this is a product of population-wide LD. Yet between family members, this is a product of within family disequilibrium. Here we investigate the effect of marker density on GS in two conifer species with differing pedigree structures.

4.2 Material and Methods

4.2.1 Experimental populations

Predictive models for GS were trained on two progeny testing populations. The first population consist of 38-year-old pedigreed coastal Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)). This population was originally established by the Ministry of Forests, Lands and Natural Resource Operations of British Columbia, Canada in 1975 and it is made up of 165 full-sib families (54 parents), from which 37 families were selected for sampling from three test sites (**Adams** (Lat. 50 24'42" N, Long. 126 09' 37"W, Elev. 576 mas), **Fleet River** (Lat. 48 39'25" N, Long. 128 05'05" W, Elev. 561 mas), and **Lost Creek** (Lat. 49 22'15" N, Long. 122 14'07" W, Elev. 424 mas)) giving a total of 1,372 trees (N≈500 per site).

The second population consisted of 1,126 38-year-old Interior spruce trees (*Picea glauca* (Moench) Voss x *Picea engelmannii* Parry ex Engelm.). This progeny test trial was established in 1972/73 by the Ministry of Forests, Lands and Natural Resource Operations of British Columbia Canada and is made up of 181 open-pollinated families, of which the best performing 25 families were selected based on their growth attributes. The trial is located on three sites (**Aleza Lake** (Lat. 54° 03' 15.7" N, Long. 122° 06' 35.4" W, Elev. 700 mas), **Prince George Tree Improvement Station (PGTIS)** (Lat. 53° 46' 17.9" N, Long.

122° 43' 07.6"W, Elev. 610 mas), and **Quesnel** (Lat. 52° 59' 27.2" N, Long. 122° 12' 30.6" W, Elev. 915 mas)).

4.2.2 Phenotyping and genotyping

Mid-rotation height measurements of the sampled trees were recorded: at age 35 for the Douglas-fir population, and at age 38 Interior spruce (HT: in meters). Estimated breeding values (EBVs) for HT were obtained in ASReml-R v4.0 (Butler et al., 2017) and used as phenotypes for the genomic prediction analysis. Wood density (WDres) measurements for the Douglas-fir population were taken indirectly, using the average of resistance measurements obtained with a Resistograph® (Instrumenta Mechanik Labor, Germany). Recordings from the Resistograph® were scaled by DBH measurements to obtain wood density indices following El-Kassaby et al. (2011). Wood density in the Interior Spruce population was measured directly in kg/m³ using X-ray densitometry (WD_{X-ray}), which uses increment cores extracted from the trees.

Genotyping of the Douglas-fir samples, using whole exome capture, was performed in a commercial facility (RAPiD Genomics®, Florida, US), probes were designed based on the available Douglas-fir transcriptome assembly (Howe et al., 2013). A total 'pool' of 56,454 SNPs, with <0.40 heterozygosity, was used in this study. Thistlethwaite et al. (2017) describe the genotyping process in more detail. For further explanation on exome capture see, Neves et al. (2013).

The Interior spruce samples were genotyped via multiplexed, high-throughput Genotyping-by-Sequencing (GBS) following Elshire et al. (2011) and Chen et al. (2013), on the Illumina HiSeq 2000 at the Cornell University Genomics Core Laboratory (see, Gamal El-Dien et al. (2015) for a complete description of the genotyping process).

4.2.3 Estimated breeding value (EBV) calculation and ABLUP accuracy

EBVs were calculated in ASReml-R v4.0 (Butler et al., 2017) via linear mixed model analysis. For the Douglas-fir population the following model was used:

$$y = X\beta + Z_1a + Z_2sa + Z_3s(rep) + Z_4sf + Z_5f + e \quad (1)$$

Where y is the phenotypic trait measurement, β is a vector of fixed effects (including mean and site effects), a is a vector of individual random additive genetic effects $\sim N(0, A\sigma_a^2)$, sa is a site x additive genetic interaction $\sim N(0, I\sigma_{sa}^2)$, $s(rep)$ is a vector of the block effect within site $\sim N(0, I\sigma_{s(rep)}^2)$, sf is a random effect site x family interaction $\sim N(0, I\sigma_{sf}^2)$, f is the effect of family $\sim N(0, I\sigma_f^2)$, and e is the random residual effect $\sim N(0, I\sigma_e^2)$, X, Z_{1-5} are incidence matrices assigning fixed and random effects to each observation. I is the identity matrix and A the average numerator relationship matrix.

Narrow-sense heritability was calculated as $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_{sa}^2 + \sigma_{sf}^2 + \sigma_f^2 + \sigma_e^2)$, where σ_a^2 , σ_{sa}^2 , σ_{sf}^2 , σ_f^2 and σ_e^2 are the variances of additive genetic, site x additive genetic, site x family, family, and residual effects, respectively.

For the Interior Spruce population a similar mixed model was used, without family effects:

$$y = X\beta + Z_1a + Z_2sa + Z_3s(rep) + e \quad (2)$$

where y is the phenotypic measurement of the analyzed trait, β is a vector of fixed effect (i.e., the overall mean and the site effect), a is a vector of random additive effects $\sim N(0, A\sigma_a^2)$, sa is a site x additive genetic interaction $\sim N(0, I\sigma_{sa}^2)$, $s(rep)$ is a vector of the block effect within site $\sim N(0, I\sigma_{s(rep)}^2)$, e is a the random residual effect $\sim N(0, I\sigma_e^2)$, and X, Z_{1-3} are incidence matrices assigning fixed and random effects to each observation. I is the identity matrix and A the average numerator relationship matrix.

Narrow-sense heritability for Interior Spruce was calculated as $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_{sa}^2 + \sigma_e^2)$, where σ_a^2 , σ_{sa}^2 , and σ_e^2 are the variances of additive genetic, site x additive genetic, and residual effects, respectively.

ABLUP cross-validation for both species was performed in ASReml-R v4.0 (Butler et al., 2017). 10 fold cross-validation was performed, using randomly sampled individuals from all 3 sites (Adams, Fleet River and Lost Creek for Douglas-fir; Aleza Lake, PGTIS and Quesnel for Interior spruce) to construct the model, and the remainder to compose the validation set. Prediction accuracy for ABLUP was calculated as the correlation between the EBVs from the validation sets and their original EBVs calculated from Eqs.1

and 2, for Douglas-fir and Interior spruce respectively. ABLUP prediction accuracy was compared to GS prediction accuracy for all SNP set totals.

4.2.4 Genomic selection and Cross-validation

The GS method used in the analysis was ridge regression (RR-BLUP) (Whittaker et al., 2000), and was implemented using the R package ‘bigRR’ (Shen et al., 2014). The genomic predicted EBVs (GEBVs) for height, were calculated as the sum of all marker effects within each individual tree. Marker effects were estimated using the following mixed model, from Henderson (1976):

$$y_{EBV} = \mathbf{1}\mu + \mathbf{Z}g + e \quad (3)$$

Where y_{EBV} is the vector of n tree EBV records for height, $\mathbf{1}$ is a vector of 1, μ is the intercept, g is the vector of random marker effects, \mathbf{Z} is the design matrix for the random marker effects, and e is the residual random effects vector. In the RR-BLUP procedure, the residuals and marker effects are presumed to follow normal distributions with constant variance, i.e. $e \sim N(0, I\sigma_e^2)$ and $g \sim N(0, I\sigma_g^2)$, where I is an identity matrix. Marker effect solutions are calculated according to the following equation:

$$\hat{g} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1} \mathbf{Z}'y \quad (4)$$

Where $\lambda = \sigma_e^2/\sigma_g^2$ is the ridge penalization parameter. Marker effects are assumed to be distributed equally, and as such all are uniformly shrunk towards zero.

Predictive accuracy was used to determine performance of this GS method. Predictive accuracy was determined as the mean of the replications of the Pearson product-moment correlation between estimated breeding values (EBVs) of the validation set and their genomic estimated breeding values (GEBVs) or $r(\text{EBV}, \text{GEBV})$. Validation was applied as a replicated randomly assigned 10-fold cross validation repeated 10 times with randomly sampled SNPs. In which 9/10 folds were used to train the model, the other fold constituting the validation population. Information from the 3 sites were pooled.

4.2.5 Effective population size estimation

The effective population size (N_e) for the Douglas-fir and Interior spruce populations were estimated using an Excel program developed by Dr. M. Lstiburek (Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic), that was based on Lindgren et al.'s (1997) status number concept.

4.2.6 Random marker sampling

The effect of the number of markers on predictive accuracy was ascertained by carrying out a random sampling method for choosing markers from the total 'pool' containing 56,454 SNPs for Douglas-fir and 62,190 for Interior spruce. Sets with SNP totals in the range of 200- 50,000 were tested and replicated 10 times, randomly sampling SNPs for each repetition. The cross-validation processes of the RR-BLUP model was then performed using these randomly sampled SNP sets. This analysis was carried out on the height and wood density phenotypes. Assuming a genome length of ~2000cM for both Douglas-fir (Krutovsky et al., 2004) and Interior spruce (Pavy et al., 2008, 2017; Pelgas et al., 2011) (an approximation based on *Picea glauca*, *Pinaceae* data), the average marker densities tested ranged from 0.05-25 markers/cM.

4.3 Results

4.3.1 Marker Number Effect

The average accuracy of height GEBVs ranged from 0.63 (SE 0.040) to 0.87 (SE 0.008) for Douglas-fir, and 0.31 (SE 0.028) to 0.63 (SE 0.019) for Interior spruce, representative of GS using SNP sets of 200 and 50,000 respectively for each species. Similarly the average accuracy of wood density GEBVs ranged from 0.62 (SE 0.023) to 0.83 (SE 0.011) for Douglas-fir, and 0.29 (SE 0.040) to 0.62 (SE 0.017) for Interior spruce, using SNP set totals of 200 and 50,000 respectively. Accordingly, the effect of marker number on the prediction accuracy of multi-site cross-validation data, shows a clear trend of increased predictive accuracy with increase in markers. Although the magnitude of prediction accuracy gains begins to plateau around a threshold of around 10,000-15,000 markers for both Douglas-fir and Interior spruce (Fig. 4.1).

Despite the random selection of markers, there is also little variation in predictive accuracy among SNP sets of the same total, as indicated by the small error bars in Fig. 4.1. On average, the accuracies for height increased by a factor of 1.03 and 1.08 when the number of markers was doubled, for Douglas-fir and Interior spruce respectively. For wood density the accuracies increased on average by a factor of 1.02 and 1.07 when the number of markers was doubled, for Douglas-fir and Interior spruce respectively. Fig. 4.1 is a graphical summary of the results, depicting the average predictive accuracy of 100 replicates per SNP set and 10 random replicates for each SNP set total, and their standard errors.

4.3.2 Pedigree and Relatedness Effect

The full-sib structure of the Douglas-fir models, out-performed those of the half-sib structure of the Interior spruce models for all SNP set totals (Fig. 4.1). The prediction accuracies for the full-sib models were on average, 1.58 and 1.54 times larger than the half-sib models for height and wood density respectively. Prediction accuracy varied more within traits when using half-sib compared to full-sib pedigrees. As represented by SNP set total error bars in Fig. 4.1. On average the standard deviations for half-sib models were 2.59 times larger than those of the full-sib models for height, and 1.75 times larger for wood density.

4.4 Discussion

Generally, prediction accuracies for height and wood density GEBVs of both Douglas-fir and Interior spruce, increased with increasing number of SNPs (Fig. 4.1). The data suggests that increasing the number of markers will provide more accurate predictions. However in the light of genotyping costs and efficiency, it is prudent to use as few markers as possible without loss of significant accuracy. Ideally this would mean that the minimum number of markers should be the same as the number of linkage groups. This number may vary according to the type of breeding population, and genomic data collected. Although the saturation point for this data was around 10,000 markers for both Douglas-fir and Interior spruce using EBVs as the model input, certainly the N_e of the populations need to be taken into account. Data with a higher N_e (>93) may require more markers to obtain similar accuracies.

The GEBV results here coalesce similar results obtained in previous breeding program studies and generalized simulations (Asoro et al., 2011; Beaulieu et al., 2014b; Lenz et al., 2017; Solberg et al., 2008; Sonesson and Meuwissen, 2009; Tan et al., 2017; Wang et al., 2017; Zapata-Valenzuela et al., 2012). The GS prediction accuracies for both height and wood density, parallel those obtained by ABLUP and by GS in previous studies on these species (Gamal El-Dien et al., 2015; Thistlethwaite et al., 2017). These results from models trained on EBVs, led us to conclude that, even at a relatively low density, the markers used were able to capture the genetic relationship as effectively as the pedigree; and in the case of HT at 35 years for Douglas-fir, more effectively. As a result of this finding, future selections may be conducted more efficiently and without the need for a structured pedigree. This will consequently speed up the breeding process by eliminating the need to conduct specific crosses (El-Kassaby and Lstibůrek, 2009). Furthermore, prior studies modeling deregressed EBVs, removing the parental average, in Douglas-fir showed extremely low GS prediction accuracies and largely undetectable LD both within and between generations (Thistlethwaite et al., 2017, 2019).

The role of short-range marker-QTL LD on GS prediction accuracy was imperceptible compared to the effect of relatedness. This was a somewhat anticipated result given the fact that most conifers are relatively undomesticated even within breeding programs, and are known to be highly outbred species with large N_e . These characteristics actively preclude population-wide LD (Isik, 2014; Neale and Savolainen, 2004). Corroborating evidence for can be seen in Fig. 4.1, the variance of the prediction accuracies is modest despite SNPs being randomly selected. This leads us to believe that the SNPs are tracing genetic relatedness as opposed to marker-QTL LD. In addition, other studies have drawn similar conclusions that marker selection strategies have little to no effect on prediction accuracy for growth and wood quality traits (Beaulieu et al., 2014a; Lenz et al., 2017; Zapata-Valenzuela et al., 2012). That is to say, that marker sets made up of only those markers with the largest effects show only a minor advantage over using all markers available (Beaulieu et al., 2014a; Lenz et al., 2017). As reported by Lenz et al. (2017), this observation is likely due to those high-effect markers having higher MAF, and hence higher information value allowing them to trace genetic relationships efficiently. In addition Zapata-Valenzuela et al. (2012) found no

discernable difference in prediction accuracy when using subsets of markers rather than the whole complement of markers available to them. This was regardless of whether or not those markers were in association with the trait in question. Going further, considering at the metadata presented in Fig. 1.2, it can be seen that marker density at the levels currently applied appears to have little to no bearing on GS prediction accuracy. Studies with less dense marker coverage, in some cases over 20 times less dense than the highest, display similarly high prediction accuracies. Since it is unlikely that such few markers could detect that amount of LD in such vast genomes, it is an indication that these markers are tracking something other than LD.

Even in species which do not display the complexity of conifer genomes, similar trends due to marker reduction have been seen. Using two Eucalyptus species and their F₁ hybrids, Tan et al. (2017) found no major advantage in using more than 5,000 SNPs compared to using the full data set of 41,304 SNPs available to them. Lorenz et al. (2012) noted that in a barley (*Hordeum vulgare* L.) population with high LD, the number of markers used could be reduced to 384, with minor impact on prediction accuracy for disease resistance. Marker selection via a k-means clustering algorithm to sample LD space, had only very modest advantages over random sampling. Additionally in dairy breeding, multiple studies have shown that prediction accuracy can be maintained despite significantly reducing marker density (Su et al., 2012; Zhang et al., 2011). In these cases it is the effect of low N_e (Hayes et al., 2009a) which is driving the prediction accuracy. With low N_e , fewer independent genomic segments arise, decreasing the number of markers required to track these segments (Wang et al., 2017). These populations all exhibit closed breeding and subsequently have lower N_e than our sampled populations.

Complex traits are thought to be largely governed by noncoding variants seemingly affecting gene regulation and expression (Boyle et al., 2017; Pickrell, 2014). The number of such variants are thought to be very large, fairly evenly distributed across the genome and have small effect sizes. The heritability of complex traits is thus similarly spread out throughout the entire genome (Boyle et al., 2017). The upshot of this is the implication that a large proportion of all genes contribute to variation in complex traits. This is at odds with the expectation that trait variants are located within specific and biologically relevant genes

and pathways (Boyle et al., 2017). Yet Tan et al. (2017) generally found that SNPs located in intergenic regions provided the better prediction accuracies over those located within/ near genic regions, or using all SNPs available. They attributed this to a slower decline of LD in these regions compared to other locations, causing markers in intergenic regions to better trace QTLs over longer genomic segments than markers with a higher rate of LD decay. Similarly Boyle et al. (2017) in their summary report state that SNPs that contribute most to the heritability, are often spread widely across the genome and are not closely located to genes with trait-specific functions.

Although genome-wide variation in LD is virtually unknown in conifers, data from other plant species suggest that LD is higher but also more variable in intergenic compared to genic regions. There are several reasons for this. First, intergenic regions in many species are replete with repetitive elements, mainly various LTR transposable elements, and such regions are often highly heterochromatic and show reduced cross-over rates. Second, complex structural variation generated by the action of repetitive elements will further limit cross over due to lack of sequence homology in intergenic regions, directing recombination to genomic regions with high gene densities. Both of these features are well documented in maize, a species that also has a relatively high repetitive genome fraction (Dooner and He, 2008; Fu et al., 2002). As noted above, levels of linkage disequilibrium in conifers are poorly studied to date. Small, targeted re-sequencing of exomic regions have found that LD decays relatively rapidly, over a few kb at the most (Heuertz et al., 2006; Larsson et al., 2013; Pyhajarvi et al., 2007). This is to be expected if most recombination is largely directed towards genic regions. However, since genic regions only constitute at most a percent or so of a typical conifer genome, these rates are likely not representative for most of the genome. As an example, Fu et al. (2002) concluded that the repetitive DNA in maize, while constituting the bulk of the genome, likely contributes little if anything to genetic length.

And so, it is perhaps not unusual that despite relatively large marker totals in use here (and an even greater total used in Thistlethwaite et al. (2017)) and for Douglas-fir those markers being located in the exomic regions, LD was not successfully traced.

4.5 Summary

The presupposition of genomic selection (GS) is that predictive accuracies should be based on population-wide LD. However in species with large, highly complex genomes the limitation of marker density may preclude the ability to resolve LD. Here we investigate such an effect on two conifer species with ~ 20 Gbp genomes, Douglas-fir (*Pseudotsuga menziesii* Mirb. (Franco)) and Interior spruce (*Picea glauca* (Moench) Voss x *Picea engelmannii* Parry ex Engelm.). Random sampling of markers was performed to obtain SNP sets with totals in the range of 200- 50,000, this was replicated 10 times. RR-BLUP was deployed as the GS method to test these SNP sets, and 10-fold cross-validation was performed on each for: 1,321 Douglas-fir trees, representing 37 full-sib F₁ families from 3 sites in British Columbia, Canada; and 1,126 Interior spruce trees, representing 25 open pollinated families located on 3 sites also in British Columbia, Canada.

As marker number increased, so too did GS predictive accuracy for both conifer species. However a plateau of gains began around 10,000 – 15,000 markers for both Douglas-fir and Interior spruce. Despite random marker selection, little variation in predictive accuracy was observed. Douglas-fir prediction accuracies were higher on average than those of Interior spruce, reflecting the full-sib nature of the Douglas-fir population versus the half-sib Interior spruce population.

Although possibly advantageous within an advanced breeding population, reducing marker density cannot be recommended for carrying out GS in conifers. Significant LD was not detected using 50,000 SNPS, only the relatedness of the populations studied. Dramatically increasing marker density would enable said markers to track LD in these large, genetically diverse genomes. As well as providing a model that could be used across populations, breeding programs, and traits.

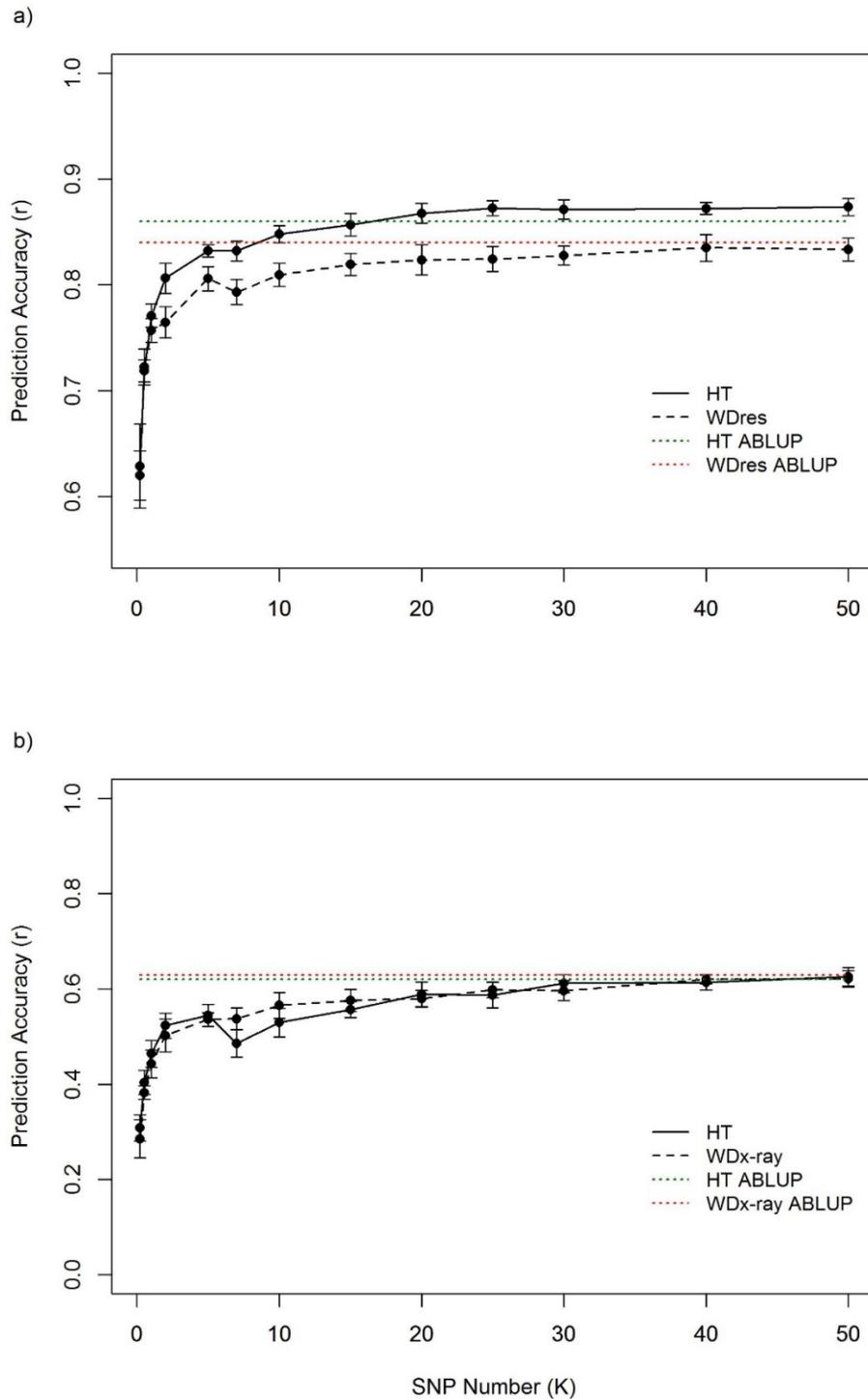


Figure 4.1 Effect of marker density on RR-BLUP prediction accuracy for multi-site data for height and wood resistance, with ABLUP for comparison, for: a) Douglas-fir (HT at 35 years, WD_{res} at 38 years), and b) Interior spruce (HT at 38 years, WD_{x-ray} at 38 years).

5 Conclusion

5.1 Research novelties and applications

The incorporation of genomic selection (GS) into commercial forestry would mark a significant juncture for forest tree breeders. Having already garnered much success in animal and crop breeding industries, it has the potential to substantially reduce lengthy breeding cycles whilst increasing gain per unit time. The push for more affordable genotyping technologies has precipitated the introduction of GS into forestry research, and will continue to make GS a viable option for breeding programs. Some important attributes concerning the application of GS to forest tree breeding have been explored in this dissertation, and will afford insight into future utilization.

The use of exome capture as a genotyping platform in all three studies (Chapters 2-4), allowed the discernment of its potential use for GS as applied to species with large genomes, in this case conifers. As a cost-effective method, it has been proposed that exome capture would provide equivalent GS results, without the price tag of genotyping array. Indeed exome capture proved very efficient in the capture of pedigree relationships, however marker-QTL LD was not as easily resolved. Although the latter conclusion was the result of both the genotyping platform used, and the genetic architecture of conifers and their influence on the GS method (Chapter 4). Chapter 2 represents the first application of exome capture data to GS in forest trees, which was validated over space and time.

If incorporated into forest tree breeding programs, GS will inevitably be applied to assess and select trees from progeny generations. Yet most studies, aside from work performed by Isik et al. (2016) and Bartholomé et al. (2016), have used within-generation cross-validation in their studies. Chapter 3 explores the effect of inter-generational validation, the role of the relatedness between training and validation sets, and among-family effects on GS prediction accuracy.

Chapter 4 investigates the effect of marker density on GS prediction accuracy in two conifer species, with differing family structures and genotyping platforms. It highlights the view that GS applied to conifers (and other species with large, outbred genomes) results in prediction accuracies that represent

the method's ability to resolve genetic relationships rather than LD. Deregression analysis of Douglas-fir from Chapters 2 and 3, support this idea. Important conclusions concerning the number of markers required to capture LD are discussed, as is the use of exome capture for GS.

5.2 Conclusions regarding research objectives

- To compare ABLUP and two GS methods; ridge regression best linear unbiased predictor (RR-BLUP) and generalised ridge regression (GRR) (Chapter 2).

The RR-BLUP and GRR models produced similar predictive accuracies across the studied traits and were comparable to ABLUP (Chapter 2). It was found that using EBVs for GS model training caused the retention of family means. The resulting predictive accuracies for both GS models were high due to the capture of both contemporary and historical pedigree information, yet little was garnered about the genetic architecture of each trait and therefore the suitability of the GS methods. Subsequently deregressed EBVs were used for GS model training, removing the parent average effect in order to disassociate the pedigree information from the marker-QTL LD. This resulted in prediction accuracies which for all practical purposes were 0.0 for both GS methods. When the influence of pedigree was removed, the generated models were no longer able to provide useful predictions, resolve LD or provide information on the genetic architecture of the traits considered.

- Assay GS prediction accuracy across environments/ spatial divisions and time (Chapter 2).

The predictive accuracy of GS models varied across the spatial divisions described in Chapter 2. On average the combined/ multi-site GS analyses produced slightly higher prediction accuracies than the single site analyses (Fig. 2.2a and b). The effect of increasing the training population size by combining sites had a positive effect on prediction accuracy, in agreement with Grattapaglia's (2014) recommendation. In addition, the multi-site approach incorporated the present GxE in the model, resulting in further improvement. Similar prediction accuracies to the within-site analyses were observed in the cross-site GS analyses (with models trained on EBVs) (Fig. 2.2a and b). An indication that within the partially controlled

experimental environment in which the trees were growing, the GxE effects were minimal. Height at 12 years was deemed an acceptable age at which moderately accurate predictions can be made concerning future height (age-age) and wood density (trait-trait). However, GS models should be periodically updated with phenotypic data throughout tree lifetime for more accurate predictions of mature traits.

- Review the potential of exome capture for use as a genotyping platform in GS studies concerning conifer species (Chapters 2, 3, and 4).

Exome capture was moderately successful as a genotyping platform with regards to the application of genomic selection in Douglas-fir, with a major caveat. Although exome capture genomic data produced some significant predictive accuracies over the various validation procedures (Chapters 2, 3, and 4), these were largely a result of relatedness rather than the LD. The results suggest that the population of SNP markers used, along with their low coverage across the Douglas-fir genome was not successful for capturing the LD with the causal genes underpinning the studied attributes. Denser or alternative marker generation methods such as whole genome sequencing or other dense unordered SNP genotyping methods such as GBS are needed, as is a larger SNP array. In this instance, exome capture provided enough markers to successfully capture/track the pedigree (contemporary and historical) and thus it was useful for genetic variance decomposition of conifer traits, and providing better genetic parameter estimates.

Boyle et al. (2017) assert that SNPs contributing most to the heritability of complex traits, are often spread widely across the genome and are not necessarily closely located to genes with trait-specific functions. Given this, in addition to the findings presented here regarding the lack of marker-QTL LD resolution, it would be remiss not to err on the side of caution when recommending exome capture for GS studies. It may be that denser marker information might capture the elusive LD, or it might be the case that LD should be sought out in additional genomic regions.

- Determine intergenerational prediction accuracy using progeny as a validation population (Chapter 3).

The cross-generational GS validation of juvenile height in Douglas-fir (Chapter 3), provided results that almost matched the ABLUP predictive accuracy. However, it was discovered that the predictive accuracy was driven by the relatedness between the training and validation sets, and even more so in capturing among-family effects. Whilst this relationship is exploited for selection in current breeding programs, the ultimate aim of using GS should be to capture true LD across populations and traits to uncover presently unknown variation, and possibly unknown traits (Beaulieu et al., 2014a; Grattapaglia, 2017).

- Investigate the effect of kinship on GS prediction accuracy (Chapters 2 and 3).

Deregression of the BVs reduced the accuracies of all cross-validation procedures dramatically (Chapters 2 and 3), indicating that the models were tracking pedigree rather than marker-QTL LD. Deregression may subsequently be preferential only in advanced breeding programs (Grattapaglia, 2017). Given this insight, in order to capture short-range marker-QTL LD in outbred conifer species, it is recommended that deregression be used to remove parental average effects as well as high density marker sets.

- Study the effect of marker density on GS prediction accuracy in two conifer species (Chapter 4).

Chapter 4 demonstrates that increasing marker density has the effect of increasing GS prediction accuracy of multi-site cross-validation data. The magnitude of gain begins to plateau around 10,000-15,000 markers for both Douglas-fir and Interior spruce (Fig. 4.1). Despite using a method of random marker selection, there was little variation in predictive accuracy, indicated by the small error bars in Fig. 4.1.

While the number of SNPs here (Chapters 2-4), derived from sequence capture, represents one of the highest in forest trees GS studies (Grattapaglia, 2017), we have yet to observe enough marker-QTL LD in our GS methods. Many more markers may be required for this to be resolved since LD in Douglas-fir survives only over short distances as in other conifer species (Neale and Savolainen, 2004).

Although advantageous within a population, reducing marker density may not be the most effective or economical method of carrying out GS, especially in conifers. As mentioned previously, we have yet to trace LD with the current array of markers available, only genetic relationships. Given the genetic diversity

of conifer species, it would perhaps be more prudent to create a denser marker array that can be used across populations and breeding programs (Grattapaglia, 2017). Increasing the number of markers in such a way could enable us to tease apart the genetic relationship from the LD, and to investigate multiple traits including unanticipated traits (Beaulieu et al., 2014a). If we can do this, the higher density of markers would offset somewhat, the effects of marker-QTL LD decay due to selection and recombination over multiple generations (Beaulieu et al., 2014a; Grattapaglia, 2017; Solberg et al., 2008). Low density marker arrays would have more impact in more advanced breeding programs (Grattapaglia, 2017).

5.3 Future research directions

The term “paradigm shift” has been affiliated with the use of GS in forest tree breeding for some years now. Indeed it is an appropriate description for, as stated in the introduction of this work, the model unit of breeding analyses will shift from being the line of descent to the allele. It is of the utmost importance therefore, that this concept is actually realized as it comprises the very core of what we are striving to achieve.

The abundance of sequencing technologies and their continued advancement, has already made possible the proliferation of research into GS in forest trees that has been seen to date. As these technologies improve, inevitably so will the quality of genomic data, a necessity for conifer species (and other species with large, complex genomes). High quality, dense genotyping will likely be the key to capturing LD in these species with large, outbred populations and low population-wide LD (Isik, 2014; Neale and Savolainen, 2004).

In conjunction with high density genotyping, technological advancements in DNA sequencing will also prompt the construction of genetic maps and with it, functional analyses for highly complex genomes. Having references and association data like this will further improve GS in forest trees, by allowing the exploitation of *a priori* knowledge, regarding genetic architecture, in GS models (Morgante et al., 2018).

As inevitable as GS in forest tree breeding appears, so too will be the requirement of continued model revision and appraisal. Since, patterns of LD can change between generations, and those generations

more distantly separated show a decrease in GS prediction accuracy (Wu et al., 2015). GS models will have to be updated periodically using phenotypic information from a variety of generations. As Bartholomé et al. (2016) discuss, there is some benefit in combining phenotypes across generations which helps to infer the Mendelian sampling term. In addition, simulations have shown that higher marker densities allowed prediction accuracies to be maintained over longer generational gaps (Müller et al., 2017b).

Whilst in traditional breeding programs familial relationships are exploited for selection, it is the objective of GS to capture true LD across populations and traits to expose unidentified sources of variation (Beaulieu et al., 2014a; Grattapaglia, 2017). However, given the fact that LD breaks due to recombination after every breeding cycle, that forest tree species have lower observed levels of LD (Neale and Kremer, 2011), and given the findings presented here; higher density marker arrays and/ or improved distribution of those markers are a prerequisite for capturing LD in Douglas-fir and other such conifers (Jaramillo-Correa et al., 2015).

5.4 Strengths and limitations

At the time of publication, the number of SNPs used in Chapter 2, derived from sequence capture, represented the highest in forest trees GS studies (Grattapaglia, 2017). Chapter 2 also represents the first application of exome capture data to GS in forest trees. Although Chapters 2-4 reported some high GS prediction accuracies, these were attributed to genetic relationships rather than a result of capturing LD. Given these findings, the studies presented here provide evidence that GS in forest tree species with large, complex genomes, requires dense marker sets with many more SNPs than have been used to date. We may also conclude that exome capture presents a cost-effective tool for tracking pedigree relationships, and genetic variance decomposition.

Alternative marker generation methods such as whole genome sequencing or other dense unordered SNP genotyping methods such as GBS may be more suitable for future GS studies in conifers. Exome capture may not provide the coverage needed in order to track LD, however further investigation with larger SNP arrays is required to test this.

Alternatively, using a single, large full-sib family causes LD to extend over further distances compared to open-pollinated trees, therefore it is possible to make within family selections using this type of population as (Fuentes-Utrilla et al., 2017) have demonstrated. Therefore low density marker arrays would have more impact in more advanced breeding programs (Grattapaglia, 2017).

References

- Acheré, V., Faivre-Rampant, P., Jeandroz, S., Besnard, G., Markussen, T., Aragonés, A., Fladung, M., Ritter, E., and Favre, J.-M. (2004). A full saturated linkage map of *Picea abies* including AFLP, SSR, ESTP, 5S rDNA and morphological markers. *Theoretical and Applied Genetics* *108*, 1602–1613.
- Asoro, F.G., Newell, M.A., Beavis, W.D., Scott, M.P., and Jannink, J.-L. (2011). Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. *The Plant Genome Journal* *4*, 132.
- Atefi, A., Shadparvar, A.A., and Ghavi Hossein-Zadeh, N. (2016). Comparison of whole genome prediction accuracy across generations using parametric and semi parametric methods. *Acta Scientiarum. Animal Sciences* *38*, 447.
- Avendanó, S., Woolliams, J.A., and Villanueva, B. (2004). Mendelian sampling terms as a selective advantage in optimum breeding schemes with restrictions on the rate of inbreeding. *Genetical Research* *83*, 55–64.
- Bartholomé, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., and Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. *BMC Genomics* *17*, 604.
- Beaulieu, J., Doerksen, T., Boyle, B., Clément, S., Deslauriers, M., Beauseigle, S., Blais, S., Poulin, P.-L., Lenz, P., Caron, S., et al. (2011). Association Genetics of Wood Physical Traits in the Conifer White Spruce and Relationships With Gene Expression. *Genetics* *188*, 197–214.
- Beaulieu, J., Doerksen, T., Clément, S., MacKay, J., and Bousquet, J. (2014a). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* *113*, 343–352.
- Beaulieu, J., Doerksen, T.K., MacKay, J., Rainville, A., and Bousquet, J. (2014b). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics* *15*, 1048.
- Berger, S., Pérez-Rodríguez, P., Veturi, Y., Simianer, H., and de los Campos, G. (2015). Effectiveness of Shrinkage and Variable Selection Methods for the Prediction of Complex Human Traits using Data from Distantly Related Individuals: Predictions with Distantly Related. *Annals of Human Genetics* *79*, 122–135.
- Bodi, K., Perera, A.G., Adams, P.S., Bintzler, D., Dewar, K., Grove, D.S., Kieleczawa, J., Lyons, R.H., Neubert, T.A., Noll, A.C., et al. (2013). Comparison of Commercially Available Target Enrichment Methods for Next-Generation Sequencing. *Journal of Biomolecular Techniques : JBT* *24*, 73–86.
- Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* *169*, 1177–1186.
- Burdon, R.D. (1977). Genetic Correlation as a Concept for Studying Genotype-Environment Interaction in Forest Tree Breeding. *Silvae Genetica* *26*, 168.
- Butler, D.G., Cullis, B.R., Gilmour, A.R., Gogel, B.G., and Thompson, R. (2017). ASReml-R Reference Manual Version 4.

- de los Campos, G., Vazquez, A.I., Fernando, R., Klimentidis, Y.C., and Sorensen, D. (2013). Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genetics* 9, e1003608.
- Cappa, E.P., El-Kassaby, Y.A., Garcia, M.N., Acuña, C., Borralho, N.M.G., Grattapaglia, D., and Marcucci Poltri, S.N. (2013). Impacts of Population Structure and Analytical Models in Genome-Wide Association Studies of Complex Traits in Forest Trees: A Case Study in *Eucalyptus globulus*. *PLoS ONE* 8, e81267.
- Cappa, E.P., Stoehr, M.U., Xie, C.-Y., and Yanchuk, A.D. (2016). Identification and joint modeling of competition effects and environmental heterogeneity in three Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*) trials. *Tree Genetics & Genomes* 12, 102.
- Chen, C., Mitchell, S.E., Elshire, R.J., Buckler, E.S., and El-Kassaby, Y.A. (2013). Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genetics & Genomes* 9, 1537–1544.
- Chen, Z.-Q., Baisou, J., Pan, J., Karlsson, B., Andersson, B., Westin, J., García-Gil, M.R., and Wu, H.X. (2018). Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BMC Genomics* 19.
- Choi, M., Scholl, U.I., Ji, W., Liu, T., Tikhonova, I.R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Özen, S., Sanjad, S., et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences* 106, 19096–19101.
- De La Torre, A.R., Birol, I., Bousquet, J., Ingvarsson, P.K., Jansson, S., Jones, S.J.M., Keeling, C.I., MacKay, J., Nilsson, O., Ritland, K., et al. (2014). Insights into Conifer Giga-Genomes. *PLANT PHYSIOLOGY* 166, 1724–1732.
- Dean, C.A., and Stonecypher, R.W. (2006). Early Selection of Douglas-Fir across South Central Coastal Oregon, USA. *Silvae Genetica* 55, 135–141.
- Dooner, H.K., and He, L. (2008). Maize Genome Structure Variation: Interplay between Retrotransposon Polymorphisms and Genic Recombination. *THE PLANT CELL ONLINE* 20, 249–258.
- Dutkowski, G.W., Silva, J.C. e, Gilmour, A.R., and Lopez, G.A. (2002). Spatial analysis methods for forest genetic trials. *Canadian Journal of Forest Research* 32, 2201–2214.
- Eckert, A.J., van Heerwaarden, J., Wegrzyn, J.L., Nelson, C.D., Ross-Ibarra, J., Gonzalez-Martinez, S.C., and Neale, D.B. (2010). Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics* 185, 969–982.
- Edwards, M.D., and Page, N.J. (1994). Evaluation of marker-assisted selection through computer simulation. *Theoretical and Applied Genetics* 88, 376–382.
- El-Dien, O.G., Ratcliffé, B., Klápště, J., Porth, I., Chen, C., and El-Kassaby, Y.A. (2018). Multienvironment genomic variance decomposition analysis of open-pollinated Interior spruce (*Picea glauca* x *engelmannii*). *Molecular Breeding* 38.
- El-Kassaby, Y.A. (1982). Associations between Allozyme Genotypes and Quantitative Traits in Douglas-Fir [*PSEUDOTSUGA MENZIESII* (Mirb.) Franco]. *Genetics* 101, 103–115.
- El-Kassaby, Y.A., and Lstibůrek, M. (2009). Breeding without breeding. *Genetics Research* 91, 111.

- El-Kassaby, Y.A., and Park, Y.S. (1993). Genetic Variation and Correlation in Growth, Biomass, and Phenology of Douglas-Fir Diallel Progeny at Different Spacings. *Silvae Genetica* 42, 289.
- El-Kassaby, Y.A., Mansfield, S., Isik, F., and Stoehr, M. (2011). In situ wood quality assessment in Douglas-fir. *Tree Genetics & Genomes* 7, 553–561.
- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6, e19379.
- Endelman, J.B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal* 4, 250.
- Falconer, D.S., and Mackay, T.F.C. (2009). Introduction to quantitative genetics (Harlow: Pearson, Prentice Hall).
- Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433.
- Fisher, R.A. (1919). XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* 52, 399–433.
- Fu, H., Zheng, Z., and Dooner, H.K. (2002). Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Proceedings of the National Academy of Sciences* 99, 1082–1087.
- Fu, Y., Springer, N.M., Gerhardt, D.J., Ying, K., Yeh, C.-T., Wu, W., Swanson-Wagner, R., D’Ascenzo, M., Millard, T., Freeberg, L., et al. (2010). Repeat subtraction-mediated sequence capture from a complex genome: Sequence capture in maize. *The Plant Journal* 62, 898–909.
- Fuentes-Utrilla, P., Goswami, C., Cottrell, J.E., Pong-Wong, R., Law, A., A’Hara, S.W., Lee, S.J., and Woolliams, J.A. (2017). QTL analysis and genomic selection using RADseq derived markers in Sitka spruce: the potential utility of within family data. *Tree Genetics & Genomes* 13, 33.
- Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Chen, C., Porth, I., and El-Kassaby, Y.A. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genomics* 16, 370.
- Garrick, D.J., Taylor, J.F., and Fernando, R.L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution* 41, 55.
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv:1207.3907 [q-Bio]*.
- Gezan, S.A., Osorio, L.F., Verma, S., and Whitaker, V.M. (2017). An experimental validation of genomic selection in octoploid strawberry. *Horticulture Research* 4, 16070.
- Gianola, D., and van Kaam, J.B.C.H.M. (2008). Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. *Genetics* 178, 2289–2303.

- Gianola, D., Fernando, R.L., and Stella, A. (2006). Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. *Genetics* 173, 1761–1776.
- Gianola, D., de los Campos, G., Hill, W.G., Manfredi, E., and Fernando, R. (2009). Additive Genetic Variability and the Bayesian Alphabet. *Genetics* 183, 347–363.
- Gilmour, A.R., Gogel, B.G., Cullis, B.R., and Thompson, R. (2009). ASReml User Guide Release 3.0.
- Gnrirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27, 182–189.
- Goddard, M.E., Hayes, B.J., and Meuwissen, T.H.E. (2010). Genomic selection in livestock populations. *Genetics Research* 92, 413–421.
- Grattapaglia, D. (2014). Breeding Forest Trees by Genomic Selection: Current Progress and the Way Forward. In *Genomics of Plant Genetic Resources*, R. Tuberosa, A. Graner, and E. Frison, eds. (Dordrecht: Springer Netherlands), pp. 651–682.
- Grattapaglia, D. (2017). Status and Perspectives of Genomic Selection in Forest Tree Breeding. In *Genomic Selection for Crop Improvement*, R.K. Varshney, M. Roorkiwal, and M.E. Sorrells, eds. (Cham: Springer International Publishing), pp. 199–249.
- Grattapaglia, D., and Resende, M.D.V. (2011). Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7, 241–255.
- Habier, D., Fernando, R.L., and Dekkers, J.C.M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177, 2389–2397.
- Habier, D., Fernando, R.L., and Garrick, D.J. (2013). Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. *Genetics* 194, 597–607.
- Hallander, J. (2009). Novel methods for improved tree breeding (Umeå: Swedish University of Agricultural Sciences).
- Hallander, J., and Waldmann, P. (2009). Optimum contribution selection in large general tree breeding populations with an application to Scots pine. *Theoretical and Applied Genetics* 118, 1133–1142.
- Hayes, B.J., Visscher, P.M., and Goddard, M.E. (2009a). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research* 91, 47.
- Hayes, B.J., Bowman, P.J., Chamberlain, A.J., and Goddard, M.E. (2009b). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92, 433–443.
- Heffner, E.L., Sorrells, M.E., and Jannink, J.-L. (2009). Genomic Selection for Crop Improvement. *Crop Science* 49, 1.
- Heffner, E.L., Lorenz, A.J., Jannink, J.-L., and Sorrells, M.E. (2010a). Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* 50, 1681.
- Heffner, E.L., Lorenz, A.J., Jannink, J.-L., and Sorrells, M.E. (2010b). Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* 50, 1681.

- Henderson, C.R. (1953). Estimation of Variance and Covariance Components. *Biometrics* 9, 226.
- Henderson, C.R. (1976). A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics* 32, 69.
- Heuertz, M., De Paoli, E., Kallman, T., Larsson, H., Jurman, I., Morgante, M., Lascoux, M., and Gyllenstrand, N. (2006). Multilocus Patterns of Nucleotide Diversity, Linkage Disequilibrium and Demographic History of Norway Spruce [*Picea abies* (L.) Karst]. *Genetics* 174, 2095–2105.
- Hill, W.G., Goddard, M.E., and Visscher, P.M. (2008). Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genetics* 4, e1000008.
- Howe, G.T., Yu, J., Knaus, B., Cronn, R., Kolpak, S., Dolan, P., Lorenz, W.W., and Dean, J.F. (2013). A SNP resource for Douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC Genomics* 14, 137.
- Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests* 45, 379–401.
- Isik, F., Bartholomé, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C., and Bouffier, L. (2016). Genomic selection in maritime pine. *Plant Science* 242, 108–119.
- Isik, F., Holland, J., and Maltecca, C. (2017). *Genetic data analysis for plant and animal breeding* (Cham: Springer).
- Ivanova, N.V., Fazekas, A.J., and Hebert, P.D.N. (2008). Semi-automated, Membrane-Based Protocol for DNA Isolation from Plants. *Plant Molecular Biology Reporter* 26, 186–198.
- Jannink, J.-L., Lorenz, A.J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* 9, 166–177.
- Jaramillo-Correa, J.P., Prunier, J., Vázquez-Lobo, A., Keller, S.R., and Moreno-Letelier, A. (2015). Molecular Signatures of Adaptation and Selection in Forest Trees. In *Advances in Botanical Research*, (Elsevier), pp. 265–306.
- Kang, B.-Y., Mann, I.K., Major, J.E., and Rajora, O.P. (2010). Near-saturated and complete genetic linkage map of black spruce (*Picea mariana*). *BMC Genomics* 11, 515.
- Kiezun, A., Garimella, K., Do, R., Stitzel, N.O., Neale, B.M., McLaren, P.J., Gupta, N., Sklar, P., Sullivan, P.F., Moran, J.L., et al. (2012). Exome sequencing and the genetic basis of complex traits. *Nature Genetics* 44, 623–630.
- Krakowski, J., Park, Y.S., and El-Kassaby, Y.A. (2006). Early testing of Douglas-fir: wood density and ring width. *Forest Genetics* 12, 99–105.
- Krutovsky, K.V., Troggio, M., Brown, G.R., Jermstad, K.D., and Neale, D.B. (2004). Comparative Mapping in the Pinaceae. *Genetics* 168, 447.
- Kumar, S., Chagné, D., Bink, M.C.A.M., Volz, R.K., Whitworth, C., and Carlisle, C. (2012). Genomic Selection for Fruit Quality Traits in Apple (*Malus domestica* Borkh.). *PLoS ONE* 7, e36674.

- Lande, R., and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* *124*, 743.
- Larsson, H., Källman, T., Gyllenstrand, N., and Lascoux, M. (2013). Distribution of Long-Range Linkage Disequilibrium and Tajima's D Values in Scandinavian Populations of Norway Spruce (*Picea abies*). *Genes|Genomes|Genetics* *3*, 795–806.
- Lee, W.-P., Stromberg, M.P., Ward, A., Stewart, C., Garrison, E.P., and Marth, G.T. (2014). MOSAIK: A Hash-Based Algorithm for Accurate Next-Generation Sequencing Short-Read Mapping. *PLoS ONE* *9*, e90581.
- Lenz, P.R.N., Beaulieu, J., Mansfield, S.D., Clément, S., Despons, M., and Bousquet, J. (2017). 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* *18*.
- Lerceteau, E., Szmidt, A.E., and Andersson, B. (2001). Detection of quantitative trait loci in *Pinus sylvestris* L. across years. *Euphytica* *121*, 117–122.
- Lin, Z., Hayes, B.J., and Daetwyler, H.D. (2014). Genomic selection in crops, trees and forages: a review. *Crop and Pasture Science* *65*, 1177.
- Lindgren, D., Gea, L.D., and Jefferson, P.A. (1997). Status number for measuring genetic diversity. *International Journal of Forest Genetics* *4*, 762–764.
- Lorenz, A.J., Chao, S., Asoro, F.G., Heffner, E.L., Hayashi, T., Iwata, H., Smith, K.P., Sorrells, M.E., and Jannink, J.-L. (2011). Genomic Selection in Plant Breeding. In *Advances in Agronomy*, (Elsevier), pp. 77–123.
- Lorenz, A.J., Smith, K.P., and Jannink, J.-L. (2012). Potential and Optimization of Genomic Selection for Fusarium Head Blight Resistance in Six-Row Barley. *Crop Science* *52*, 1609.
- Ma, Y., Reif, J.C., Jiang, Y., Wen, Z., Wang, D., Liu, Z., Guo, Y., Wei, S., Wang, S., Yang, C., et al. (2016). Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). *Molecular Breeding* *36*.
- Magnussen, S., and Yanchuk, A.D. (1993). Selection age and risk: finding the compromise. *Silvae Genetica* *42*, 25–40.
- Märtens, K., Hallin, J., Warringer, J., Liti, G., and Parts, L. (2016). Predicting quantitative traits from genome and phenome with near perfect accuracy. *Nature Communications* *7*, 11512.
- Matheson, A.C., and Cotterill, P.P. (1990). Utility of genotype × environment interactions. *Forest Ecology and Management* *30*, 159–174.
- Matheson, A.C., and Raymond, C.A. (1984). The impact of genotype x environment interaction on Australian *Pinus Radiata* breeding programs. *Aust. For Res.* *14*, 11–25.
- McCulloch, C.E., and Searle, S.R. (2001). *Generalized, linear, and mixed models* (New York: John Wiley & Sons).

- Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J.M.L.M., van der Zaag, P.J., Franke, A., Nilsson, M., Lehrach, H., and Brookes, A.J. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics* 10, 374–386.
- Meuwissen, T.H. (2009). Accuracy of breeding values of “unrelated” individuals predicted by dense SNP genotyping. *Genetics Selection Evolution* 41, 35.
- Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Moreau, L., Charcosset, A., and Gallais, A. (2004). Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica* 137, 111–118.
- Morgante, F., Huang, W., Maltecca, C., and Mackay, T.F.C. (2018). Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. *Heredity* 120, 500–514.
- Mrode, R.A. (2014). *Linear models for the prediction of animal breeding values* (Boston, MA: CABI).
- Müller, B.S.F., Neves, L.G., de Almeida Filho, J.E., Resende, M.F.R., Muñoz, P.R., dos Santos, P.E.T., Filho, E.P., Kirst, M., and Grattapaglia, D. (2017a). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of Eucalyptus. *BMC Genomics* 18, 524.
- Müller, D., Schopp, P., and Melchinger, A.E. (2017b). Persistency of Prediction Accuracy and Genetic Gain in Synthetic Populations Under Recurrent Genomic Selection. *Genes|Genomes|Genetics* 7, 801–811.
- Munoz, P.R., Resende, M.F.R., Huber, D.A., Quesada, T., Resende, M.D.V., Neale, D.B., Wegrzyn, J.L., Kirst, M., and Peter, G.F. (2014). Genomic Relationship Matrix for Correcting Pedigree Errors in Breeding Populations: Impact on Genetic Parameters and Genomic Selection Accuracy. *Crop Science* 54, 1115.
- Neale, D.B., and Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12, 111–122.
- Neale, D.B., and Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science* 9, 325–330.
- Neale, D.B., McGuire, P.E., Wheeler, N.C., Stevens, K.A., Crepeau, M.W., Cardeno, C., Zimin, A.V., Puiu, D., Perte, G.M., Sezen, U.U., et al. (2017). The Douglas-Fir Genome Sequence Reveals Specialization of the Photosynthetic Apparatus in Pinaceae. *Genes|Genomes|Genetics* 7, 3157–3167.
- Neves, L.G., Davis, J.M., Barbazuk, W.B., and Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal* 75, 146–156.
- Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276.
- Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584.

- Owino, F. (1977). Genotype x environment interaction and genotypic stability in loblolly pine. *Silvae Genetica* 26, 21–26.
- Patterson, H.D., and Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika* 58, 545.
- Patterson, H.D., and Thompson, R. (1974). Maximum likelihood estimation of components of variance. In *Proc. 8th Internat. Biometric Conf.*, (Washington, DC: Biometric Society), pp. 197–207.
- Pavy, N., Pelgas, B., Beauseigle, S., Blais, S., Gagnon, F., Gosselin, I., Lamothe, M., Isabel, N., and Bousquet, J. (2008). Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics* 9, 21.
- Pavy, N., Lamothe, M., Pelgas, B., Gagnon, F., Birol, I., Bohlmann, J., Mackay, J., Isabel, N., and Bousquet, J. (2017). A high-resolution reference genetic map positioning 8.8 K genes for the conifer white spruce: structural genomics implications and correspondence with physical distance. *The Plant Journal* 90, 189–203.
- Pelgas, B., Bousquet, J., Meirmans, P.G., Ritland, K., and Isabel, N. (2011). QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics* 12.
- Perez, P., and de los Campos, G. (2014). Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* 198, 483–495.
- Pickrell, J.K. (2014). Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *The American Journal of Human Genetics* 94, 559–573.
- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M., et al. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *The Plant Genome Journal* 5, 103.
- Pyhajarvi, T., Garcia-Gil, M.R., Knurr, T., Mikkonen, M., Wachowiak, W., and Savolainen, O. (2007). Demographic History Has Influenced Nucleotide Diversity in European *Pinus sylvestris* Populations. *Genetics* 177, 1713–1724.
- Ratcliffe, B., El-Dien, O.G., Klápště, J., Porth, I., Chen, C., Jaquish, B., and El-Kassaby, Y.A. (2015). A comparison of genomic selection models across time in interior spruce (*Picea engelmannii* × *glauca*) using unordered SNP imputation methods. *Heredity* 115, 547–555.
- Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-Validation. In *Encyclopedia of Database Systems*, L. Liu, and M.T. Özsu, eds. (Boston, MA: Springer US), pp. 532–538.
- Resende, M.D.V., Resende, M.F.R., Sansaloni, C.P., Petrolí, C.D., Missiaggia, A.A., Aguiar, A.M., Abad, J.M., Takahashi, E.K., Rosado, A.M., Faria, D.A., et al. (2012a). Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194, 116–128.

- Resende, M.F.R., Muñoz, P., Acosta, J.J., Peter, G.F., Davis, J.M., Grattapaglia, D., Resende, M.D.V., and Kirst, M. (2012b). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* *193*, 617–624.
- Resende, M.F.R., Munoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F., and Kirst, M. (2012c). Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics* *190*, 1503–1510.
- Resende, R.T., Resende, M.D.V., Silva, F.F., Azevedo, C.F., Takahashi, E.K., Silva-Junior, O.B., and Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. *Heredity* *119*, 245–255.
- Ritter, E., Aragonés, A., Markussen, T., Achere, V., Espinel, S., Fladung, M., Wrobel, S., Faivre-Rampant, P., Jeandroz, S., and Favre, J.-M. (2002). Towards construction of an ultra high density linkage map for *Pinus pinaster*. *Annals of Forest Science* *59*, 637–643.
- Rutkoski, J.E., Poland, J., Jannink, J.-L., and Sorrells, M.E. (2013). Imputation of Unordered Markers and the Impact on Genomic Selection Accuracy. *Genes|Genomes|Genetics* *3*, 427–439.
- Sallam, A.H., Endelman, J.B., Jannink, J.-L., and Smith, K.P. (2015). Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. *The Plant Genome* *8*.
- Sharma, A.K. (2000). Forest tree improvement in India prospects. In *Glimpses in Botany*, K.G. Mukerji, B.P. Chamola, and A.K. Sharma, eds. (New Delhi: APH Pub. Corp.), pp. 181–190.
- Shen, X., Alam, M., Fikse, F., and Ronnegard, L. (2013). A Novel Generalized Ridge Regression Method for Quantitative Genetics. *Genetics* *193*, 1255–1268.
- Shen, X., Alam, M., and Ronnegard, L. (2014). Package “bigRR”: Generalized Ridge Regression (with special advantage for $p \gg n$ cases).
- Solberg, T.R., Sonesson, A.K., Woolliams, J.A., and Meuwissen, T.H.E. (2008). Genomic selection using different marker types and densities. *Journal of Animal Science* *86*, 2447–2454.
- Sonesson, A.K., and Meuwissen, T.H. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genetics Selection Evolution* *41*, 37.
- Su, G., Brøndum, R.F., Ma, P., Guldbandsen, B., Aamand, G.P., and Lund, M.S. (2012). Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *Journal of Dairy Science* *95*, 4657–4665.
- Suren, H., Hodgins, K.A., Yeaman, S., Nurkowski, K.A., Smets, P., Rieseberg, L.H., Aitken, S.N., and Holliday, J.A. (2016). Exome capture from the spruce and pine giga-genomes. *Molecular Ecology Resources* *16*, 1136–1146.
- Tan, B., Grattapaglia, D., Martins, G.S., Ferreira, K.Z., Sundberg, B., and Ingvarsson, P.K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F1 hybrids. *BMC Plant Biology* *17*, 110.

- Thistlethwaite, F.R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M.U., and El-Kassaby, Y.A. (2017). Genomic prediction accuracies in space and time for height and wood density of Douglas-fir using exome capture as the genotyping platform. *BMC Genomics* 18, 930.
- Thistlethwaite, F.R., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., Stoehr, M.U., and El-Kassaby, Y.A. (2019). Genomic selection of juvenile height across a single-generational gap in Douglas-fir. *Heredity*.
- Thompson, R. (2008). Estimation of quantitative genetic parameters. *Proceedings of the Royal Society B: Biological Sciences* 275, 679–686.
- Thomson, M.J. (2014). High-Throughput SNP Genotyping to Accelerate Crop Improvement. *Plant Breeding and Biotechnology* 2, 195–212.
- Ukrainetz, N.K., Kang, K.-Y., Aitken, S.N., Stoehr, M., and Mansfield, S.D. (2008). Heritability and phenotypic and genetic correlations of coastal Douglas-fir (*Pseudotsuga menziesii*) wood quality traits. *Canadian Journal of Forest Research* 38, 1536–1546.
- Van Eenennaam, A.L., Weigel, K.A., Young, A.E., Cleveland, M.A., and Dekkers, J.C.M. (2014). Applied Animal Genomics: Results from the Field. *Annual Review of Animal Biosciences* 2, 105–139.
- VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F., and Schenkel, F.S. (2009). Invited Review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92, 16–24.
- Varshney, R.K., Roorkiwal, M., and Sorrells, M.E. (2017). *Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement*. (Switzerland: Springer International Publishing).
- Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M.K., Thornton, A.M., Roeb, W., Abu Rayyan, A., Loulus, S., Avraham, K.B., King, M.-C., et al. (2010). Whole Exome Sequencing and Homozygosity Mapping Identify Mutation in the Cell Polarity Protein GPM2 as the Cause of Nonsyndromic Hearing Loss DFNB82. *The American Journal of Human Genetics* 87, 90–94.
- Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F., and Xiang, J. (2017). Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. *BMC Genetics* 18.
- White, T.L., Adams, W.T., and Neale, D.B. (2007). *Forest genetics* (Wallingford: CABI).
- Whittaker, J.C., Thompson, R., and Denham, M.C. (2000). Marker-assisted selection using ridge regression. *Genet. Res.* 75, 249–252.
- Woolliams, J., and Thompson, R. (1994). A theory of genetic contributions. In *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, C. Smith, J.S. Gavora, B. Benkel, J. Chesnais, W. Fairfull, J.P. Gibson, B.W. Kennedy, and E.B. Burnside, eds. (Guelph), pp. 127–134.
- Wray, N.R., and Thompson, R. (1990). Prediction of rates of inbreeding in selected populations. *Genetical Research* 55, 41.
- Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist* 56, 330–338.

- Wu, X., Lund, M.S., Sun, D., Zhang, Q., and Su, G. (2015). Impact of relationships between test and training animals and among training animals on reliability of genomic prediction. *Journal of Animal Breeding and Genetics* 132, 366–375.
- Xu, S. (2003). Theoretical Basis of the Beavis Effect. *Genetics* 165, 2259.
- Yanchuk, A.D. (1996). General and specific combining ability from disconnected partial diallels of coastal Douglas-fir. *Silvae Genetica* 45, 37–45.
- Yeh, F.C., and Heaman, C. (1982). Heritabilities and genetic and phenotypic correlations for height and diameter in coastal Douglas-fir. *Canadian Journal of Forest Research* 12, 181–185.
- Zapata-Valenzuela, J., Isik, F., Maltecca, C., Wegrzyn, J., Neale, D., McKeand, S., and Whetten, R. (2012). SNP markers trace familial linkages in a cloned population of *Pinus taeda*—prospects for genomic selection. *Tree Genetics & Genomes* 8, 1307–1318.
- Zapata-Valenzuela, J., Whetten, R.W., Neale, D., McKeand, S., and Isik, F. (2013). Genomic Estimated Breeding Values Using Genomic Relationship Matrices in a Cloned Population of Loblolly Pine. *Genes|Genomes|Genetics* 3, 909–916.
- Zhang, Z., Ding, X., Liu, J., Zhang, Q., and de Koning, D.-J. (2011). Accuracy of genomic prediction using low-density marker panels. *Journal of Dairy Science* 94, 3642–3650.
- Zhong, S., Dekkers, J.C.M., Fernando, R.L., and Jannink, J.-L. (2009). Factors Affecting Accuracy From Genomic Selection in Populations Derived From Multiple Inbred Lines: A Barley Case Study. *Genetics* 182, 355–364.
- Zikhali, M., Leverington-Waite, M., Fish, L., Simmonds, J., Orford, S., Wingen, L.U., Goram, R., Gosman, N., Bentley, A., and Griffiths, S. (2014). Validation of a 1DL earliness per se (eps) flowering QTL in bread wheat (*Triticum aestivum*). *Molecular Breeding* 34, 1023–1033.
- Zobel, B., and Talbert, J. (1984). *Applied forest tree improvement* (New York: Wiley).