ABSTRACT

IBARRA OYARZUN, LUIS ALEJANDRO. Quantitative Genetic on Hybrid Populations of Trees: Estimations of Genetic Parameters and Implications for Breeding Strategies. (Under the direction of Dr. Gary R. Hodge and Dr. Juan Jose Acosta).

Many combinations of hybrid trees have been made in the genera *Eucalyptus* and *Pinus* for the objectives of wood and cellulose production. In South Africa and Chile, field tests with hybrid populations of trees were established, and growth and wood properties measurements recorded. Quantitative genetic parameters were estimated, fitting a model that considers the effects of parent General Hybridizing Ability (GHA) for species A and B, and Specific Hybridizing Ability (SHA) for the combination of parents), along with Genetic x Environment interactions. In addition, a stochastic simulation of full-sib hybrid progeny tests was done with various mating designs and underlying true genetic parameters, where the target was to describe the accuracy and variation of the genetic parameter estimates.

A clonal hybrid population of *Eucalyptus nitens x E. globulus* assessed in two breeding zones in Chile revealed large genetic variation in volume and wood properties. Significant genetic gain is possible for all traits analyzed. The *E. nitens* parent had a considerable impact on the volume gain of the hybrid, which implies that testing more *E. nitens* parents in future interspecific crosses will assure more volume gain. In contrast, *E. globulus* demonstrated zero GHA variance for volume, but did show important GHA variance for wood traits (Basic density (m^3/kg), Pulp Yield (%), and Specific Consumption (m^3/ADt)). In the Arauco zone, *E. globulus* demonstrated a large effect on the genetic variability of these traits; meanwhile, in the Valdivia zone, *E. nitens* and *E. globulus* were roughly similar in the size of the GHA variance. There was a strong relationship between parental performance in pure species crosses (GCA) and hybrid crosses (GHA), which indicates that parents could be selected for interspecific crosses based on pure species test results for volume and wood properties.

In clonal and seedling hybrid populations of *Pinus patula x P. tecunumanii* (low elevation, TECL) in South Africa, a large impact of *P. tecunumanii* (TECL) was observed compared to

P. patula. TECL exhibited higher GHA variance for volume gain in both populations. For new hybrid crosses, a more intense selection of parents should be made among TECL parents to increase volume gain. In a hybrid seedling population of *P. patula x P. tecunumanii* (high elevation) (PATxTECH), *P. patula* had a large impact on the volume gain and MOE, while TECH had a zero estimate of GHA variance for volume and a low GHA variance for MOE. Thus, more PAT parents should be tested in new hybrid crosses of this variety to increase volume and MOE gain.

In the stochastic simulation, several scenarios with a range of genetic variances and mating designs were tested to understand which are optimum mating designs for obtaining unbiased and precise estimates of GHA and SHA variance. A combination of 24 parents tested in 4 crosses gave the most precise estimates for GHA and SHA variance. Zero GHA variance estimates are more likely to be obtained when a low number of crosses per parent are tested, especially in the presence of high SHA variance. For GHA variances, serious underestimates can be observed under some conditions. For example, in scenarios with a low number of parents and/or low number of crosses, the frequency of zero GHA variance estimates (despite non-zero true GHA variance) was as high as 11% over 1000 iterations. In the real data analyzed from the hybrid population of *E. globulus x E. nitens* and *P. patula x P. tecunumanii* (high elevation), zero GHA variance estimates were found for one of the species for volume; in both of those populations, the mating designs utilized a low number of parents and crosses (or the combination of both).

Quantitative Genetics on Hybrid Populations of Trees: Estimations of Genetic Parameters and Implications for Breeding Strategies.

by
Luis Alejandro Ibarra Oyarzún

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Forestry and Environmental Resources

Raleigh, North Carolina
2021

APPROVED BY:

_____	_____
Dr. Gary R. Hodge Committee Co-Chair	Dr. Juan Jose Acosta Committee Co-Chair
_____	_____
Dr. Kitt G. Payn	Dr. Christian Maltecca

DEDICATION

I dedicate this research to my family, who are the most important part of my life. To my wife Vanessa, for her patience, unconditional love, and support during this journey to obtain this degree. I am very lucky and fortunate to have you in my life. To my children Benjamin and Samantha, from whom I learned that happiness is just to take them into my harm, even when I felt so tired and overwhelmed!

I especially feel gratitude for my parents Mabel and René, for teaching me that with your own effort, you can achieve any goal that you proposed and for their support when I decided to study forestry in my undergrad.

BIOGRAPHY

Luis was born in Chiloe Island, in the city of Castro, Chile, in 1982. Convinced since his youth that Forestry was the career that he wanted to pursue, he completed his high school in 2000 in the Technical Forestry College of Quilacahuin, located in the rural area of San Pablo, Chile. There he learned the basic technical aspects of plant production in the nursery, tree planting, pruning, thinning techniques, and the basis of forest inventory and data analysis. After that, he studied Forestry Engineering at the Universidad Austral of Valdivia, Chile, finishing his undergraduate program in 2008. His thesis work was Growth and Quality of *Nothofagus dombeyi* plantations established in the intermediate depression of Valdivia, under the direction of Pablo Donoso Hiriart. In 2009 he began to work in the research facility of the Chilean company Arauco Forestry (Bioforest) in the Genetics Division. He worked specifically in the evaluation of genetic field tests, using genetic parameters to estimate the performance of family and clonal progeny and predict breeding values. In the fall of 2017, he started to work as a Research Assistant with the Camcore program in the Department of Forestry & Environmental Resources at the North Carolina State University in Raleigh, with financial support from Camcore and from Bioforest (Arauco Company). The goal was to pursue a Ph.D. in Forest Genetics Breeding under the supervision of Dr. Gary Hodge and Dr. Juan Jose Acosta. During his time in Raleigh with his lovely wife Vanessa Ramirez, he became the happy father of two healthy kids, Benjamin, born in 2018, and Samantha, born in 2020. Despite the disruption of normality due to the COVID-19 pandemic, which impacted the regular operations of the University and the entire world, co-advisors and committee members worked hard with him to finish his research on time. His Ph.D. research focused on the estimates of quantitative genetic parameters in hybrid populations of trees and their implications for breeding strategies.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisors Dr. Hodge and Dr. Acosta, for their continuous support, patience, and many good pieces of advice during my Ph.D. studies. I learned so much from them, and I will always be in gratitude for that. I would like to thank Bioforest for their financial support and special thanks to Christian De Veer and Claudio Balocchi for believing in me achieving the Ph.D. degree. I also want to thank the Camcore team for their friendship and support, especially to Romeo, for their invaluable help when I most need it. I would like to thank as well to the LASA community, from where I found many friends and priceless memories. My appreciation also goes to the volunteers of the English Conversation Club, where we were lucky to find many good friends. And special thanks to Andy and Cheryl, for their support and good moments.

Finally, I would like to thank my wife, children, and parents for their love, encouragement, and constant support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1:

Introduction

1.1 Quantitative Genetics in Forest Tree Breeding

Genetic enhancement, also known as genetic breeding, can be defined as the improvement of particular traits of interest in trees, plants, and animals. The term forest tree breeding implies selection for one or more valuable characteristics or traits in tree species that have important economic value. Many of the traits improved in forest tree breeding, such as growth rate, wood properties, disease, and pest resistance, are controlled by several genes and their interactions, or in other words, they display polygenic inheritance (Burdon, 2004; Burley, 2004; Libby et al., 1969). In this regard, most of the traits mentioned above are more quantitative in nature than qualitative (or simply inherited).

Quantitative genetics is the study of the inheritance of characteristics (or traits), and it is founded on the assumption of a model in which many genes affect the trait and in which non-genetic factors may also be involved. Considering that many genes and the environment interact to result in the phenotypic expression of a trait, it would be difficult, if not impossible, to determine the action of individual gene loci affecting a trait. Quantitative genetics deals with the assessment of genetic and environmental parameters, including phenotypic and genetic variances, covariances and correlations, the estimation of heritability, and the response to selection, and normally requires artificial regeneration to be effectively accomplished (Burley, 2004; Hill, 2010).

Statistical methods were invented by authors R.A. Fisher and S.G Wright and others beginning in the early 1920s. Important methodologies, such as analysis of variance and the use of path coefficients, are used to partition observed variation and describe the resemblance between relatives. Many of the tools and methods developed in quantitative

genetics have had widespread application in disciplines ranging far from their original targets (Hill, 2010).

1.1.1 The Animal Model

The estimation of genetic parameters is fundamental to understanding if a trait of interest is under genetic control and can be inherited through selection. A statistical measure of the degree to which a trait can be inherited, called *heritability,* is one of the most important parameters in quantitative genetics (Falconer & Mackay, 1996; Lynch & Walsh, 1998). With unbalanced datasets and complex pedigree information, the estimation of this parameter, and other genetic parameters, is not possible using standard analysis of variance techniques. However, a mixed linear model was developed for many genetic applications (sometimes called the 'animal model') in which the phenotype of each individual is defined in terms of a combination of fixed and random effects, and the genetic structure is incorporated in the variances and covariances of these effects (Hill, 2010).

The analysis of quantitative data in tree breeding programs across a variety of sites can be done with this animal model, which has this general form:

$$y = \boldsymbol{X}b + \boldsymbol{Z}u + \boldsymbol{e} \qquad \text{Eq. 1.1}$$

with

$$\text{var}(\boldsymbol{y}) = \boldsymbol{ZGZ'} + \boldsymbol{R} = \boldsymbol{V};\ \text{var}(\boldsymbol{u}) = \boldsymbol{G};\ \text{and var}(\boldsymbol{e}) = \boldsymbol{R} \qquad \text{Eq. 1.2}$$

where $\boldsymbol{y}$ represents a nx1 vector of observations, n is the number of records, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are n×f, and n×r matrices are representing design matrices for the fixed and random effects in the f×1 and r×1 vectors $\boldsymbol{b}$ and $\boldsymbol{u}$, where f and r are the numbers of levels of fixed and random effects respectively. The random-effects vector $\boldsymbol{u}$, and the residual vector $\boldsymbol{e}$, are assumed to be multivariate normally distributed. The matrices $\boldsymbol{V}$, $\boldsymbol{G}$, and $\boldsymbol{R}$ represent the variance matrices of $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{e}$, correspondingly (Thompson, 2008). Since this linear model includes both fixed and random effects ($\boldsymbol{u}$ and $\boldsymbol{b}$), it is called a linear mixed model. The fixed effects are estimated by Best Linear Unbiased Estimation (BLUE). In contrast, random effects are

estimated by Best Linear Unbiased Prediction (BLUP), usually obtained using Restricted Maximum Likelihood (REML) estimates of the variances and covariances of the random effects. The solutions for both BLUE and BLUP can be obtained by the following mixed model equations, developed by Henderson (1950) and cited by Henderson (1975):

$$\begin{bmatrix} \boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z} \\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}+G^{-1}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{b}} \\ \widehat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'R^{-1}y} \\ \boldsymbol{Z'R^{-1}y} \end{bmatrix} \quad \text{Eq. 1.3}$$

This set of equations includes terms for fixed and random effects, and estimates solutions for fixed effects (***b),*** and predicts solutions for random effects (***u***) simultaneously (Mrode, 2014; Piepho et al., 2008; Thompson, 2008).

For data analysis in forest tree genetic breeding programs, there is interest in the estimation of the genetic variances and covariances in **G**, adjusting the data for the fixed effects ***b*** (e.g., sites), and there is also interest in predicting the random effects, ***u*** (e.g., breeding values).

This animal model, which can be referred to as the "individual tree model" or “individual model” when it is applied to tree species, can be written as a linear combination of the overall mean (*u*), the breeding value of an individual (*a*), and the error (*e*) in the following way:

$$y_i = u + a_i + e_i \quad \text{Eq. 1.4}$$

Using **G** (Eq. 1.3) as the variance-covariance matrix for ***u***, the expected covariance between relatives in additive genetic effects is G=A*σ_a^2, and thus **u**~N (0, **G**). In this way, **A** is the additive genetic relatedness matrix, which can be derived from a mating pedigree or from genomic relatedness data. By using the additive relatedness matrix (**A**), this animal model relates trait covariances across individuals to their genetic relatedness and is able to separate the additive genetic variance (σ_a^2, variance in a_i Eq. 1.4) from the error variance (σ_e^2, variance in e_i Eq. 1.4) (Isik et al., 2017; Thomson et al., 2018). Then, the total phenotypic variance is the sum of the additive and residual variances:

$$\sigma_p^2 = \sigma_a^2 + \sigma_e^2 \quad \text{Eq. 1.5}$$

1.1.2 Genetic Parameters Estimations

The number of genetic parameter estimates that we can obtain for a breeding population of trees will depend on the degree of relatedness from which we have information. For example, in open-pollinated (OP) trials, we will be able only to separate the additive effect of each mother from the total phenotypic variance, at least when only pedigree information is available. Other methods using molecular marker data could allow the breeder to reconstruct the precise relatedness among individuals, which could then be used to form a pedigree relationship of the trees. This general approach has sometimes been called "Breeding without breeding" (e.g., see El-Kassaby & Lstibůrek, (2009)), highlighting the fact that complex pedigrees can be extracted from simple mating schemes, without the need for "breeding" with complex mating designs.

The term General Combining Ability (GCA) refers to the average performance of a progeny from a single parent (in the OP example, from a mother tree) relative to the population mean. Using the variation associated with the population of GCA effects, breeders can estimate the additive genetic variance for a given trait in a population. GCA variance corresponds to one-quarter of the additive genetic variance. Therefore, the parent's breeding value (BV) is 2-times the GCA (2*GCA) since the parent contributes, on average, one-half of its total additive genetic value to each progeny.

In control-pollinated trials (CP), where for the progeny population, the identity of both parents of each progeny is known, in addition to the GCA of the parents, the Specific Combining Ability (SCA) or a particular cross can be estimated. SCA is a term that refers to the tendency of specific combinations of parents to produce progeny with an average performance that deviates from the expected parent values (i.e., the sum of the two GCAs). Progeny that performs better than the sum of the GCA of the parents will have a positive SCA value, and a negative SCA value reflects that the progeny perform worse than the expected mid-parent value (Fukatsu et al., 2014; Lynch & Walsh, 1998; White et al., 2007). The variance among SCA values corresponds to one-quarter of the dominance genetic variance.

In OP genetic tests, only additive variance can be quantified, and in CP genetic tests, both additive and dominance genetic variance can be quantified. In clonal field tests, additional non-additive genetic variance (epistasis) can also be estimated. If tests with common genetic entries have been planted in multiple environments, then the separation of the variance components can be done for additive, dominance, and epistasis effects along with their respective interactions with the environment. An example of a full linear mixed model (LMM), with the GCA, SCA, Clone, and the interaction of each term with the environment (E) is presented below:

$$y_{ijklm} = \mu + E_i + GCA_j + GCA_k + SCA_{jk} + Clone_l + GCAxE_{ij} + SHAxE_{ijk} + ClonexE_{il} + err_{ijklm} \quad \text{Eq. 1.6}$$

Where:

y is the m^{th} observation of the l^{th} clone within the k^{th} family at the i^{th} Environment; u is the overall population mean; E_i is the fixed effect of i^{th} Environment; GCA_j or GCA_k is the random GCA effect for j^{th} female or k^{th} male parent; SHA_{ij} is the random SCA_{jk} effect of the j^{th} and k^{th} parents; $Clone_l$ is the random clone within family effect for the l^{th} clone; $GCAxE_{ij}$ is the random GCA by Environment interaction of i^{th} environment and j^{th} female or k^{th} male GCA ; $SHAxE_{ijk}$ is the random SCA by Environment interaction of i^{th} environment and jk^{th} family; $ClonexE_{il}$ is the random clone by Environment interaction of i^{th} Environment and l^{th} clone and err_{ijklm} is the random error.

All the effects in this model (Eq. 1.6), except for overall mean (μ) and site effect (E_i), are assumed to be random and independently distributed.

Through the estimates of the variance components of GCA, SCA, Clone and their interaction with the environment (different sites where the trees were established), important genetic parameters can be obtained by following the derivation of Falconer & Mackay (1996). The additive, dominance, total genetic, and phenotypic variances are estimated as follow:

Additive variance : $\sigma_a^2 = 4\sigma_{GCA}^2$ Eq. 1.7

Dominance variance: $\sigma_d^2 = 4\sigma_{SCA}^2$ Eq. 1.8

Genetic variance : $\sigma_G^2 = 2\sigma_{GCA}^2 + \sigma_{SCA}^2 + \sigma_{Clone}^2$ Eq. 1.9

Phenotypic variance : $\sigma_{phen}^2 = \sigma_G^2 + 2\sigma_{GCAxE}^2 + \sigma_{SCAxE}^2 + \sigma_{ClonexE}^2 + \sigma_{err}^2$ Eq. 1.10

The narrow-sense heritability (h^2), broad-sense heritability (H^2), and dominance effect (d^2) are usually estimated with the appropriate variance component estimates in the following way:

Narrow-sense heritability : $h^2 = \sigma_a^2/\sigma_{phen}^2$ Eq. 1.11

Proportion of Dominance : $d^2 = \sigma_d^2/\sigma_{phen}^2$ Eq. 1.12

Broad-sense heritability : $H^2 = \sigma_G^2/\sigma_{phen}^2$ Eq. 1.13

The type-B genetic correlation (rB), described by Burdon (1977), is a measure of the degree of Genotype × Environment interaction (GxE) when we test the same trait in genotypes established in different environments (sometimes called multi-environment trial analysis, MET). The magnitude of the type-B genetic correlation ranges from 0 to 1, where values near 1.0 indicate a near-perfect correlation between the performance of the genotypes for the trait of interest measured across the different environments (and, therefore, a low level of GxE). A type-B genetic correlation near zero indicates a weak correlation between the performance of the genotypes across environments (i.e., a high level of GxE). In other words, a stable ranking of the genotypes across sites is an indication of low GxE (rB near 1.0), and unstable rankings across sites indicate high GxE (rB near to zero). The amount of GxE is a crucial parameter for tree breeders to know to optimize genotype selection for stable performance across future planting sites (Hodge & Dvorak, 2015).

Type-B genetic correlations can be obtained for all sources of genetic variation: total genetic effects, clone within family, parent, and cross estimated as:

Total genetic effects : $rB_G = (\sigma_G^2)/(\sigma_G^2 + 2\sigma_{GCAxE}^2 + \sigma_{SCAxE}^2 + \sigma_{ClonexE}^2)$ Eq. 1.14

Clone within family : $rB_{Clone} = \sigma_{Clone}^2/(\sigma_{Clone}^2 + \sigma_{ClonexE}^2)$ Eq. 1.15

GCA : $rB_{GCA} = \sigma_{GCA}^2/(\sigma_{GCA}^2 + \sigma_{GCAxE}^2)$ Eq. 1.16

SCA : $rB_{SCA} = \sigma_{SCA}^2/(\sigma_{SCA}^2 + \sigma_{SCAxE}^2)$ Eq. 1.17

The predicted breeding values of each parent can be estimated as twice the GCA, since each parent contributes one-half the breeding value of its progeny, considering that each parent contributed 1n allele to the 2n alleles belonging to each progeny (White et al., 2007).

1.1.3 Genetic Parameters in Hybrid Populations of Trees

In hybrid populations, genetic values can be estimated related to the genetic worth of the parents from the two species and the effect of the specific combinations in these interspecific crosses. These values and their related variances are analogous to those from a pure species scenario. In a hybrid combination of trees, the combining ability of the parent is called General Hybridizing Ability (GHA), which is the genetic worth of a hybrid parent based on the average performance of its progeny relative to the population mean. The Specific Hybridizing Ability (SHA) is the deviation of a hybrid progeny family relative to the sum of the GHA values of their parents (Nikles et al., 1991; White et al., 2007).

The estimation of each variance component, related to the GHA for parents of species A and B, the SHA of the specific parent-species combination across their respective interaction with the environment it is estimated in a model which incorporates the effect of both parents. An example of the model (LMM) is presented next:

$$y_{ijkl} = \mu + E_i + GHA_{A_j} + GHA_{B_k} + SHA_{jk} + GHA_A xE_{ij} + GHA_B xE_{ik} + SHAxE_{ijk} + err_{ijkl}$$

Eq. 1.18

Where:

y is the l^{th} observation of the jk^{th} hybrid family at the i^{th} Environment; u is the overall population mean; E_i is the fixed effect of i^{th} Environment; GHA_{A_j} or GHA_{B_k} is the random GHA effect for j^{th} Species A female or k^{th} Species B male parent; SHA_{jk} is the random SHA effect of the j^{th} and k^{th} parents; $GHA_A xE_{ij}$ or $GHA_B xE_{ik}$ is the random GHA by Environment interaction of i^{th} environment and j^{th} Species A female or k^{th} Species B male parent; $SHAxE_{ijk}$ is the random SHA by Environment interaction of i^{th} environment and j^{th} and k^{th} parents and err_{ijkl} is the random error. All the effects in this model, with the exception of

the overall mean and the site effect, were assumed to be random and independently distributed.

With the variance components obtained from the full hybrid model, the genetic parameters can be derived, in a similar way to what is done with pure species, using GHA variance of each parental species, the SHA variance for the specific hybrid crosses, and their respective interactions with the environment in the following manner:

Species A GHA variance : $\sigma^2_{a_A} = 4\sigma^2_{GHA_A}$ Eq. 1.19

Species B GHA variance : $\sigma^2_{a_B} = 4\sigma^2_{GHA_B}$ Eq. 1.20

Proportion of dominance : $\sigma^2_d = 4\sigma^2_{SHA}$ Eq. 1.21

Genetic variance : $\sigma^2_G = 2(\sigma^2_{GHA_A} + \sigma^2_{GHA_B}) + 4\sigma^2_{SHA}$ Eq. 1.22

Phenotypic variance : $\sigma^2_{phen} = \sigma^2_G + \sigma^2_{GHA_AxE} + \sigma^2_{GHA_BxE} + \sigma^2_{SHAxE} + \sigma^2_{err}$ Eq. 1.23

Often, authors working with hybrid populations will estimate a “hybrid narrow-sense heritability” for each parental species by multiplying the parental GHA variance by 4 and then dividing by the total phenotypic variance (e.g., Dieters et al., 1997; Madhibha et al., 2013; Mitchell et al., 2013; van den Berg et al., 2015; Zhu et al., 2017). A word of caution here, the sum of the heritabilities and the proportion of dominance will be much higher than the broad-sense heritability (H^2), since this approach calculates two "full" narrow-sense heritability estimates for the two different species. These two "hybrid narrow-sense heritabilities" are mostly used to compare the size of the GHA variance associated with each parent species in the familiar form of a ratio of "additive" variance over the phenotypic variance. However, this comparison could be made without estimating these heritabilities, just looking at GHA variances directly.

To clarify this, the narrow-sense heritability of each species (h^2_A or h^2_B) broad-sense heritability (H^2) and dominance effect (d^2) are sometimes estimated following Falconer & Mackay (1996) derivations as:

Narrow-sense heritability for species A : $h^2_A = 4\,\sigma^2_{GHA_A}/\sigma^2_{phen}$ Eq. 1.24

Narrow-sense heritability for species B : $h^2_B = 4\,\sigma^2_{GHA_B}/\sigma^2_{phen}$ Eq. 1.25

Dominance effect : $d^2 = 4\,\sigma^2_{SHA}/\sigma^2_{phen}$ Eq. 1.26

Broad-sense heritability : $H^2 = \sigma^2_G/\sigma^2_{phen}$ Eq. 1.27

Similar to the approach used for pure species, the GxE of each hybrid genetic variance component can be measured through type-B genetic correlations for all genetic effects in the following manner:

All genetic effects : $rB_G = (\sigma^2_G)/(\sigma^2_G + 2\sigma^2_{GHA_AxE} + 2\sigma^2_{GHA_BxE} + 4\sigma^2_{SHAxE})$ Eq. 1.28

GHA species A : $rB_{GHA_A} = \sigma^2_{GHA_A}/(\sigma^2_{GHA_A} + \sigma^2_{GHA_AxE})$ Eq. 1.29

GHA species B : $rB_{GHA_B} = \sigma^2_{GHA_B}/(\sigma^2_{GHA_B} + \sigma^2_{GHA_BxE})$ Eq. 1.30

SHA : $rB_{SHA} = \sigma^2_{SHA}/(\sigma^2_{SHA} + \sigma^2_{SHAxE})$ Eq. 1.31

With this model, a predicted GHA value of each species-parent tested in a hybrid combination is obtained.

1.1.4 Relationship Between GCA and GHA

Forest tree breeders working with hybrids will often maintain breeding populations of both parental species, and periodically, new hybrid crosses are made between these populations. Usually, intra-species recurrent selection is made in the parental populations and shown to be efficient for hybrid improvement by selecting the best parent for a given trait within each population for future interspecific species crosses. This method is known as Recurrent Selection for General Combining Ability (RSGCA), and it requires hybrid crosses only for deployment. First, however, it is necessary to know if there is a correspondence between the pure species GCA value of each parent with their hybrid GHA to see if the selection of parents based on intra-specific crosses will identify parents that will also be superior for inter-specific crosses. A high correlation between these parameters (GHA and GCA) indicates that parents can be selected for interspecific crosses based on their pure species performance for a given trait.

When the GCA is not a good predictor of the GHA, due to a high non-additive effect, such as dominance (SHA or SCA) or epistasis effects, the Reciprocal Recurrent Selection (RRS)

is another breeding strategy that can be used to develop new hybrid populations. This method is known to many breeders, and it was defined by Comstock et al. (1949) as a method of selection of inbred lines through SCA values in maize.

Both breeding strategies have been adopted successfully in forestry through the years, with interspecific crosses of species of the genus Populus, Eucalyptus, and Pinus, to mention some of them.

Parental GCA values can be compared with their GHA value via genetic correlations or Pearson correlations to evaluate the consistency of the parental performance evaluated as pure species parents and as hybrid parents, and such correlations will be designated r_{HP} in this manuscript. A high correlation would indicate that parental GCA could be a good indicator for GHA performance for a given trait. In a scenario where GCA is a good predictor of GHA, interspecific crosses can be planned based on the parent performance in pure and/or hybrid species performance. Without at least a moderately strong correlation between GCA and GHA, future interspecific crosses could be done only based on GHA values obtained for each parent based on their hybrid progeny performance, giving a breeder the only option of using RRS for new hybrid progeny tests.

1.1.5 Quantitative Genetic Parameter Estimates in Hybrid Tree Populations

In this research, two different populations of hybrid trees were evaluated through quantitative genetic analysis to examine the performance of hybrid progeny in growth and wood properties traits across different environments in the southern hemisphere.

One of these populations of trees consists of hybrids between *Eucalyptus nitens x E.* globulus, located in Chile (Chapter 3), where several hybrid progeny families were clonally replicated and tested across 27 field tests in two contrasting breeding zones. In total, more than 1,200 clones were evaluated for volume and three wood properties important for pulp production.

The other hybrid population consists of hybrids between *Pinus patula x P. tecunumanii* in South Africa and Eswatini (Chapter 4). Two South African companies have been evaluating growth and timber strength in these populations. Each company had a distinct population of hybrids formed from different *P. patula* and *P. tecunumanii* parents. One company tested 137 families (40 and 97 hybrid families of two different varieties of the hybrid) as seedling progeny across a total of 5 tests, and the other company tested 26 hybrid families with clonal replication in 2 tests.

In both the *E. nitens x E. globulus* and the *P. patula x P. tecunumanii* populations, important genetic parameters were estimated, including σ^2_{GHA} for both species, σ^2_{SHA}, and H^2. Also, for the two populations, where possible, predicted GHA values for different parents were compared with their homologous parental GCA values to discover any possible relationship between hybrid performance and the known parental performance as pure species. These genetic parameters, including the GCA-GHA correlation (r_{HP}), can help breeders understand how much genetic gain can be made for a given trait and can be used to evaluate possible implications for the hybrid breeding strategies of these varieties.

A stochastic simulation (Chapter 2) was performed, testing different scenarios related to the hybrid genetic parameter estimates. These scenarios were composed for several combinations of values of genetic parameters, intended to represent a range of possible real situations for hybrid tree populations, and for a number of different mating designs, with variable numbers of parents and crossed tested. The goal was to examine how the accuracy and precision of the parameter estimates would change related to variations in the mating design, fluctuation in the number of parents and crosses, different sizes of genetic variances ($\sigma^2_{GHA_A}$, $\sigma^2_{GHA_B}$, and σ^2_{SHA}). For each scenario, 1000 iterations were run, and the behavior of the estimates in terms of accuracy, precision, and frequency of severe errors in estimation was described.

In summary, in this dissertation, quantitative genetic methods were applied to two different hybrid populations of trees to estimate important genetic parameters in growth and wood properties, and to examine relationships between the parental performance in hybrid and pure species combinations. In addition, a stochastic simulation testing of different

scenarios was done to elucidate to what degree the results obtained in the true hybrid populations studied can be expected to be accurate, precise, and reliable.

1.2 References

Burdon, R. D. (1977). Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. Silvae Genet., 26(5–6), 168–175.

Burdon, R. D. (2004). Tree Breeding, Principles: Forest Genetics and Tree Breeding; Current and Future Signposts. Encyclopedia of Forest Sciences, 44(1995), 1538–1545. https://doi.org/10.1016/b0-12-145160-7/00074-0

Burley, J. (2004). Tree Breeding, Principles: A Historical Overview of Forest Tree Improvement. In Encyclopedia of Forest Sciences (Vol. 21, Issue 31, pp. 1532–1538). Elsevier. https://doi.org/10.1016/B0-12-145160-7/00073-9

Dieters, M. J., Nikles, D. G., Toon, P. G., & Pomroy, P. (1997). Genetic parameters for F1 hybrids of Pinus caribaea var. hondurensis with both Pinus oocarpa and Pinus tecunumanii. Canadian Journal of Forest Research, 27(7), 1024–1031. https://doi.org/10.1139/x97-053

El-Kassaby, Y. A., & Lstibůrek, M. (2009). Breeding without breeding. Genetics Research, 91(2), 111–120. https://doi.org/10.1017/S001667230900007X

Falconer, D. . S., & Mackay, T. F. . (1996). Introduction to quantitative genetics (4th edn). Pearson. https://doi.org/10.1002/bimj.19620040211

Fukatsu, E., Hiraoka, Y., Matsunaga, K., Tsubomura, M., & Nakada, R. (2014). Genetic relationship between wood properties and growth traits in Larix kaempferi obtained from a diallel mating test. Journal of Wood Science, 61(1), 10–18. https://doi.org/10.1007/s10086-014-1436-9

Henderson, C. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics, 31(2), 423–447. https://www.jstor.org/stable/2529430

Hill, W. G. (2010). Understanding and using quantitative genetic variation. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1537), 73–85. https://doi.org/10.1098/rstb.2009.0203

Hodge, G. R., & Dvorak, W. S. (2015). Provenance variation and within-provenance genetic parameters in Eucalyptus urophylla across 125 test sites in Brazil, Colombia, Mexico, South Africa and Venezuela. Tree Genetics and Genomes, 11(3). https://doi.org/10.1007/s11295-015-0889-3

Isik, F., Holland, J., & Maltecca, C. (2017). Genetic Data Analysis for Plant and Animal Breeding. In Genetic Data Analysis for Plant and Animal Breeding. https://doi.org/10.1007/978-3-319-55177-7

Libby, W. J., Stettler, R. F., & Seitz, F. W. (1969). Forest Genetics and Forest-Tree Breeding. Annual Review of Genetics, 3(1), 469–494. https://doi.org/10.1146/annurev.ge.03.120169.002345

Lynch, M., & Walsh, B. (1998). Genetic and Analysis of Quantitative Traits. Sinauer. https://books.google.com/books?id=UhCCQgAACAAJ

Madhibha, T., Murepa, R., Musokonyi, C., & Gapare, W. (2013). Genetic parameter estimates for interspecific Eucalyptus hybrids and implications for hybrid breeding strategy. New Forests, 44(1), 63–84. https://doi.org/10.1007/s11056-011-9302-8

Mitchell, R. G., Wingfield, M. J., Hodge, G. R., Steenkamp, E. T., & Coutinho, T. A. (2013). The tolerance of Pinus patula × Pinus tecunumanii, and other pine hybrids, to Fusarium circinatum in greenhouse trials. New Forests, 44(3), 443–456. https://doi.org/10.1007/s11056-012-9355-3

Mrode, R. . (2014). Linear Models for the Prediction of Animal Breeding Values. In CAB INTERNATIONAL (Ed.), Linear Models for the Prediction of Animal Breeding Values 3rd Edition (Vol. 3).

Nikles, D. G., Council, A. F., & Newton, R. S. (1991). Correlations of breeding values in pure and hybrid populations of hoop pine and some southern pines in Queensland and relevance to breeding strategies. Proceedings of the 11th Meeting of the Research Working Group 1 of the Australian Forestry Council, Coonawarra, South Australia, 1–5.

Piepho, H. P., Möhring, J., Melchinger, A. E., & Büchse, A. (2008). BLUP for phenotypic

selection in plant breeding and variety testing. Euphytica, 161(1–2), 209–228. https://doi.org/10.1007/s10681-007-9449-8

Thompson, R. (2008). Estimation of quantitative genetic parameters. Proceedings of the Royal Society B: Biological Sciences, 275(1635), 679–686. https://doi.org/10.1098/rspb.2007.1417

Thomson, C. E., Winney, I. S., Salles, O. C., & Pujol, B. (2018). A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. PLOS ONE, 13(10), e0197720. https://doi.org/10.1371/journal.pone.0197720

van den Berg, G. J., Verryn, S. D., Chirwa, P. W., & Deventer, F. Van. (2015). Genetic Parameters of Interspecific Hybrids of Eucalyptus grandis and E. urophylla Seedlings and Cuttings. Silvae Genetica, 64(1–6), 291–308. https://doi.org/10.1515/sg-2015-0027

White, T. L., Adams, W. T., & Neale, D. B. (2007). Forest Genetics. In CABI Publishing. CABI Publishing. https://doi.org/10.1038/167764a0

Zhu, Y., Wu, S. J., Xu, J., Lu, Z., Li, G., Hu, Y., Yang, X., & Bush, D. (2017). Genetic parameters for growth traits and stem-straightness in Eucalyptus urophylla × E. camaldulensis hybrids from a reciprocal mating design. Euphytica, 213(7), 1–16. https://doi.org/10.1007/s10681-017-1923-3

Chapter 2:

Simulation of Quantitative Genetic Parameters in Hybrid Tree Populations

2.1 Introduction

In forestry, hybridization is the successful mating between individual trees of two different species (White et al., 2007), typically resulting in progeny with intermediate characteristics of its parents. In some cases, the hybrid progeny performs better than the average of both parent species, and this phenomenon is called *mid-parent heterosis*. In other cases, the hybrid progeny performs better than the best parent species. This behavior is called *high-parent heterosis* (White et al., 2007), which can be very desirable in developing a new variety of trees.

In natural forests, mating between trees occurs spontaneously when pollen is exchanged between male and female reproductive systems. Sometimes, different species of the same genus may occur in the same environment, with one population overlapping the other, and in these situations, hybrid crosses could happen naturally (Zobel & Talbert, 1984).

In a tree breeding program, making hybrids by crossing trees of two different species is often done with the goal of obtaining specific traits from each parental species (Luan et al., 2013), such as growth rate, resistance to diseases, adaptability to climatic conditions and wood quality. However, the successful hybridization of trees can be challenging to achieve. It is necessary to consider the compatibility of both species, flowering time, and also the direction of pollination, that is, whether to use species A as the female and species B as the male, or vice versa (Potts & Dungey, 2004; Stockwell & Richter, 1947). Often in the development of hybrid crosses of a particular species combination, comparatively few of them are successful, and relatively few parents may be progeny tested compared to the number of parents typically tested in pure classical species breeding programs. In the

hybrid scenario, it can also be difficult to obtain balanced mating designs for progeny testing, with similar numbers of parents and crosses to accurately estimate the genetic worth of the trees as hybrid parents. This matter has to be considered when full-sib crosses are planned in a hybrid breeding program.

Under these circumstances, the estimation of quantitative genetic parameters associated with a given trait or the genetic value of the parents in crosses could be problematic since the accuracy of those estimates depends mainly on the numbers of parents and progeny tested.

In a pure species breeding scenario, *heritability* is a genetic parameter used to explain which part of the total observed phenotypic variation (σ^2_{ph}) is caused by total genotypic variance (σ^2_G) or by additive variance (σ^2_a). The proportion of the phenotypic variation caused by genotypic variation is called *broad-sense heritability* (H^2), and the proportion of the additive genetic variation is called *narrow-sense heritability* (h^2) (Bos & Caligari, 2008; Falconer & Mackay, 1996; Lynch & Walsh, 1998; Yadav et al., 2020). The higher the h^2 value is in a population for a specific trait, the higher the genetic gain that a breeding program can achieve (Hodge & Acosta, 2020).

The genetic worth of a parent can be defined as the ability to pass favorable genes or characteristics to their progeny during the process of fertilization, and some authors refer to this genetic worth as *combining ability* (Bison et al., 2007; Kabir et al., 2014). Two types of combining ability are defined, General Combining Ability (GCA) and Specific Combining Ability (SCA). The term GCA is used to quantify the average performance of progeny related to a single parent compared to the population mean. GCA corresponds to half of a parent breeding value (BV) since a parent contributes, on average, one-half of its total additive genetic value to its progeny. SCA is used to quantify the degree to which particular combinations of parents produce progeny that could perform better or worse than expected based on the sum of the parents GCA values (or the mean of the parental breeding values) (Fukatsu et al., 2014; Lynch & Walsh, 1998; White et al., 2007).

The precise estimation of these genetic parameters is very important in the future prediction gain of any cultivar. The phenotypic variation of a population is influenced by its pool of genes and by the environment. The correct separation of these two main drivers of variation allows better predictions of the breeding values and the Genotype × Environment interaction (GxE). GCA can be considered a measure of the parents' additive genetic effects that can be passed from one generation to another. Precise estimates of GCA variance can help to make precise predictions of specific GCA random effects, thus improving the ability to make correct selections of future parent candidates in a breeding program. These factors will impact the expected genetic gain for a given trait in the long term, which can be quantified by the breeder's equation (Jonas & de Koning, 2015; White et al., 2007).

In the same way, a correct estimation of the SCA variance is essential not to confound this effect in the estimation of genetic gain. The accurate estimation of those parameters depends mainly on an optimum mating design. A large number of parents will allow a better estimation of the true GCA variance of a population, and a large number of crosses per parent will allow a better estimation of the SCA variance (Bos & Caligari, 2008; White et al., 2007). The same principles apply in a breeding program with hybrid trees, where in that context, the combining ability of a parent is called General Hybridizing Ability (GHA), which is the genetic worth of a hybrid parent based on the average performance of its progeny, and the Specific Hybridizing Ability (SHA), which is the mean deviation of hybrid progeny relative to the sum of the GHA values of their parents (Volker et al., 2008). Ideally, breeders would like to have large mating designs, but in reality, breeders face limitations in terms of infrastructure, resources, and time.

In a scenario where a low rate of successful mating is likely in the generation of hybrid progeny, it is important for a breeder to ask, how many parents should be tested to assure unbiased and precise prediction? in how many crosses? To answer these questions, which operationally requires a considerable effort of resources and time (in years scale), a stochastic simulation can be done to optimize breeding and selection strategy to maximize genetic gain over one or more breeding cycles. Computer simulations are often used to evaluate the accuracy of predictions, effectiveness, the response of selection, among

others, testing different real-world scenarios (Céron-Rojas & Crossa, 2018). With complex and unbalanced experimental designs, simulation results will often provide a more reliable guide to optimal designs than will asymptotic variances (Shen et al., 1996).

In this study, we simulated several mating designs to produce hybrid progeny under different scenarios with varying numbers of parents and number of crosses per parent, as well as distinct variances of GHA and SHA effects. The scenarios in the simulation were chosen based on real-life examples observed in hybrid breeding programs. They were analyzed under a high number of iterations to determine how the estimates of genetic parameters behave across a range of GHA and SHA variances and distinct mating designs with low to high number of parents and crosses. The goal was to understand what level of confidence a breeder can have using parameter estimates to make hybrid breeding strategy decisions. Through comparison of simulated data in different scenarios, we evaluated the expected genetic parameters across several numbers of parents and crosses, with the main objective to understanding which are the optimum design parameters to be recommended for obtaining unbiased and precise estimates of GHA and SHA in the current and future hybrid breeding programs.

2.2 Material and Methods

2.2.1 Overview of Simulation

A stochastic simulation of hybrid progeny following Mendelian laws was written in R software (Team R. Core., 2019), based on an algorithm for simulation of pure species full-sib progeny, described by Lstibůrek, Hodge, & Lachout (2015). Several scenarios were simulated, assuming normal distribution and zero mean of the phenotypes for volume growth, considering a number of mating designs and diverse initial additive variances from the pure species parents of the hybrids, expressed in a set of General Hybridizing Abilities (GHA) values for each parent species and another set of Specific Hybridizing Ability (SHA) values for the interspecific crosses. The effect of the environment was considered through

type-B genetic correlations, which is an important measure of Genotype × Environment interaction (GxE), ranging from 0 to 1, with values near zero indicating weak agreement of genetic performance across environments, and therefore high levels of GxE (Gezan et al., 2017). The Type-B genetic correlation (rB_g) is a ratio between the genetic variance (σ_g^2) and the sum of the genetic variance σ_g^2 and the genetic x environment variance (σ_{gxe}^2), thus the general formula can be written as $rB_g = \sigma_g^2/(\sigma_g^2 + \sigma_{gxe}^2)$.

In pure species breeding, the narrow-sense heritability (h^2) is often estimated as 4-times the General Combining Ability (GCA) variance divided by phenotypic variance. Similarly, dominance variance can be quantified as d^2, equal to 4 times Specific Combining Ability (SCA) variance divided by phenotypic variance. In a hybrid context, the GHA variance (σ_{GHA}^2) is analogous to pure species GCA variance, and the SHA variance (σ_{SHA}^2) is analogous to pure species SCA variance (Mitchell et al., 2013; van den Berg et al., 2015; Volker et al., 2008).

The initial genetic parameters assumed in this study were based on typical values observed for standardized 8-year-old volume in tropical forest tree species (Hodge & Dvorak, 2012). These values were used to establish the size of the *true* total phenotypic variance and the GHA of both parents, along with their interaction with the environment (GxE) for those parameters. Studies of hybrid populations of *Eucalyptus* and *Pinus* species in growth traits, where estimates of additive and dominance variances were provided (Dieters et al., 1997; van den Berg et al., 2015), were used to establish the size of the true SHA variances and associated GxE for the simulation. Based on the average of total phenotypic variance (σ_{ph}^2) for standardized volume in tropical forest tree species, the size of the true phenotypic variance was set to σ_{ph}^2= 2500. The GHA variances for the pure species parents were set to σ_{GHA}^2 = 125 for Species A parents (which would correspond to a h_A^2 = 0.20 = [4*125/2500] in pure species), and for species B parents, two different GHA variances were assumed, σ_{GHA}^2 = 125 and 250 (corresponding to h_B^2 = 0.10 and 0.30 in pure species). Similarly, two SHA variances were assumed, σ_{SHA}^2 = 125 and 250 (corresponding to d^2 = 0.20 and 0.40 in hybrid populations estimates).

Every scenario considered full-sib partial diallel designs with 12, 24, or 48 parents (Npar) participating from 2 to 8 crosses (Ncross). There were 12 total combinations of input parameters called "Scenarios," tested with 1000 repetitions each (iterations within Scenarios). To indicate which genetic parameter came from Species A or B, a subscript followed by the letters A or B was incorporated into each parameter notation. For instance, the GHA value of Species A was denoted GHA_A, and so on with the other parameters. The number of offspring, number of sites, and the type B genetic correlation (rB_g) were fixed across scenarios. The input parameters can be seen in the following Table 2.1.

Table 2.1. Input parameters for the simulated data generation of each scenario. Species A, with fixed parameters across scenarios, and Species B with a range of combinations of $\sigma^2_{GHA_B}$, number of parents (Npar), and number of crosses (Ncross). The letter "E" represents the environment effect in interaction with GHA and SHA variances. Two levels of SHA variance (σ^2_{SHA}) were tested across scenarios.

Scenario	$\sigma^2_{GHA_B}$	σ^2_{SHA}	$\sigma^2_{GHA_{BxE}}$	σ^2_{SHAxE}	σ^2_{err}	$Npar_B$	$Ncross_B$
1	62.5	125	15.6	31.3	2109.4	12	8
2	62.5	125	15.6	31.3	2109.4	24	4
3	62.5	125	15.6	31.3	2109.4	48	2
4	62.5	250	15.6	62.5	1953.1	12	8
5	62.5	250	15.6	62.5	1953.1	24	4
6	62.5	250	15.6	62.5	1953.1	48	2
7	187.5	125	46.9	31.3	1953.1	12	8
8	187.5	125	46.9	31.3	1953.1	24	4
9	187.5	125	46.9	31.3	1953.1	48	2
10	187.5	250	46.9	62.5	1796.9	12	8
11	187.5	250	46.9	62.5	1796.9	24	4
12	187.5	250	46.9	62.5	1796.9	48	2
Fixed Parameters: $\sigma^2_{GHA_A}$= 125; $\sigma^2_{GHA_{AxE}}$= 31.3; σ^2_{ph}= 2500; rB_g= 0.8; $Npar_A$= 24; $Ncross_A$= 4; N siblings= 200							

In summary, Species A parameters were set up with intermediate values of GHA variance ($\sigma^2_{GHA_A}$= 125), with 24 parents tested in an average of 4 crosses each, and Species B parameters were established with a low and a high GHA variance ($\sigma^2_{GHA_B}$= 62.5 and 187.5), and a range of number of parents and number of crosses, as can be seen in Table 2.1 above. The phenotypic variance was fixed with a value of σ^2_{ph}= 2500, rB_g with a value of 0.8 for GHA and SHA by environment interaction, respectively. The number of sites simulated was 4, with a total of 200 siblings per family across the 4 test sites.

2.2.2 Generation of Hybrid Population Data

For each scenario, the simulated dataset generated was composed of a full-sib hybrid population with equal family size for all combinations (Table 1, N siblings). The total number of offspring genotypes was 19200, distributed in 4 field tests, each with 4800 trees. The number of parents and crosses vary across the scenarios (Table 2.1), with the crosses formed from 3 different incomplete diallel designs, according to the number of parents and crosses (Appendix D).

Each trial was a randomized complete block design (RCBD), with single tree plots (STP), with 24 replications each. The data simulation assumed 100% survival in all tests.

The hybrid crosses of parents from Species A and B were randomly assigned, assuming no relatedness among the pure species parents. Additive and dominance effects were sampled from the normal distribution, assuming the infinitesimal genetic model (Barton et al., 2017; M. Hamilton, 2009; Kerr, Dieters, Tier, et al., 2004; Mrode, 2014).

The mendelian sampling effect was estimated according to previous derivations (Hoeschele & VanRaden, 1991; Mulder et al., 2007; Norris et al., 2009; Walsh & Lynch, 2018), is estimated as follow:

$$a = 0.5\left(a_{Spp_A} + a_{Spp_B}\right) + m_a \qquad \text{Eq. 2.1}$$

$$d = f_{Spp_A,Spp_B} + m_d \qquad \text{Eq. 2.2}$$

Where a_{Spp_A} and a_{Spp_B} were the genetic additive effects of parents from Species A female and Species B male parent respectively, f_{Spp_A,Spp_B} was the combination of genes of female and male parents from Species A and B, m_a and m_d represents the mendelian sample of additive and dominance effects separately. Under infinitesimal model, m_a is independent of parental breeding values and has a mean zero (Walsh & Lynch, 2018). In the same way m_d has an expected value of zero (Hoeschele & VanRaden, 1991).

2.2.3 Model and Variance Components

The variance components of each simulated population were estimated, using the restricted maximum-likelihood approach (REML) conducted in R software with ASReml-R package (Butler et al., 2017), by fitting the following model:

$$y_{ijkl} = \mu + E_i + GHA_{A_j} + GHA_{B_k} + SHA_{jk} + GHA_A xE_{ij} + GHA_B xE_{ik} + SHAxE_{ijk} + err_{ijkl}$$

Eq. 2.3

Were *y* is the *l*th observation of the *jk*th hybrid family at the *i*th Environment; u is the overall population mean; E_i is the fixed effect of i^{th} Environment; GHA_{A_j} or GHA_{B_k} is the random GHA effect for j^{th} Species A female or k^{th} Species B male parent; SHA_{jk} is the random SHA effect of the j^{th} and k^{th} parents; $GHA_A xE_{ij}$ or $GHA_B xE_{ik}$ is the random GHA by Environment interaction of i^{th} environment and j^{th} Species A female or k^{th} Species B male parent; $SHAxE_{ijk}$ is the random SHA by Environment interaction of i^{th} environment and j^{th} and k^{th} parents and err_{ijkl} is the random error. All the effects on this model, except for the overall mean and site effect, were assumed to be random and independently distributed.

2.2.4 Accuracy of Predictions Across Scenarios

A comparison between the true variances, sample variances, and estimated variances was made across scenarios to examine the precision and accuracy of the estimates.

In this simulation, the values of the true variances (σ^2) where the initial parameters of each scenario (Table 2.1), sample variances ($\hat{\sigma}^2$) were the variances of the true random effects generated in a given iteration of the simulation, and estimated variances ($\hat{\hat{\sigma}}^2$) were the variances obtained from the REML solutions of the simulated data under the model presented in Eq. 2.3. For each iteration within scenarios, σ^2 and $\hat{\sigma}^2$ for each random effect were recorded, and $\hat{\hat{\sigma}}^2$ values were estimated.

A *statistical* definition of *"bias"* is any trend away from or deviation from the true value in a data collection (Šimundić, 2013). This data collection represents a sample obtained from a

population, with “population” indicating the complete universe of data of the phenomena studied, and it is characterized by the true distribution of the random variable (Weisberg, 2010). Considering that studying an entire population is not feasible, due to limitations in budget and time, researchers study the phenomena of interest by representative samples of a population (Šimundić, 2013). In this way, estimates of the parameters of a population arise from these samples. An estimate of some parameter, such as the mean of a given population, is biased if the estimate “on average” deviates from the true value of the parameter (Weisberg, 2010). This definition implies that bias is a number that could be positive or negative (Sutherland et al., 1982).

In genetic breeding, a goal can be to change the population mean of a particular trait. In this case, the breeder will be interested in a random variable expressing the genetic value of particular individuals in the population. Using, for example, a forest population of trees, we can say that u is the true breeding value (TBV) and $\hat{u}$ is the estimated breeding value (EBV) for a specific trait of a single tree (Legarra & Reverter, 2018). With these concepts, we can define the “*accuracy”* of the EBV estimates as a Pearson correlation between EBV and TBV (Singh & Singh, 2015), calculated as $r(u,\hat{u}) = cov(u,\hat{u})/\sqrt{var(\hat{u})var(u)}$ across a series of individuals (Legarra & Reverter, 2018). Consequently, bias is the difference between the averages of EBV and TBV estimated by $b_0 = \bar{\hat{u}} - \bar{u}$ and dispersion of the slope of the regression of TBV on EBV, being $b_1 = cov(u,\hat{u})/var(\hat{u})$. The same concept was applied on estimated GHA and SHA variance components, where in the simulation the variance estimates from each of the one thousand iterations are analogous to the breeding value predictions. In this case, true variances values are underestimated when $b_0 < 0$, and are overestimated when $b_0 > 0$. In the same way, values of $b_1 < 1$ represent an overestimation of predicted variances (Legarra & Reverter, 2018; Macedo et al., 2020).

In concordance with the Central Limit Theorem (CLT) in a stochastic simulation, each iteration is a sample of the population, with N = 1000 samples for each Scenario, with the sample variances ($\hat{\sigma}^2$), REML estimated variances ($\hat{\hat{\sigma}}^2$) and the means of both of those all random variables on their own (Martinson & Core, 2018). Under these circumstances, the mean of the iteration variances (i.e $\mu_{\hat{\sigma}^2_{GHA}}$ or $\mu_{\hat{\hat{\sigma}}^2_{SHA}}$) will be the estimates of true GHA and

SHA variances, and the standard deviations of the iteration variances (i.e $\sigma_{\hat{\hat{\sigma}}^2_{GHA}}$ or $\sigma_{\hat{\sigma}^2_{SHA}}$) will be an indicator of the fluctuation of the iterations around the true GHA and SHA variances values, in other words, a parameter to describe the bulk interval of where we could find the true variances. By summarizing these parameters of $\hat{\sigma}^2$and $\hat{\hat{\sigma}}^2$ from each Scenario and being compared with the σ^2 (initial parameters of each Scenario) is how the simulation will be evaluated, in terms of whether the sample and REML estimates variances are precise and accurate estimates of the true value on each scenario.

To understand the frequency of how often breeders might have important errors in the estimation of the quantitative genetic parameters, a ratio of the $\hat{\hat{\sigma}}$ estimate, and the σ true value was calculated. The variances will be plotted in σ values ($\sigma = \sqrt{\sigma^2}$) to facilitate the comparisons between scenarios, making it easier to represent the different GHA and SHA values on the same scale and the same graphs. This ratio was called Q-value ($\hat{\hat{\sigma}}/\sigma$) in this manuscript, and was estimated for both species through scenarios, with $\mathrm{Q}_{\mathrm{GHA_A}}$ equal to the ratio of σ GHA REML estimate ($\hat{\hat{\sigma}}_{\mathrm{GHA_A}}$) over the true σ GHA of Species A ($\sigma_{\mathrm{GHA_A}}$), and similarly for species B. The ratio of $\hat{\hat{\sigma}}/\sigma$ for the SHA variation was denoted $\mathrm{Q}_{\mathrm{SHA}}$. Using the distribution of the predictions observed over the Q-values, one can examine how often the estimates of the genetic parameters across scenarios were severely under or overestimated. In the same way, it can be inspected how often estimates of zero variation were obtained in scenarios where the true variance was never zero.

Assuming that $\hat{\hat{\sigma}}$ is an unbiased estimator, Q-values are expected to be equal to 1. Q values greater than 1 indicate over-estimation of true values, and lower than 1 indicates an under-estimation of the GHA and SHA variances studied in this simulation. In other words, Q-values different from 1 would indicate the possible “bias” direction of the estimates.

2.3 Results

2.3.1 Mean Sample Variances and Mean REML Variances Estimates in Comparison with Target Variance.

Across all of the scenarios, the mean of the sample variances ($\hat{\sigma}^2$) and the mean of the REML variances estimates ($\hat{\hat{\sigma}}^2$) were very accurate, with the mean of these values being extremely close to the true target variances (with all means of variances within ± one SD of the 1K iterations) for all the parameters within all the scenarios.

For example, in Scenario 1, with a target value for GHA_A of $\sigma^2_{GHA_A}$= 125, the mean sample GHA_A variance across the 1K iterations was $\hat{\sigma}^2_{GHA_A}$= 126.4, with a standard deviation of $\sigma_{\hat{\sigma}^2_{GHA_A}}$= 36.3 for the independent 1K estimates. Similarly, the mean REML GHA_A variance was $\hat{\hat{\sigma}}^2_{GHA_A}$= 126.7 with a standard deviation of 50.9. Looking for the GHA_B estimates in the same scenario, with a target value of $\sigma^2_{GHA_B}$= 62.5, the mean sample GHA_B variance was $\hat{\sigma}^2_{GHA_B}$= 62.2 with a standard deviation of 25.5, and the mean REML GHA_B variance was $\hat{\hat{\sigma}}^2_{GHA_B}$= 61.4 with an SD of 36. Finally, for the SHA parameter, with a target value of σ^2_{SHA}= 125, the mean sample SHA variance was $\hat{\sigma}^2_{SHA}$ =125.1 ± 18.4, and the mean REML SHA variance was $\hat{\hat{\sigma}}^2_{SHA}$= 125.4 ± 26.2. The rest of the genetic parameters in Scenario 1 were accurately estimated as well, with mean sample variance ($\hat{\sigma}^2$) and mean REML variances ($\hat{\hat{\sigma}}^2$) of all the parameters extremely close to their respective target values. These results are representative of all Scenarios and can be examined in Appendix A.

The REML variances estimates were approximately normally distributed in almost all scenarios; the exceptions were scenarios with low GHA_B variance and high SHA variance, a topic that will be discussed later in the analysis of the distributions of the estimates summarized in Figure 2.3.

2.3.2 Relationship of Samples Variances and REML Variance Estimates Across Iterations.

In order to illustrate the relationship of the sample parameters and REML estimates across the 1K iterations in each Scenario, a set of scatterplots was made. For these figures, the sample standard deviation ($\hat{\sigma}$) and REML standard deviation estimate ($\hat{\hat{\sigma}}$) were plotted, rather than sample variance ($\hat{\sigma}^2$) and REML variance estimate ($\hat{\hat{\sigma}}^2$).

These plots represented the square root of the sample, and REML estimates variances across all iterations within Scenarios for GHA and SHA effects, and this transformation was done to facilitate plotting and the comparison of results from all Scenarios simultaneously on a single graph. Figure 2.1 presents the relationship for GHA_B (REML $\hat{\hat{\sigma}}_{GHA_B}$ vs $\hat{\sigma}_{GHA_B}$), and Figure 2.2 present the relationship for GHA_A (REML $\hat{\hat{\sigma}}_{GHA_A}$ vs $\hat{\sigma}_{GHA_A}$).

Throughout the different Scenarios, which were described previously, the Species B parents vary from $Npar_B$= 12, 24, and 48, with the average number of crosses per parent varying from NC_B= 8, 4, and 2, respectively, associated with the different mating design described in Appendix D.

The REML $\hat{\hat{\sigma}}_{GHA_B}$ were better estimates of the sample $\hat{\sigma}_{GHA_B}$ values as the number of crosses increased (from 2 to 8 in this simulation), no matter if the target GHA_B variance was low or high ($\sigma^2_{GHA_B}$ value of 62.5 or 187.5). This can be observed in Figure 2.1 by comparing the change of the magnitude of the correlation ("r") between these parameters across scenarios that hold the other parameters constant.

For example, in Figure 2.1, by contrasting the correlation of Scenarios 1 and 3, which have the same GHA_B and SHA variance, the correlation of $\hat{\hat{\sigma}}_{GHA_B}$ and $\hat{\sigma}_{GHA_B}$ was r= 0.69 with 8 crosses per parent, and r= 0.33 with only 2 crosses per parent. For Scenario 2, with 4 crosses per parents, the correlation was r= 0.52.

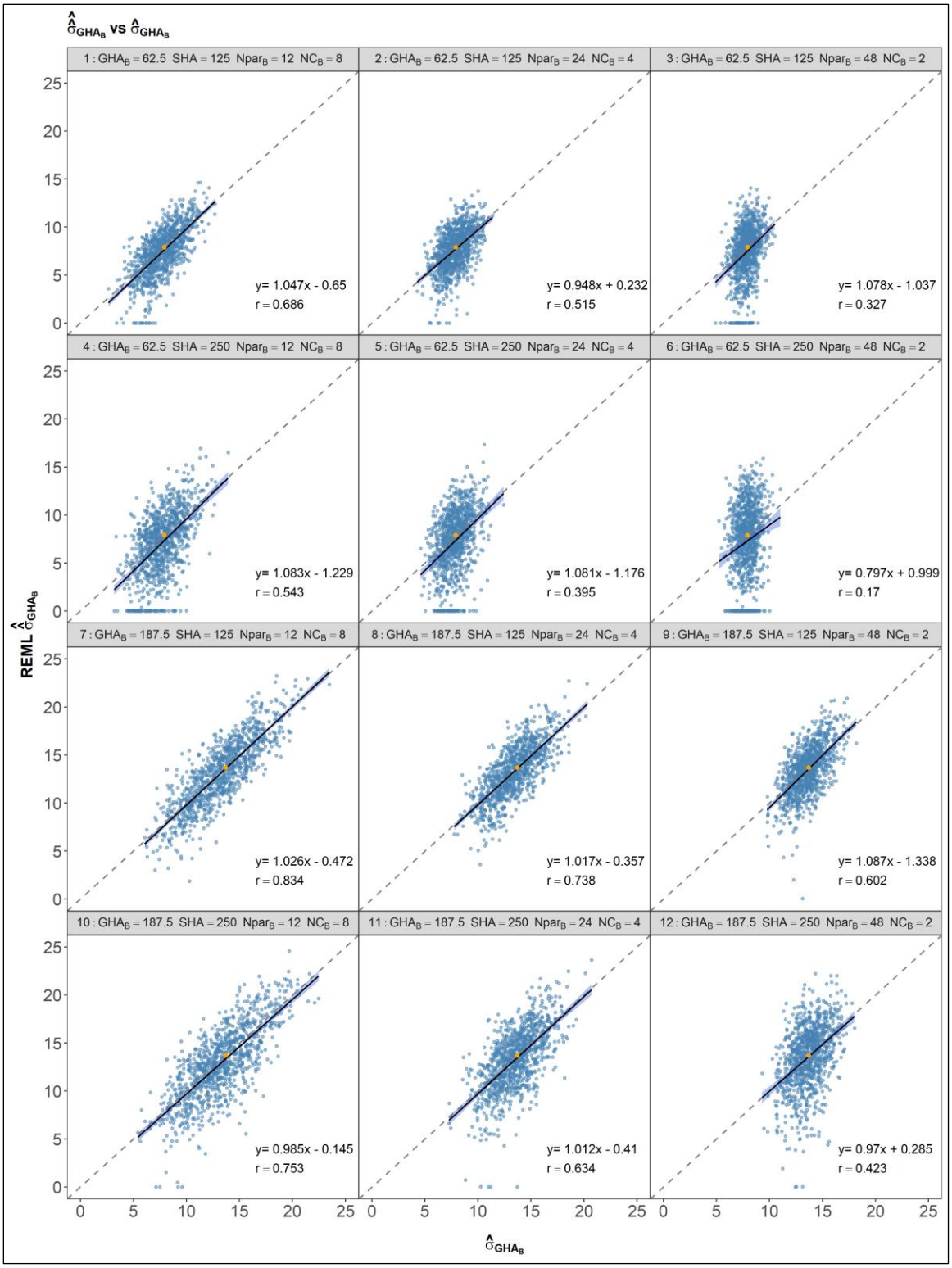


Figure 2.1. Scatterplots of predicted REML $\hat{\hat{\sigma}}_{GHA_B}$ vs $\hat{\sigma}_{GHA_B}$ (σ_{GHA} from species B), where orange dot represent the true σ_{GHA_B} value of each scenario. Blue dots represent the result from 1000 simulations within the 12 scenarios. The segmented line is the reference line of equivalence in each scenario. The GHA_A parameters where constant across scenarios, with a GHA_A= 125, $Npar_A$= 24, NC_A= 4. Npar and NC were abbreviations of Number of Parents and Number of Crosses for each species, respectively.

Similar patterns can be seen for other sets of scenarios with the same parameters but a different number of crosses. For example, in Scenarios 7, 8, and 9, where the correlation of $\hat{\hat{\sigma}}_{GHA_B}$ and $\hat{\sigma}_{GHA_B}$ was r= 0.83, 0.74, and 0.60 with 8, 4, and 2 crosses per parent, respectively. Similar patterns were also observed in Scenarios 4 to 6 and Scenarios 10 to 12, with a higher level of SHA variance (σ^2_{SHA}= 250 versus 125), where increasing the number of crosses per Species B parent substantially increased the correlation of $\hat{\hat{\sigma}}_{GHA_B}$ and $\hat{\sigma}_{GHA_B}$ in every case.

Among the different scenarios, the number of Species A parents and the number of crosses per parent was held constant at $Npar_A$ = 24 and NC_A = 4, and it can be observed that varying the number of parents, crosses, or GHA variance in Species B does not have much impact on the correlation of $\hat{\hat{\sigma}}_{GHA_A}$ and $\hat{\sigma}_{GHA_A}$, which is more and less constant across Scenarios, as can be seen in Figure 2.2. However, the same general trend as was discussed before can be observed in GHA_A relationship when Scenarios with a low number of crosses were compared with Scenarios with a high number of crosses of Species B. For example, the correlation of $\hat{\hat{\sigma}}_{GHA_A}$ and $\hat{\sigma}_{GHA_A}$ of Scenarios 3, 6, 9, and 12 with $Npar_B$= 48 and NC_B= 2 was always lower than in Scenarios 1, 4, 7, 10 with $Npar_B$= 12 and NC_B= 8 (Figure 2.2). Nevertheless, the decrease in the magnitude of the correlations was generally small; for example, the largest decrease was found comparing Scenarios 7 and 9, where the correlation of REML $\hat{\hat{\sigma}}_{GHA_A}$ and $\hat{\sigma}_{GHA_A}$ went from r = 0.68 to 0.55, respectively. The smallest decrease was found comparing Scenarios 4 and 5 versus Scenario 6, where the correlation went from r= 0.54 and 0.57 to r = 0.54, almost no decrease at all.

Another interesting observation was that the variation of both the sample $\hat{\sigma}$ parameter and the REML $\hat{\hat{\sigma}}$ estimates around the true target parameter were lower in the scenarios with low GHA_B variance (Scenarios 1 to 6, with $\sigma^2_{GHA_B}$= 62.5), and higher in scenarios with high GHA_B variance (Scenarios 6 to 12 with $\sigma^2_{GHA_B}$= 187.5) (Figure 2.1). This pattern was also present for GHA_A relationships, where the variation around the true value of $\hat{\sigma}$ and $\hat{\hat{\sigma}}$ was slightly larger in Scenarios 6 to 12 than in Scenarios 1 to 6 (Figure 2.2).

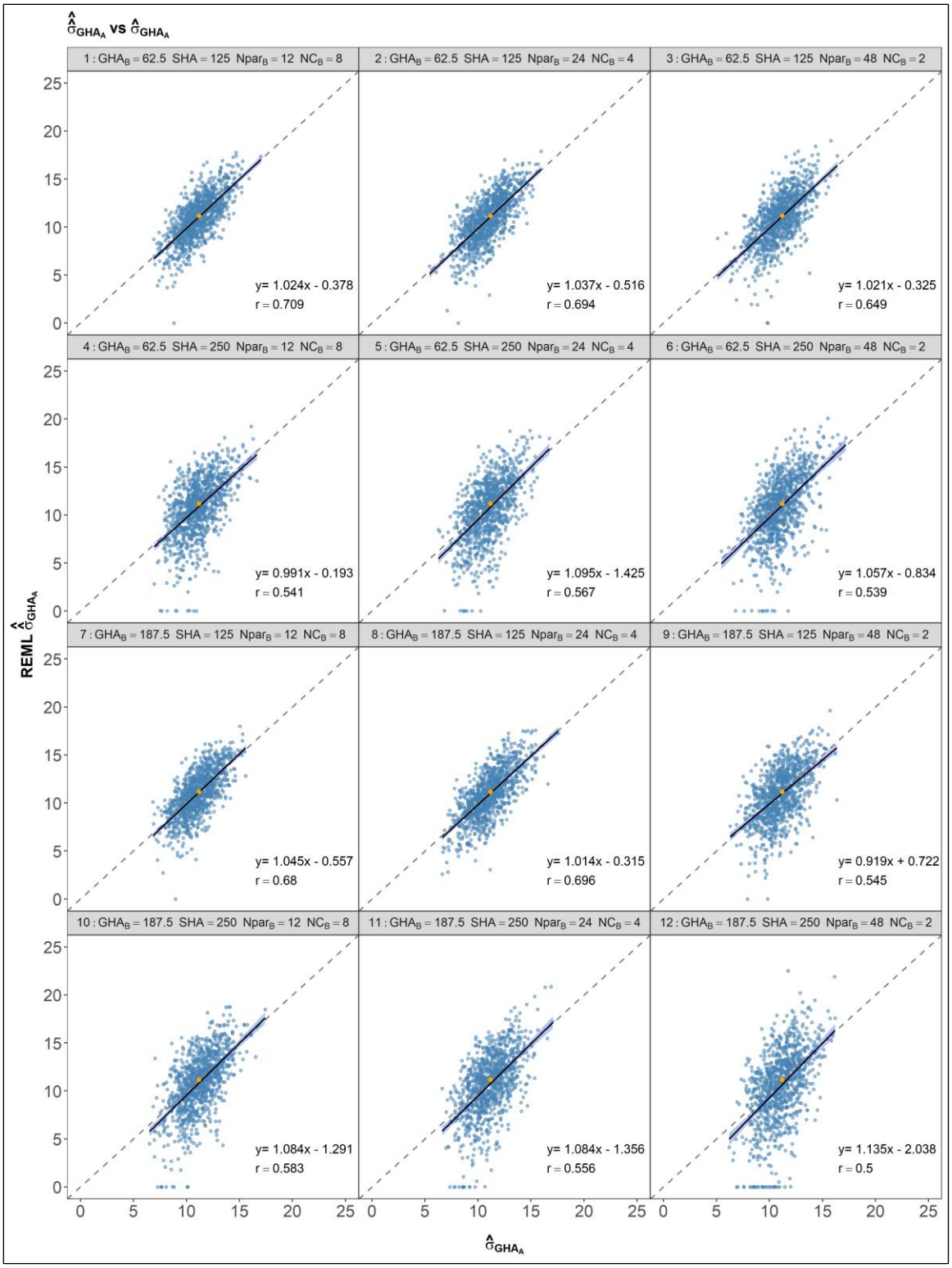


Figure 2.2. Scatterplots of predicted REML $\hat{\hat{\sigma}}_{GHA_A}$ vs $\hat{\sigma}_{GHA_A}$ (σ_{GHA} from species A), where orange dot represent the true σ_{GHA} value of each scenario. Blue dots represent the result from 1000 simulations of 12 scenarios. The segmented line is the reference line of equivalence in each scenario. GHA_B parameters vary across scenarios, being showed in every plot title. Npar and NC were abbreviations of Number of Parents and Number of Crosses for each species, respectively.

Finally, there was an important effect of $Npar_B$ and NC_B on the variation of both the sample $\hat{\sigma}$ parameter and the REML $\hat{\hat{\sigma}}$ estimates around the true target parameter. As $Npar_B$ increases from 12 to 24 to 48, there was a notable reduction of the variation of the sample $\hat{\sigma}$ (Figure 2.1, compare Scenarios 1 to 2 to 3, 4 to 5 to 6, etc.).

This was not unexpected; a sample of 48 parents should give a more precise estimate of $\sigma^2_{GHA_B}$ than a sample of 12 parents. However, the reduction of the variation of the REML $\hat{\hat{\sigma}}$ estimates as the number of B parents increases were much lower than sample $\hat{\sigma}$ parameter. That was slightly difficult to discern visually in Figure 2.1, but still can be seen by comparing Scenarios 1 to 2 to 3, where the variation among sample $\hat{\sigma}$ estimates clearly diminishes, while the variation among REML $\hat{\hat{\sigma}}$ estimates appears relatively constant. This pattern was also observed through comparing scenarios 4 to 5 to 6, where again the sample $\hat{\sigma}$ showed a lower dispersion than REML $\hat{\hat{\sigma}}$. In the Scenarios with higher GHA_B (Scenarios 7 to 8), this pattern was more challenging to see, but it is still there.

The patterns described in the Figure 2.1 inspection can also be seen by checking the mean values over 1000 iterations presented in Appendix A. For example, comparing Scenarios 7 to 8 to 9 with $Npar_B$= 12, 24, 48, respectively, the standard deviation of the 1K sample $\hat{\sigma}^2_{GHA_B}$ estimates were 78, 54, and 38. In comparison, the standard deviation of the 1K REML $\hat{\hat{\sigma}}^2_{GHA_B}$ estimates were 94, 73, and 67, a smaller decrease in variation as the number of parents increased.

The effect of $Npar_B$ and NC_B on the variation of the REML $\hat{\hat{\sigma}}$ estimates for GHA_A appear to be relatively small. There was almost no increase in the visible dispersion among $\hat{\hat{\sigma}}_{GHA_A}$ estimates, moving from Scenarios 1 to 2 to 3 or from Scenarios 4 to 5 to 6, representing scenarios with low GHA_B variance. There was a slight increase among REML $\hat{\hat{\sigma}}_{GHA_A}$ estimates moving from Scenarios 7 to 8 to 9, and from Scenarios 10 to 11 to 12, which represent scenarios with high GHA_B variance. But still, the impact was small, and this was confirmed as well by inspection of the means and standard deviations among $\hat{\hat{\sigma}}^2_{GHA_A}$ estimates in Appendix A, where the standard deviation of the 1K REML $\hat{\hat{\sigma}}^2_{GHA_A}$ estimates were

51, 54, and 59, a very small impact in the variation of the GHA_A estimates by increasing the number of parents.

2.3.3 Frequency of Severe Errors in GHA Variance Parameter Estimates: Histograms of Q = $\hat{\hat{\sigma}}/\sigma$

As with the scatterplots in Figure 2.1 and Figure 2.2, histograms to examine the frequency of severe errors in parameter estimates are presented, comparing true standard deviation (σ) and REML standard deviation estimates ($\hat{\hat{\sigma}}$), calculated from the true variance (σ^2) and REML variance estimates ($\hat{\hat{\sigma}}^2$). Figures 3 and 4 show histograms of Q-values for the REML $\hat{\hat{\sigma}}$ estimates for GHA_B and GHA_A, respectively. Since Q = $\hat{\hat{\sigma}}/\sigma$, where σ is the true parameter, Q-values near 1 represent good estimates of the true genetic variance. Very high Q-values represent cases where the estimated genetic variation was "too high," and very low Q values represent cases where the estimated genetic variation was "too low." Either of these situations could mislead the breeder attempting to develop mating and testing strategies. In the histograms, Q-values were aggregated into bins of 0, 0.25, 0.50, ..., up to 2.25. In this study, Q-values of 0.50 and 1.50 were classified as *errors* (50% too high or too low) and are represented as orange bars. Q-values in the 1.75 class and above or in the 0.25 class and below were classified as *severe errors* and are represented as red bars. Q-values in the 1.00 class represent *precise* estimates of the true value. Q-values in the 0.75 and 1.25 classes represent *reasonably precise* estimates of this value; all of those are represented in blue bars.

By inspecting the distribution plots in Figure 2.3, it can be seen that the size of the GHA_B variance had an important impact on the distribution of the predictions. Comparing Scenarios 1 to 6 (with true variance $\sigma^2_{GHA_B}$= 62.5) to Scenarios 7 to 12 (with true $\sigma^2_{GHA_B}$= 187.5), it is clear that the distribution of Q-values is much broader in the scenarios with lower $\sigma^2_{GHA_B}$. The scenarios with low $\sigma^2_{GHA_B}$ had a much higher proportion of errors and severe errors in REML $\hat{\hat{\sigma}}^2$ estimates for GHA_B ($\hat{\hat{\sigma}}^2_{GHA_B}$), which can be observed graphically in the frequency of orange and red bars in Figure 2.3.

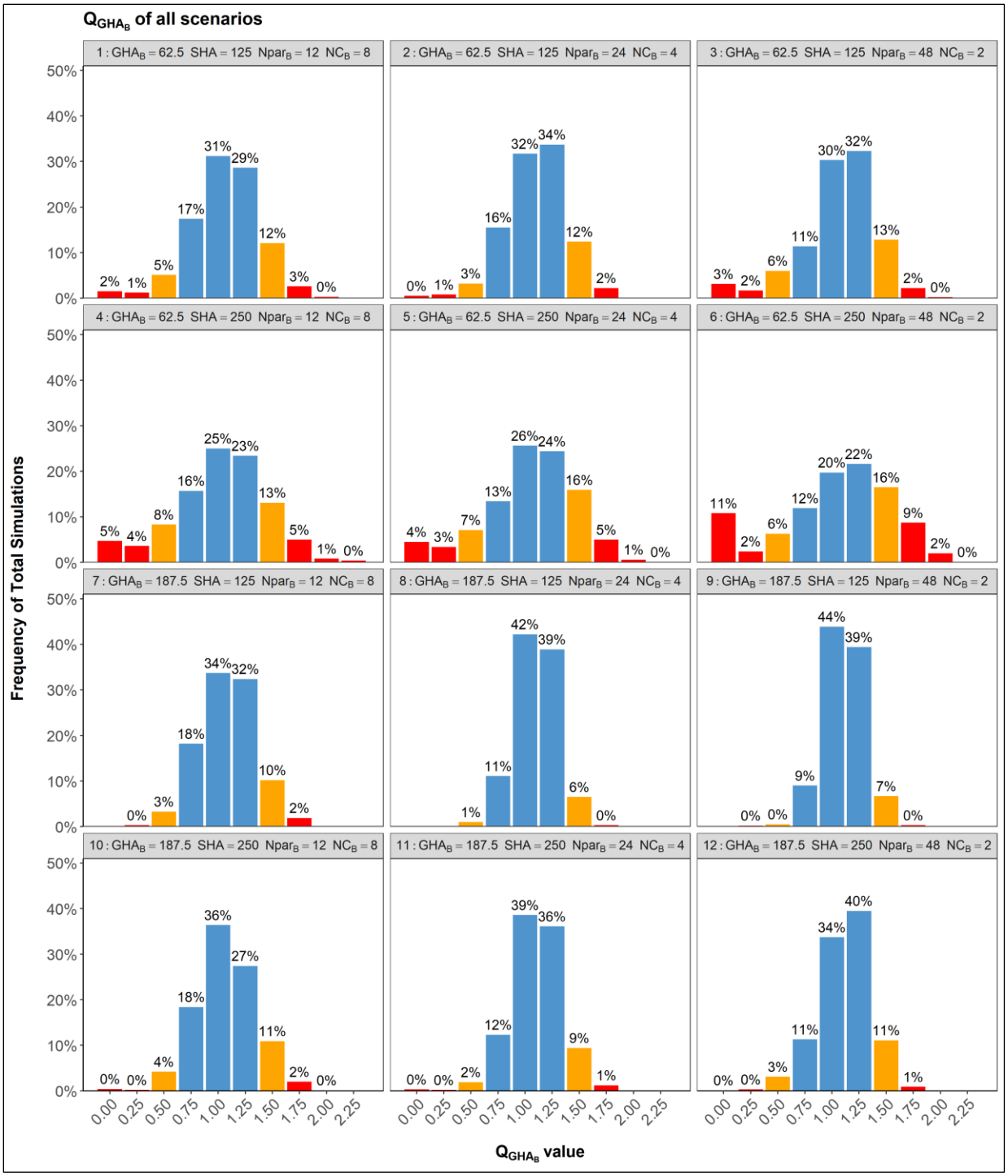


Figure 2.3. Histogram of Q values for GHA_B (GHA from species B). $Q = \hat{\hat{\sigma}}/\sigma$, where $\hat{\hat{\sigma}}$ = REML estimate of σ_{GHA_B}, and σ = the true parameter value for σ_{GHA_B}. Histograms show results from 1000 simulations of 12 scenarios with different genetic parameters and mating designs. Red bars represent a high deviation of true GHA values, and orange bars represent a moderate deviation of true GHA values.

In the scenarios with high $\hat{\hat{\sigma}}^2_{GHA_B}$, there was a very low frequency of errors and severe errors. For example, in Scenario 10, with true $\sigma^2_{GHA_B}$= 187.5, in 11% of the iterations the $\hat{\hat{\sigma}}_{GHA_B}$ was overestimated and 2% of the times severely overestimated, which represent a total of 13% of overestimated values, and in only around 4% of the iterations was $\hat{\hat{\sigma}}_{GHA_B}$ underestimated. In Scenario 4, with exactly the same parameters as Scenario 10, except for the size of true GHA_B variance, with a value of $\sigma^2_{GHA_B}$= 62.5, there was a much higher frequency of both errors and severe errors with 19% of overestimates and 17% of underestimate values across iterations, a considerable difference in the precision of the estimates.

Comparing Scenarios 4 and 5 with Scenario 6, there is a clear increase in the frequency of errors and severe errors in the estimates of $\hat{\hat{\sigma}}_{GHA_B}$, with $Npar_B$ = 48 and NC_B = 2. For example, Scenario 4 had 19% overestimates and 17% underestimates, but Scenario 6 had 27% overestimates and 19% underestimates. Scenario 6 was also the worst case for the frequency of $\hat{\hat{\sigma}}_{GHA_B}$= 0, with 11% of the estimates falling into this class.

Through all scenarios with low $\sigma^2_{GHA_B}$ (Scenarios 1 to 6), on average, the GHA_B variance was in the range of precise to reasonably precise in 69% (blue bars on Figure 2.3), overestimated in 19%, and underestimated in 12% of the iterations, with higher proportions of both overestimate and underestimate errors in the presence of high SHA variance. More importantly, across Scenarios 1 to 6, there were from 2% to 11% of the iterations where estimates of GHA_B variance were $\hat{\hat{\sigma}}^2_{GHA_B}$= 0. This would be a critical error, and this kind of result, in genetic analysis of hybrid progeny, could seriously mislead a tree breeder. A zero GHA variance estimate, if accepted as true, would be interpreted to mean that one of the parental species has no genetic influence on a given trait, suggesting that this factor could be completely ignored in developing a breeding strategy for a hybrid variety.

On the other hand, in the scenarios with high $\sigma^2_{GHA_B}$ (Scenarios 7 to 12), there was only 1% of the iterations with severe underestimates, or severe overestimates, and there was a near-zero chance of estimating $\hat{\hat{\sigma}}^2_{GHA_B}$= 0.

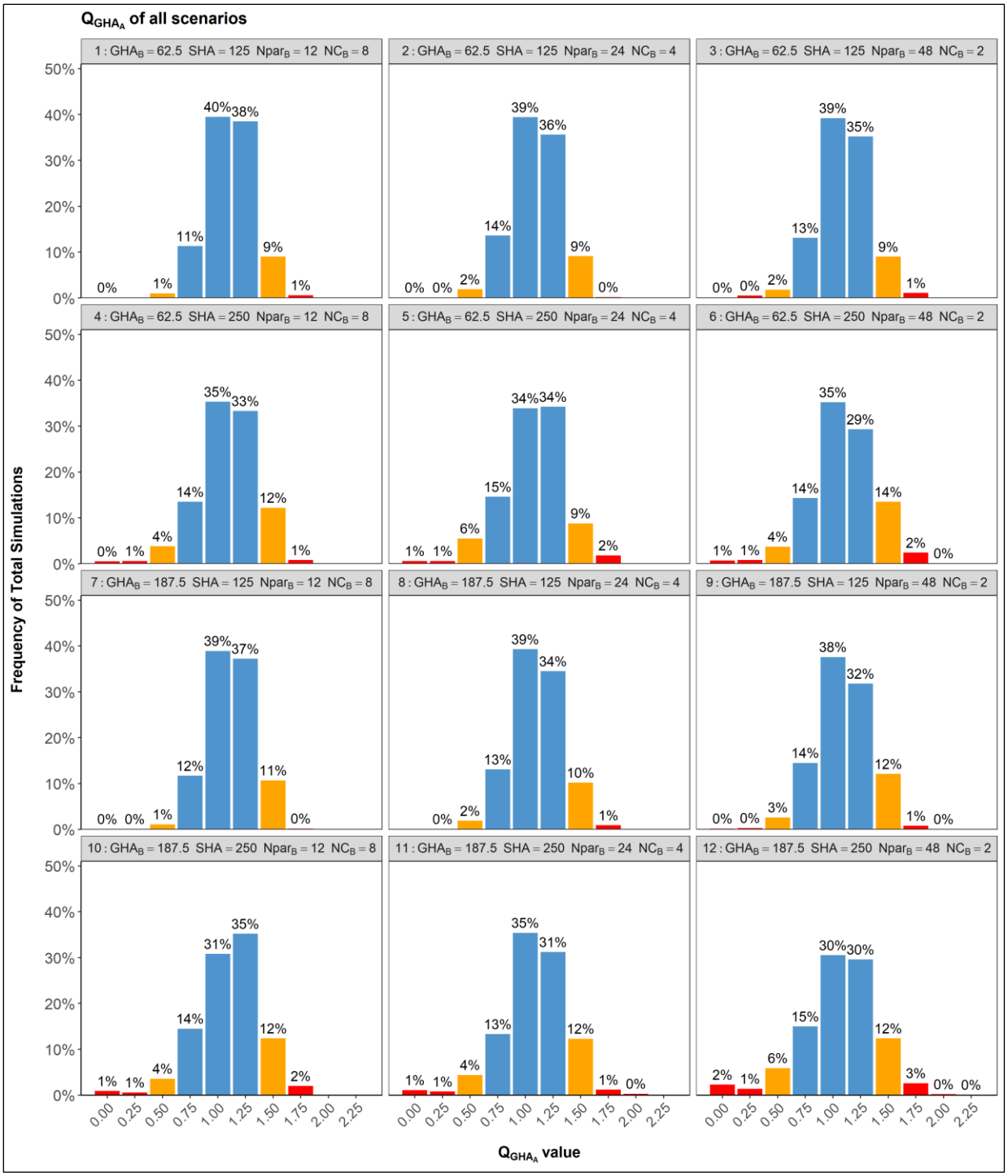


Figure 2.4. Histogram of Q values for GHA_A (GHA from species A. $Q = \hat{\hat{\sigma}}/\sigma$, where $\hat{\hat{\sigma}}$ = REML estimate of σ_{GHA_A}, and σ = the true parameter value for σ_{GHA_A}. Histograms show results from 1000 simulations of 12 scenarios with different genetic parameters and mating designs of species B. Red bars represent a high deviation of true GHA values, and orange bars represent a moderate deviation of true GHA values.

In approximately 87% of the iterations (ranging from 82 to 92%), the $\hat{\hat{\sigma}}^2_{GHA_B}$ was a precise or reasonably precise estimate of the true variance $\sigma^2_{GHA_B}$ (the blue bars, Figure 2.3).

Figure 2.4 presents Q-plots for the $\hat{\hat{\sigma}}_{GHA_A}$ estimates, with all scenarios having constant $Npar_A$ = 24 and NC_A= 4, with a true target variance of $\hat{\hat{\sigma}}^2_{GHA_A}$= 125. As seen with the REML σ estimates for GHA_B ($\hat{\hat{\sigma}}_{GHA_B}$), there appears to be a slight increase in the frequency of overestimates and underestimates of $\hat{\hat{\sigma}}_{GHA_A}$ in scenarios with $Npar_B$ = 48 and NC_B = 2 (Scenarios 3, 6, 9, and 12), but in general, the distributions of the estimates across all 12 scenarios were very similar. More importantly, the distribution of estimates in Figure 2.4 (with true $\sigma^2_{GHA_A}$= 125) appear intermediate between the distributions of estimates for Scenarios 1 to 6 in Figure 3 (with true $\sigma^2_{GHA_B}$= 62.5) and Scenarios 7 to 12 in Figure 2.3 (with true $\sigma^2_{GHA_B}$= 187.5). This suggests that the primary factor affecting the precision of variance component estimates is the size of the true variance, but the number of parents and number of crosses per parent does have some impact, particularly if GHA variance is small.

In the scenarios with high SHA variance (σ^2_{SHA}= 250, Scenarios 7 to 12), there were very few severe overestimates or underestimates of $\hat{\hat{\sigma}}_{GHA_A}$ (depicted in red bars, Figure 2.4), with frequencies ranging from 1 to 3%, and from 0% to 2% estimates of $\hat{\hat{\sigma}}^2_{GHA_A}$= 0. Overall, 84% of iterations gave estimates of $\hat{\hat{\sigma}}_{GHA_A}$ that were precise or reasonably precise, and this value increased to 85% in Scenarios 1 to 6 with low SHA variance, a more stable behavior than the GHA_B estimates.

2.3.4 SHA Variance Estimates

The SHA variance (σ^2_{SHA}) was the most accurately estimated parameter across all scenarios investigated in this study and were almost completely unaffected by different levels of σ^2_{SHA}, $\sigma^2_{GHA_A}$, $\sigma^2_{GHA_B}$, or numbers of parents. Scatterplots of $\hat{\hat{\sigma}}_{SHA}$ and $\hat{\sigma}_{SHA}$ (Appendix B) showed much narrower distributions than corresponding plots for GHA variances, and Q-plots of $\hat{\hat{\sigma}}/\sigma$ for SHA variance showed much lower frequencies of overestimate or underestimate errors (Appendix C). On average, across all scenarios, 99% of the $\hat{\hat{\sigma}}_{SHA}$ estimates were near

to the true value (i.e., precise or reasonably precise). This is most likely due to the fact that in all scenarios, there were always 96 family combinations (see diallel in Appendix D), which allowed a very good prediction of this parameter. As mentioned above, the size of the SHA variance did affect the precision of the GHA variance estimates and the frequency of errors in estimation, but this had very little impact on the estimation of the σ^2_{SHA} itself.

2.4 Discussion

2.4.1 Genetic Parameters Assumptions

The genetic parameters in this simulation were all based on growth traits typical of forest trees. Volume growth is one of the most important traits to breed for in almost all pure species and hybrid varieties across the world, with high economic returns in timber volume (m^3/land unit) obtained at the moment of harvest.

The genetic parameter values were set based on examples of growth traits observed in tropical pure species of *Pinus* and hybrid populations in the genera of *Eucalyptus* and *Pinus*. In these populations, narrow-sense heritabilities on the order of h^2 = 0.10 to 0.30 are commonly observed and were used as a basis for the size of the parental GHA variance. In this simulation, GHA variances ranged from σ^2_{GHA} = 62.5 to 187.5. General Combined Abilities (GCA) variances of the same size would be equivalent to parental heritability ranging from h^2 = 0.10 to 0.30 (for example, (4*62.5)/2500 = 0.10, where 2500 is the total phenotypic variance used across the scenarios, and 62.5 is the lowest parental GHA tested).

There is evidence of important SHA variation in pine and eucalyptus hybrids, particularly for growth traits. Although not many examples are reported in the literature specifically about the size of SHA variance, sometimes there is a combined effect of both parental GHA variances described (as a hybrid additive effect), and dominance variance is often reported as a proportion of the additive effect or a proportion of the total phenotypic variance. Across

many studies of hybrid populations of Eucalyptus and Pinus, it is not uncommon to find important SHA variance.

For example, in *E. grandis x E. urophylla* hybrids, for the trait of volume growth, estimates of the proportion of dominance range from $d^2 = 0.23$ to 0.27 (van den Berg et al., 2015, 2018).

In another *Eucalyptus* study, Chen *et al.* (2018) reported results for the hybrid between *Eucalyptus tereticornis x Eucalyptus urophylla*. For an experiment with a 10x10 mating design, with 59 families measured at 10-years, they found that SHA variance was intermediate between the GHA variance for the parent species. Zhu *et al.* (2017), in a hybrid population of *E. urophylla x E. camaldulensis,* using a reciprocal matting design of 6x6, with 36 families measured at 8.3 years for height and DBH, reported SHA variance higher than GHA variance for both species and both traits. In a study of *E. grandis x E. urophylla* where the pedigree of 1,000 genotypes was reconstructed through 33,398 SNPs and 24,001 DArt markers, de Lima *et al.* (2019) found a substantial non-additive effect in growth traits (height, DBH, and volume) where the d^2/h^2 ratio was sometimes 1 or higher than 1, depending on the reconstruction algorithm.

For the *Pinus* genus, Dieters *et al.* (1997) reported results for hybrids of subtropical pines *P. caribaea x P. oocarpa* and *P. caribaea x P. tecunumanii*. For both hybrid combinations, they examined a set of 61 families from an 11x6 factorial mating design, and they estimated the proportion of dominance as $d^2 = 0.07$ and 0.18, respectively. In a related study looking at older data and using the same 61 families plus an additional 20 families from a 5x4 factorial mating design, Gwaze *et al.* (2000) reported that the ratio of dominance to additive variance (σ_d^2/σ_a^2) for *P. caribaea x P. oocarpa* was $\sigma_d^2/\sigma_a^2 = 2.46$ for hgt and $\sigma_d^2/\sigma_a^2 = 1.08$ for DBH. In that study, the GHA variance was reported as the combined aggregate for the parent species, which implies that SHA variance was bigger than the average GHA variance for the parent species (σ_a^2). For the *P. caribaea x P. tecunumanii* hybrid, it was found a $\sigma_d^2/\sigma_a^2 = 0.91$ for height 0.71 for DBH, which implies again that SHA variance was an important effect, roughly similar to the results of *P. caribaea x P. oocarpa* variety. Finally, Belaber *et al.* (2018) studied a population of P. *elliottii x P. caribaea*, with a mating design of 16x21 parents, with

133 families, and they found that the size of the SHA variance was very similar to the average of the parental GHA variances for DBH growth.

All these studies showed an important SHA variation, supporting the hypothesis that this parameter is very significant in hybrid populations. Thus, it makes sense to include from medium to high values of true SHA variances in the different scenarios simulated to examine this effect on the precision and accuracy of all parameter estimates. For this reason, relatively high SHA variances with σ^2_{SHA} = 125 to 250 were assumed, which correspond to a proportion of dominance ranging from d^2 = 0.20 and d^2 = 0.40.

2.4.2 REML Variance Estimates Compared to the True and Sample Variances

The dispersion of the sample estimates $\hat{\sigma}_{GHA_B}$ and the REML estimates $\hat{\hat{\sigma}}_{GHA_B}$ around the true value $\sigma^2_{GHA_B}$, observed in Figure 2.1, was lower in scenarios with low GHA_B variance (scenarios 1 to 6) and higher in scenarios with high GHA_B variance (scenarios 6 to 12). This can be interpreted or summarized as the higher the true genetic variance, the lower will be the precision of a given variance estimate from a single iteration of an experiment. Over the long run, of course, the mean-variance estimates would still be precise, being around the true value.

The size of the SHA variance also had a small effect on the precision of the REML estimates of the GHA_A and GHA_B variances (Figure 2.1 and 2.2, Appendix A), and a mild effect on the correlation of the sample $\hat{\sigma}_{GHA}$ and $\hat{\hat{\sigma}}_{GHA}$ estimates. In summary, high SHA variance results in less precise estimates of the GHA variance parameters. This is seen by comparing Scenarios with the same parameters for $\sigma^2_{GHA_B}$, $Npar_B$ and NC_B but with different σ^2_{SHA} (e.g., Scenarios 1 vs 4, or 9 vs 12). For example, Scenarios 1 and 4 are exactly the same, except for σ^2_{SHA} = 125 and 250, respectively. Dispersion of the REML estimates $\hat{\hat{\sigma}}^2_{GHA_B}$ is greater in Scenario 4 than Scenario 1 (see Figure 2.1), and the standard deviation of the 1000 estimates is greater as well (SD= 46 versus SD= 36, see Appendix A). The same pattern was observed in Scenarios 9 vs. 12, and in all pairs of Scenarios differing only in the size of σ^2_{SHA}.

A somewhat analogous phenomenon was noted in a study by Kerr *et al*. (2004). They mentioned that evidence in the literature suggests that the correlation between purebred GCA and hybrid GHA performance decreases with increasing nonadditive genetic effects between parental populations. In the current study, the sample GHA variance and REML GHA variance estimate showed a decrease in the magnitude of the correlation when the SHA variance increases and the dispersion of the REML estimate increases.

Across the scenarios, a higher number of crosses per parent demonstrated a higher correlation between the sample GHA variance and the REML GHA estimates, no matter if the GHA or SHA was high or low. Roy (2000) examined partial-diallel mating designs in pure species scenarios, looking at GCA and SCA variance estimation and genetic value prediction, and he concluded: 1) that the presence of SCA variance substantially decreases the precision of estimates and increases chances of misinterpretation, and 2) that more than 8 or 10 crosses per parent should be unnecessary. In the current study, 4 and 8 crosses per parent in Species B demonstrated better quality GHA_B variance estimates than scenarios with 2 crosses per parent and the same GHA_B variance and SHA variance. In a Monte Carlo simulation study for breeding-population advancement, King & Johnson (1993) modeled and compared five different mating schemes. Two of these schemes were random mating designs, one with 2 and the other with 8 crosses per parent. It was found that having more crosses per parent can provide more genetic gain over five generations of breeding. This suggests that more crosses per parent in a hybrid breeding program would also result in a higher precision of the GHA estimates and a higher gain of the breeding population.

The dispersion of the REML $\hat{\hat{\sigma}}^2_{GHA_A}$ around the true value $\sigma^2_{GHA_A}$ was stable across scenarios, suggesting that 24 parents tested in 4 crosses will provide reasonably precise estimates of the genetic parameters for that species, regardless of the genetic parameters or specific mating design (number of parents and number of crosses per parent in the other species. This result was consistent with what was found in the REML estimates for species B, $\hat{\hat{\sigma}}^2_{GHA_B}$, where 24 parents tested in 4 crosses generally had lower standard deviations around the mean estimates than 12 parents with 8 crosses, or 48 parents with 2-crosses per parent, for scenarios with the same true parameters (see Appendix A).

SHA variance estimates were generally very precise and showed very stable performance across all scenarios. The correlation of the sample $\hat{\sigma}_{SHA}$ and the REML estimate $\hat{\hat{\sigma}}_{SHA}$ ranged from 0.50 up 0.74, showing a slight increase with increasing number of crosses per parent (Appendix C), along with a slight decrease in the standard deviation of $\hat{\hat{\sigma}}^2{}_{SHA}$. The higher level of SHA variance (σ^2_{SHA} = 250) also resulted in lower correlations between sample $\hat{\sigma}_{GHA}$ and REML estimate $\hat{\hat{\sigma}}_{GHA}$, and an increase in the standard deviation of the $\hat{\hat{\sigma}}^2{}_{GHA}$.

2.4.3 Factors Affecting the Frequency of Critical Errors

Across most scenarios, there was a relatively low frequency of "errors" and "severe errors" in the estimates of $\sigma^2_{GHA_A}$, $\sigma^2_{GHA_B}$, and σ^2_{SHA}. There were essentially no severe errors for SHA variance and only around 1% errors, and these were all overestimated. For GHA variances, with the exception of estimates of $\sigma^2_{GHA_B}$ in Scenarios 1-6, there was typically a very low frequency of errors. Across all 12 scenarios for GHA_A variance and Scenarios 7 to 12 for GHA_B variance, there were on average approximately 11.6% overestimates, with 1.2% severe overestimates and only 3.6% underestimates with 0.8% severe underestimates.

Only when GHA variance for one of the species is low is there some significant probability of error in the variance estimate, as Scenarios 1 to 6 with true $\sigma^2_{GHA_B}$= 62.5 (comparable to a pure species heritability of 0.10), especially with a high SHA value.

In Scenarios 1 to 3, with a moderate SHA variance of σ^2_{SHA} = 125, there was around a 14.9% chance of an overestimate error for $\sigma^2_{GHA_B}$, with 2.5% severe overestimates and around a 7.7% chance of underestimating error, with 2.9% severe underestimates. These errors were noticeably higher than the average of the 12 scenarios for GHA_A variance and Scenarios 7 to 12 for GHA_B, especially the severe errors, which were 2 to 4 times more considerable.

Now, for Scenarios 4 to 6, with high SHA variance (σ^2_{SHA} = 250), the frequency of overestimated errors for $\sigma^2_{GHA_B}$ was 17% with severe overestimates of 9.8%. These Scenarios had, by far, a much higher proportion of error and severe errors than the average

of the 12 scenarios for GHA_A variance and Scenarios 7 to 12 for GHA_B, with the severe underestimates frequency increased around 10 times. The worst-case scenario for errors in estimates of $\sigma^2_{GHA_B}$ was Scenario 6, with high SHA variance, high number of parents ($Npar_B$ = 48), and a low number of crosses for species B parents (NC_B = 2).

Other studies in the literature, while not directly comparable to the current study, seem to indicate that high levels of non-additive genetic variance and low numbers of crosses per parent in a full-sib mating design can lead to complications in quantitative genetic analyses.

In genomic studies in pure species, traits with high heritability often show higher genomic predictive accuracy than traits with low heritability (Ge et al., 2020). In this study, Species B was set with a low h^2 in Scenarios 1 to 6, corresponding to a hybrid h^2 = 0.10. This low h^2 for Species B resulted in a high frequency of error and severe error in the estimates, as was discussed previously. High heritability, or a high proportion of additive variance, would generally be associated with lower levels of dominance (or SCA variance). Similarly, De Almeida Filho *et al*. (2016) reported a decrease in the accuracy of the genomic prediction when the proportion of dominance variance (d^2) increases. Following this idea to the current study with hybrids, we could infer that higher SHA variance and lower GHA variance would decrease the accuracy of GHA variance estimates. In the Q-values histograms (Figure 2.3), this was observed in the spread of Q-values; higher GHA_B resulted in a lower spread of Q-values (Figure 2.3, scenario 7 to 12), being the estimates near to the true GHA (σ_{GHA}). The opposite was observed in low GHA_B scenarios (Figure 2.3, scenarios 1 to 6), where the frequency of error estimates was even greater in the scenarios with high SHA (Scenarios 4 to 6).

A number of studies with hybrid breeding populations, cited by Zhu *et al*. (2017), report that SHA variance was the main contributor to the total genetic variance, being larger than GHA variances in hybrid populations of *E. urophylla x E. grandis*, *E. urophylla x E. pellita* for volume, and hybrids of *E. grandis x E. urophylla*, *E. grandis x E. tereticornis* and *E .grandis x E. camaldulensis* for growth traits. For the hybrid population of *E. urophylla x E. camaldulensis,* Zhu *et al.* (2017) had an estimate of σ^2_{GHA} = 0 for one of the species for DBH at 8.3 years old, with a moderate level of SHA variance (corresponding to an estimate of

σ^2_{SHA} = 138 in this study). That study used a complete diallel of 6 x 6 parents, i.e., each parents of each species was tested in 6 crosses. This kind of scenario was not examined in the current study, but Scenario 1, probably the most comparable, with $Npar_B$ = 12 and NC_B = 8, had a frequency of underestimate errors = 8%, serious underestimates = 3%, and 2% of estimates of zero. It seems likely that the frequency of errors would be higher with just 6 parents of each species. In presence of a high SHA variance, breeders must be cautious on the level of confidence in parental GHA variance estimates.

2.4.4 Severe "errors" in the Variance Estimates: Implication in Breeding Strategies

The estimation of variance components for GHA of the two parent species, SHA, and the interactions of those with the environment (GxE) are important for hybrid breeding strategy development, that is, the decisions a breeder will make to implement a hybrid breeding program. If initial parameter estimates for a particular hybrid variety indicate large differences in the magnitude of the GHA variances, for example, the $\hat{\hat{\sigma}}^2_{GHA_A}$ is 2-times bigger than $\hat{\hat{\sigma}}^2_{GHA_B}$, a breeder might decide to include more parents for species A in the mating design than for species B in order to increase the selection intensity and expected gain from species A for the given trait.

Errors in the estimates of GHA variance could also affect the accurate estimation of gain for a breeding program. Perhaps the most serious possible error, and which was observed in the simulation results, is when we obtain severe underestimates or even estimates of GHA variance equal to zero for one of the parent species. This result has been observed with real data for hybrid trees; for example, van den Berg *et al*. (2017) reported zero GHA variance for the *E. grandis* parent for volume growth.

If a breeder obtains an estimate of $\sigma^2_{GHA_A}$= 0 or an estimate of $\sigma^2_{GHA_A}$ that is near zero, the expected inference would be that Species A has a very low impact on the gain for a given trait, and the breeder might ignore the opportunity to select parents of this species. This could be a serious error if, in fact, the true $\sigma^2_{GHA_A}$was not zero. In the other direction, we could have to deal with the overestimation of the GHA, which was observed in a lower

proportion of cases across the scenarios simulated. In this case, with an overestimate of GHA variance, a breeder will still likely test a number of parents of that species and will obtain gain, but lower than was expected. Furthermore, this would also lead to the eventual correction of the GHA variance estimate to a more precise (and lower) value. From this perspective, underestimates of σ^2_{GHA} are more critical errors than overestimates of σ^2_{GHA}.

2.4.5 Selection of Mating Strategy Based on Simulation Results.

The simulation results across the different scenarios of number of parents, number of crosses per parent, and the effect of several sizes of GHA and SHA variances suggest that to provide accurate estimates of variance components, a good hybrid mating design would be an incomplete diallel with at least 24 parents of each species, tested in 4 crosses. For the purposes of making gain, a population of between 20 to 50 parents should be large enough to sustain progress from selection through several generations of breeding in a pure species (McKeand & Bridgwater, 1998). A breeding population size of 20 and 50 is likely to maintain uncommon neutral alleles for many generations (Kang 1979, White *et al*. 2007). This can be a useful measure to assess an appropriate population size for multiple generations of recurrent selection, considering that selection criteria could change over time when new traits are selected or due to environmental changes, thus that what was once neutral alleles may become desirable in the future. A study of sustainability and robustness of genetic gain and diversity for breeding populations for Norway spruce on Swedish, conducted by Rosvall *et al*. (1998), concluded that populations with no fewer than 24 members are sustainable through ten breeding cycles. The conclusion was supported by the observation that inbreeding and inbreeding depression were not substantial, and the drop in additive gain per generation was negligible. Jayawickrama & Jefferson (1999), in a study of stochastic simulation of genetic advancement of multiple traits, stated that a strategy based on several populations of 20-30 unrelated parent clones might be viable for several generations, maintaining genetic diversity.

Now thinking of moving forward with a hybrid population, many breeding strategies are available to consider. For example, Reciprocal Recurrent Selection (RRS) is a breeding procedure where both pure species are improved cyclically for hybrid performance. This method allows capturing general and specific combining ability (GHA and SHA) to some extend from one cycle to another. However, this method is time-consuming and costly since it requires the recombination of individuals from the parental species, based on the hybrid progeny performance (Rezende et al., 2014). A method derived for RRS is the Reciprocal Recurrent Selection with Forward Selection (RRS-FS), where hybrid and pure species crosses are evaluated in the same generation. Selection is then made based on GCA and GHA values (Kerr, Dieters, & Tier, 2004; Perron, 2008). Other methods for hybrid breeding are Synthetic (SYN), where an F_1 population of hybrid is maintained by outcrossing, which means treated as a new species variety going forward.

If the long-term hybrid breeding strategy is RRS (improving both pure species for hybrid performance), 24 parents of each species would likely be a sufficient genetic base to maintain gain for many cycles of breeding, especially in the presence of a high SHA effect, since this method captures both GHA and SHA variation. van den Berg *et al*. (2018), in a study comparing two-hybrid breeding strategies and their effect in the realized gain in a hybrid population of *Eucalyptus grandis x E*. *urophylla* recommend the use of RRS-FS, since in this population for tree volume, the non-additive variance was higher than the additive variance (GHA). In that study, a total of 24 *E*. *grandis* and 23 *E*. *urophylla* selected parents were mated in different combinations.

Alternatively, if the long-term hybrid breeding strategy was the use a “synthetic” species, also called “composite” (i.e., a new hybrid variety, treated as a pure species going forward), 48 parents should also be sufficient to maintain cycles of hybrid selections, especially when the trait has high GHA variance and low SHA variance. Kerr *et al*. (2004) recommended a size of 100 parents for RRS scheme and 200 parents for SYN scheme in a simulation study comparing different breeding strategies in forest tree hybrid populations. That study used a very conservative (large) number of parents for RRS and SYN breeding schemes and did not examine the impact of variable numbers of parents and crosses. As discussed before,

many authors agreed with the use of 24 to 50 parents to assure gain in a long-term pure species breeding program, so in a hybrid breeding program, 2 x 24 parents per species = 48 parents should be enough to maintain gain and diversity in a synthetic population. Thus, more than 50 parent it is not a realistic number in a hybrid mating design, where incompatibility of crosses, abortion or asynchronous flowering time are some of the issues to obtain successful hybrid crosses.

In the current study, the precision of the GHA_B variance estimates was on average higher with 24 parents and 4 crosses per parent, and the estimates of GHA_A variance (based on a fixed 24 parents and 4 crosses per parent) were very stable across all scenarios, regardless of variation in the other genetic parameters. Scenarios with $Npar_B$ = 48 and NC_B = 2 did not markedly improve any of the variance component estimates. In some scenarios led to worse estimates of GHA_B variance when that value was low. Nevertheless, 24 parents of each species and 4 crosses per parent would still require a substantial effort to achieve that successful number of crosses per parent, noting that some particular interspecific crosses might fail, and the number of effective crosses could be lower.

2.5 Conclusions

For all scenarios examined, over 1000 iterations, the means of the sample variances and the REML estimated variances were exceptionally close to all true target variances for both GHA variances and SHA variance.

The different scenarios tested demonstrated that the size of the GHA variance, the number of parents, and the number of crosses per parent were all important factors in the precision of the GHA variance estimates. The current results indicated that, in general, good precision of the GHA variance estimates was achieved with 24 parents tested in 4 crosses. With more parents, the GHA variance should be better estimated, but maintaining at least 4 crosses per parent is recommended, which could be difficult to accomplish in a hybrid population. A smaller number of crosses per parent could affect the precision of the GHA variance

estimates dramatically, especially in the presence of high SHA variation for a given trait. This is not uncommon in some hybrid breeding programs of Eucalyptus species for growth traits, described in studies of Madhibha *et al*. (2013) and Zhu *et al*. (2017).

The size of the GHA variance is perhaps the most critical factor affecting the frequency of errors in variance component estimates. A low GHA variance will result in a larger spread of the GHA estimated variance relative to the true GHA variance. A combination of low GHA variance with high SHA variance could produce an unreliable GHA variance estimate far from the true GHA variance in the population. In this study, in scenarios with low GHA variance and high SHA variance, the probability of zero variance estimates ranged from 4% to 11%, and the probability of severe GHA underestimates variance ranged from 7 to 13%, with higher error rates associated with lower numbers of crosses per parent.

Breeders working with hybrids should be aware of the effects of GHA variance (or hybrid heritability), SHA variance, the number of parents, and the number of crosses per parent on the precision of the variance component estimates and the frequency of errors in the estimates. This can help a breeder to be aware of how realistic a zero GHA variance estimate could be expected to be in a hybrid population for a specific trait of interest and assess whether or not to ignore this effect in the formulation of mating and selection strategies.

2.6 References

Barton, N. H., Etheridge, A. M., & Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. Theoretical Population Biology, 118, 50–73. https://doi.org/10.1016/j.tpb.2017.06.001

Belaber, E. C., Gauchat, M. E., Reis, H. D., Borralho, N. M., & Cappa, E. P. (2018). Genetic Parameters for Growth, Stem Straightness, and Branch Quality for Pinus elliottii var. elliottii × Pinus caribaea var. hondurensis F1 Hybrid in Argentina. Forest Science, 64(Xx), 1–14. https://doi.org/10.1093/forsci/fxy021

Bison, O., Ramalho, M. A. P., Peçanha Rezende, G. D. S., Aguiar, A. M., & de Resende, M. D. V. (2007). Combining ability of elite clones of Eucalyptus grandis and Eucalyptus urophylla with Eucalyptus globulus. Genetics and Molecular Biology, 30(2), 417–422. https://doi.org/10.1590/S1415-47572007000300019

Bos, I., & Caligari, P. (2008). Selection methods in plant breeding: 2nd edition. In Selection Methods in Plant Breeding: 2nd Edition. https://doi.org/10.1007/978-1-4020-6370-1

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2017). ASReml-R Reference Manual Version 4. ASReml-R Reference Manual, 176. http://www.homepages.ed.ac.uk/iwhite/asreml/uop.

Céron-Rojas, J. J., & Crossa, J. (2018). Linear Selection Indices in Modern Plant Breeding Foreword by Daniel Gianola. https://doi.org/10.1007/978-3-319-91223-3

Chen, S., Weng, Q., Li, F., Li, M., Zhou, C., & Gan, S. (2018). Genetic parameters for growth and wood chemical properties in Eucalyptus urophylla × E. tereticornis hybrids. Annals of Forest Science, 75(1), 16. https://doi.org/10.1007/s13595-018-0694-x

De Almeida Filho, J. E., Guimarães, J. F. R., E Silva, F. F., De Resende, M. D. V., Muñoz, P., Kirst, M., & Resende, M. F. R. (2016). The contribution of dominance to phenotype prediction in a pine breeding and simulated population. Heredity, 117(1), 33–41. https://doi.org/10.1038/hdy.2016.23

de Lima, B. M., Cappa, E. P., Silva-Junior, O. B., Garcia, C., Mansfield, S. D., & Grattapaglia,

D. (2019). Quantitative genetic parameters for growth and wood properties in Eucalyptus “urograndis” hybrid using near-infrared phenotyping and genome-wide SNP-based relationships. PLoS ONE, 14(6), 1–25. https://doi.org/10.1371/journal.pone.0218747

Dieters, M. J., Nikles, D. G., Toon, P. G., & Pomroy, P. (1997). Genetic parameters for F1 hybrids of Pinus caribaea var. hondurensis with both Pinus oocarpa and Pinus tecunumanii. Canadian Journal of Forest Research, 27(7), 1024–1031. https://doi.org/10.1139/x97-053

Falconer, D. . S., & Mackay, T. F. . (1996). Introduction to quantitative genetics (4th edn). Pearson. https://doi.org/10.1002/bimj.19620040211

Fukatsu, E., Hiraoka, Y., Matsunaga, K., Tsubomura, M., & Nakada, R. (2014). Genetic relationship between wood properties and growth traits in Larix kaempferi obtained from a diallel mating test. Journal of Wood Science, 61(1), 10–18. https://doi.org/10.1007/s10086-014-1436-9

Ge, F., Jia, C., Bao, P., Wu, X., Liang, C., & Yan, P. (2020). Accuracies of genomic prediction for growth traits at weaning and yearling ages in yak. Animals, 10(10), 1–11. https://doi.org/10.3390/ani10101793

Gezan, S. A., de Carvalho, M. P., & Sherrill, J. (2017). Statistical methods to explore genotype-by-environment interaction for loblolly pine clonal trials. Tree Genetics & Genomes, 13(1), 1. https://doi.org/10.1007/s11295-016-1081-0

Gwaze, D. P., Dungey, H. S., Dieters, M. J., Toon, P. G., & Nikles, D. G. (2000). Interspecific pine hybrids. I. Genetic parameter estimates in Australia. Forest Genetics, 7(1), 11–20.

Hamilton, M. (2009). Population Genetics. United Kingdom: Wiley. https://doi.org/10.1017/CBO9781107415324.004

Hodge, G. R., & Acosta, J. J. (2020). An algorithm for genetic analysis of full-sib datasets with mixed-model software lacking a numerator relationship matrix function, and a comparison with results from a dedicated genetic software package. Forests, 11(11),

1–21. https://doi.org/10.3390/f11111169

Hodge, G. R., & Dvorak, W. S. (2012). Growth potential and genetic parameters of four Mesoamerican pines planted in the Southern Hemisphere. Southern Forests, 74(1), 27–49. https://doi.org/10.2989/20702620.2012.686192

Hoeschele, I., & VanRaden, P. M. (1991). Rapid Inversion of Dominance Relationship Matrices for Noninbred Populations by Including Sire by Dam Subclass Effects. Journal of Dairy Science, 74(2), 557–569. https://doi.org/10.3168/jds.S0022-0302(91)78203-9

Jayawickrama, K. J. S., & Jefferson, P. A. (1999). Stochastic simulation of genetic advancement of multiple traits using sequential culling. New Zealand Journal of Forestry Science, 29(1), 25–43.

Jonas, E., & de Koning, D. J. (2015). Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Frontiers in Genetics, 5(FEB), 1–8. https://doi.org/10.3389/fgene.2015.00049

Kabir, M., G.N, A., Nwagu, B.I., Adeyinka, I.A., Shehu, D.M., Galadima, & Yahaya, H. (2014). General combining ability (GCA), specific combining ability (SCA) and reciprocal effects on average daily gain in body weights at various ages of rabbit in northern guinea savannah zone of Nigeria. IOSR Journal of Agriculture and Veterinary Science, 7(4), 48–51. https://doi.org/10.9790/2380-07414851

Kang, H. (1979). Long-term tree breeding. In Proceedings 15th Southern Forest Tree Improvement Conference; Mississippi State University: MS: Eastern Tree Seed Laboratory, USDA Forest Service, Macon, GA 31202, June 19- 21, 1979. Southern Forest Tree Improvement Committee Sponsored Publication No. 37:66-72.

Kerr, R. J., Dieters, M. J., & Tier, B. (2004). Simulation of the comparative gains from four different hybrid tree breeding strategies. Canadian Journal of Forest Research, 34(1), 209–220. https://doi.org/10.1139/x03-180

Kerr, R. J., Dieters, M. J., Tier, B., & Dungey, H. S. (2004). Simulation of hybrid forest tree breeding strategies. Canadian Journal of Forest Research, 34(1), 195–208.

https://doi.org/10.1139/x03-181

King, J., & Johnson, G. (1993). Monte Carlo simulation models of breeding-population advancement. Silvae Genetica, 42(2), 68–78.

Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method 01 Mathematical Sciences 0104 Statistics. Genetics Selection Evolution, 50(1), 1–18. https://doi.org/10.1186/s12711-018-0426-6

Lstibůrek, M., Hodge, G. R., & Lachout, P. (2015). Uncovering genetic information from commercial forest plantations—making up for lost time using "Breeding without Breeding." In Tree Genetics and Genomes (Vol. 11, Issue 3). https://doi.org/10.1007/s11295-015-0881-y

Luan, Q. F., Li, Y. J., Jiang, J. M., Yue, H. F., & Diao, S. F. (2013). Discrimination of Pinus taeda × P. caribaea var. hondurensis between its allele-species and hybrids using near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 21(3), 203–211. https://doi.org/10.1255/jnirs.1051

Lynch, M., & Walsh, B. (1998). Genetic and Analysis of Quantitative Traits. Sinauer. https://books.google.com/books?id=UhCCQgAACAAJ

Macedo, F. L., Reverter, A., & Legarra, A. (2020). Behavior of the Linear Regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models. Journal of Dairy Science, 103(1), 529–544. https://doi.org/10.3168/jds.2019-16603

Madhibha, T., Murepa, R., Musokonyi, C., & Gapare, W. (2013). Genetic parameter estimates for interspecific Eucalyptus hybrids and implications for hybrid breeding strategy. New Forests, 44(1), 63–84. https://doi.org/10.1007/s11056-011-9302-8

Martinson, D. G., & Core, C. (2018). Quantitative Methods of Data Analysis for the Physical Sciences and Engineering. Cambridge University Press. https://doi.org/10.1017/9781139342568

McKeand, S. E., & Bridgwater, F. E. (1998). A strategy for the third breeding cycle of loblolly pine in the southeastern U.S. Silvae Genetica, 47(4), 223–234.

Mitchell, R. G., Wingfield, M. J., Hodge, G. R., Steenkamp, E. T., & Coutinho, T. A. (2013). The tolerance of Pinus patula × Pinus tecunumanii, and other pine hybrids, to Fusarium circinatum in greenhouse trials. New Forests, 44(3), 443–456. https://doi.org/10.1007/s11056-012-9355-3

Mrode, R. . R. A. R. . (2014). Linear Models for the Prediction of Animal Breeding Values. In CAB INTERNATIONAL (Ed.), Linear Models for the Prediction of Animal Breeding Values 3rd Edition (Vol. 3).

Mulder, H. A., Bijma, P., & Hill, W. G. (2007). Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. Genetics, 175(4), 1895–1910. https://doi.org/10.1534/genetics.106.063743

Norris, D., Ngambi, W., & Mbajiorgu, C. A. (2009). Estimates of Additive and Non-Additive Genetic Variances with Varying Levels of Inbreeding. Journal of Biological Sciences, 9(3), 254–258. https://doi.org/10.3923/jbs.2009.254.258

Perron, M. (2008). A strategy for the second breeding cycle of Larix x marschlinsii in Québec, Canada including experiments to guide interspecific tree breeding programme. Silvae Genetica, 57(4–5), 282–291. https://doi.org/10.1515/sg-2008-0043

Potts, B. M., & Dungey, H. S. (2004). Interspecific hybridization of Eucalyptus : key issues for breeders and geneticists. Forest Research, 27(2), 115–138. https://doi.org/10.1023/A:1025021324564

Rezende, G. D. S. P., de Resende, M. D. V, & de Assis, T. F. (2014). Eucalyptus Breeding for Clonal Forestry. In T. Fenning (Ed.), Challenges and Opportunities for the World's Forests in the 21st Century (pp. 393–424). Springer Netherlands. https://doi.org/10.1007/978-94-007-7076-8_16

Rosvall, O., Lindgren, D., & Mullin, T. J. (1998). Sustainability robustness and efficiency of a multi-generation breeding strategy based on within-family clonal selection. Silvae

Genetica, 47(5–6), 307–321.

Roy, D. (2000). Plant Breeding: Analysis and Exploitation of Variation. Alpha Science International. https://books.google.com/books?id=bLhdFLvg0KYC

Shen, P.-S., Cornelius, P. L., & Anderson, R. L. (1996). Planned Unbalanced Designs for Estimation of Quantitative Genetic Parameters II : Nested Matings. Journal of Agricultural, Biological, and Environmental Statistics, 1(4), 490–505.

Šimundić, A. M. (2013). Bias in research. Biochemia Medica, 23(1), 12–15. https://doi.org/10.11613/BM.2013.003

Singh, B. D., & Singh, A. K. (2015). Marker-assisted plant breeding: Principles and practices. In Marker-Assisted Plant Breeding: Principles and Practices. https://doi.org/10.1007/978-81-322-2316-0

Stockwell, P., & Richter, F. I. (1947). Hybrid Forest Trees. Yearb. Agr. 1943– 1947, 465–472.

Sutherland, I., Anderson, S., Auquier, A., Hauck, W. W., Oakes, D., Vandaele, W., & Weisberg, H. I. (1982). Statistical Methods for Comparative Studies: Techniques for Bias Reduction. In Biometrics (Vol. 38, Issue 3). https://doi.org/10.2307/2530082

Team R. Core. (2019). R: A language and environment for statistical computing. In R Foundation for Statistical Computing. https://doi.org/10.1108/eb003648

van den Berg, G. J., Verryn, S. D., Chirwa, P., & van Deventer, F. (2017). Genetic parameters and genotype by environment interaction of Eucalyptus grandis populations used in intraspecific hybrid production in South Africa. Southern Forests: A Journal of Forest Science, 79, 287–295.

van den Berg, G. J., Verryn, S. D., Chirwa, P. W., & Deventer, F. Van. (2015). Genetic Parameters of Interspecific Hybrids of Eucalyptus grandis and E. urophylla Seedlings and Cuttings. Silvae Genetica, 64(1–6), 291–308. https://doi.org/10.1515/sg-2015-0027

van den Berg, G. J., Verryn, S. D., Chirwa, P. W., & van Deventer, F. (2018). Realised genetic

gains and estimated genetic parameters of two Eucalyptus grandis × E. urophylla hybrid breeding strategies. Southern Forests, 80(1), 9–19. https://doi.org/10.2989/20702620.2016.1263010

Volker, P. W., Potts, B. M., & Borralho, N. M. G. (2008). Genetic parameters of intra- and inter-specific hybrids of Eucalyptus globulus and E. nitens. Tree Genetics and Genomes, 4(3), 445–460. https://doi.org/10.1007/s11295-007-0122-0

Walsh, B., & Lynch, M. (2018). Evolution and Selection of Quantitative Traits. OUP Oxford. https://books.google.com/books?id=L2liDwAAQBAJ

Weisberg, H. I. (2010). Bias and Causation. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470631102

White, T. L., Adams, W. T., & Neale, D. B. (2007). Forest Genetics. In CABI Publishing. CABI Publishing. https://doi.org/10.1038/167764a0

Yadav, S., Jackson, P., Wei, X., Ross, E. M., Aitken, K., Deomano, E., Atkin, F., Hayes, B. J., & Voss-Fels, K. P. (2020). Accelerating genetic gain in sugarcane breeding using genomic selection. Agronomy, 10(4), 1–21. https://doi.org/10.3390/agronomy10040585

Zhu, Y., Wu, S. J., Xu, J., Lu, Z., Li, G., Hu, Y., Yang, X., & Bush, D. (2017). Genetic parameters for growth traits and stem-straightness in Eucalyptus urophylla × E. camaldulensis hybrids from a reciprocal mating design. Euphytica, 213(7), 1–16. https://doi.org/10.1007/s10681-017-1923-3

Zobel, B. J., & Talbert, J. T. (1984). Applied Forest Tree Improvement. The Blackburn Press.

Chapter 3:

Quantitative Genetics of a Hybrid Population of *Eucalyptus nitens x Eucalyptus globulus*: Estimation of Genetic Parameters and Implications for Breeding Strategies

3.1 Introduction

The genus *Eucalyptus* contains over 700 species, many of which are widely planted for industrial forestry purposes. As a genus, *Eucalyptus* is appreciated for its fast-growing, valuable wood properties and wide adaptability. Several *Eucalyptus* species were introduced successfully since the first quarter of the 1800s into many countries, including India, France, Portugal, Congo, South Africa, Zimbabwe, Brazil, and Chile. Today *Eucalyptus* species are one of the largest sources of woody biomass globally and are the most extensive plantation hardwoods used for pulpwood and timber production (Madhibha et al., 2013).

Chile is one of the top ten countries in terms of land dedicated to forestry plantations (Salas et al., 2016). In Chile, the most common hardwood species used across the country are *E. globulus* and *E. nitens*, representing a large proportion of the total forest land base of 2.3 million hectares, with 25.3% planted with *E. globulus* and 11.9% with *E. nitens* (INFOR, 2020). Both Eucalyptus species are originally from Australia and are well adapted to the Chilean soil and weather after generations of breeding. Breeding work of these species began with provenance tests and progressed through cycles of improvement using both open-pollination and control-pollination. In the last decade, many operational programs began utilizing clonal propagation of these species for some of their commercial deployment. The choice between *E. nitens* and *E. globulus* for commercial plantations across the country depends mainly on the environment. *E. globulus* is considered the premium species for pulp and paper production due to its promising growth and excellent overall pulpwood properties. However, the relatively poor frost tolerance of this species in

the south of the country restricts its use. On frost-prone sites, *E. globulus* is replaced by *E. nitens*, a more frost-tolerant species (Humphreys et al., 2008), which also exhibits extremely fast growth. Under these circumstances, the hybridization of these two species was the next step in Chilean breeding programs.

In forestry, hybridization is understood as a successful mating between individual trees of two different species (White et al., 2007), typically resulting in progeny with intermediate characteristics of its parents. In some cases, the hybrid progeny performs better than the average of parents, and this phenomenon is referred to as *mid-parent heterosis*. In other cases, the hybrid progeny performs better than the best parent. This behavior is called *high-parent heterosis* (White et al., 2007), which is very desirable in developing a new commercial population or variety of trees.

Interspecific hybrid crosses are often made with the goal of combining specific traits from each parental species (Luan et al., 2013), such as growth rate, resistance to diseases, adaptability to climatic conditions, and wood quality. However, the successful hybridization of trees can be difficult, and it is necessary to consider the compatibility of the two species, flowering time, and also the direction of pollination, that is, whether to use species A as the female and species B as the male, or vice versa (Potts & Dungey, 2004; Stockwell & Richter, 1947). Often in the development of hybrid crosses, relatively few of them are successful, with the result that not as many parents have the chance to be tested for their value as a hybrid parent in comparison with the number of parents that can be tested in classic pure species breeding programs. In this scenario, it is not easy to obtain a balanced progeny test design with a similar number of parents and crosses to precisely estimate the genetic parameters and parental genetic values. This matter has to be considered when diallel crosses are planned in a hybrid breeding program.

In the Chilean forestry context, the goal of hybridization of these species was the combination of frost tolerance and high growth rate from *E. nitens* with the favorable wood properties of *E. globulus*, thus exploiting the best characteristics of both species (Griffin et al., 2000; Humphreys et al., 2008; Volker et al., 2008). A complicating factor in the hybridization of *E. globulus* and *E. nitens* is a structural pre-zygotic barrier due to flower

size. *Eucalyptus globulus* has a much larger flower than *E. nitens*, and this constrained the directions of the crosses, where *E. nitens* must be the female and *E. globulus* the male pollen donor. Even with crosses done in this direction, typically, a low number of seeds are obtained (Humphreys et al., 2008; Potts et al., 2000).

Despite the complicated process of obtaining hybrid progeny of these two Eucalyptus species, successful full-sib hybrid progenies have been tested by ARAUCO Forestry Company in Chile, using clonal propagation of progeny from the successful crosses. The parents of those crosses came from pure species breeding programs, where the *E. globulus* parents were selected from several controlled-pollination (CP) field tests, and the *E. nitens* parents were selected from several open-pollinated (OP) tests. From both pure and hybrid trials, information on breeding values of growth and wood properties was obtained from two major breeding zones from ARAUCO.

In the pure species programs of *E. nitens* and *E. globulus* and the hybrid between them, the target traits for improvement have been growth and wood properties, as these are the most valuable economic traits in Chilean plantations. The wood properties related to pulp and paper production are Pulp Yield (%), Specific Consumption (m^3/ADt, where ADt stands for **A**ir-**D**ry Metric **T**ons of wood and is the product of wood volume (m^3), and Basic Density (kg/m^3). Basic density is also an important trait for solid wood products, where typically, a higher basic density is associated with increased stiffness and strength of the wood (S. Chen et al., 2018). In order to rapidly assess a large number of trees, non-destructive wood samples were used along with near-infrared reflectance spectroscopy (NIRS) to obtain estimates of this hybrid population's physical and chemical wood properties, including pulp yield basic density and specific consumption.

In this research study, a large clonal population of *E. nitens x E. globulus* hybrid was evaluated, using quantitative genetics methods to estimate important genetic parameters for tree volume and wood properties, and then to predict clonal genetic values, General Hybridizing Ability (GHA) for parents of both species, and Specific Hybridizing Ability (SHA) for the specific full-sib crosses. Hybrid genetic values were compared with analogous genetic values from pure species testing to elucidate if there is a relationship with the

known parental performance as pure species and evaluate the possible implications for the hybrid breeding strategies of this variety.

3.2 Material and Methods

3.2.1 Description of Field Tests and Locations

A large full-sib clonal population of *Eucalyptus nitens x Eucalyptus globulus* (GloNi) was established in the middle-south of Chile, from the Bio-Bio to Los Lagos Region. The total number of sites was 27, distributed in two breeding zones, Arauco and Valdivia (Figure 3.1).

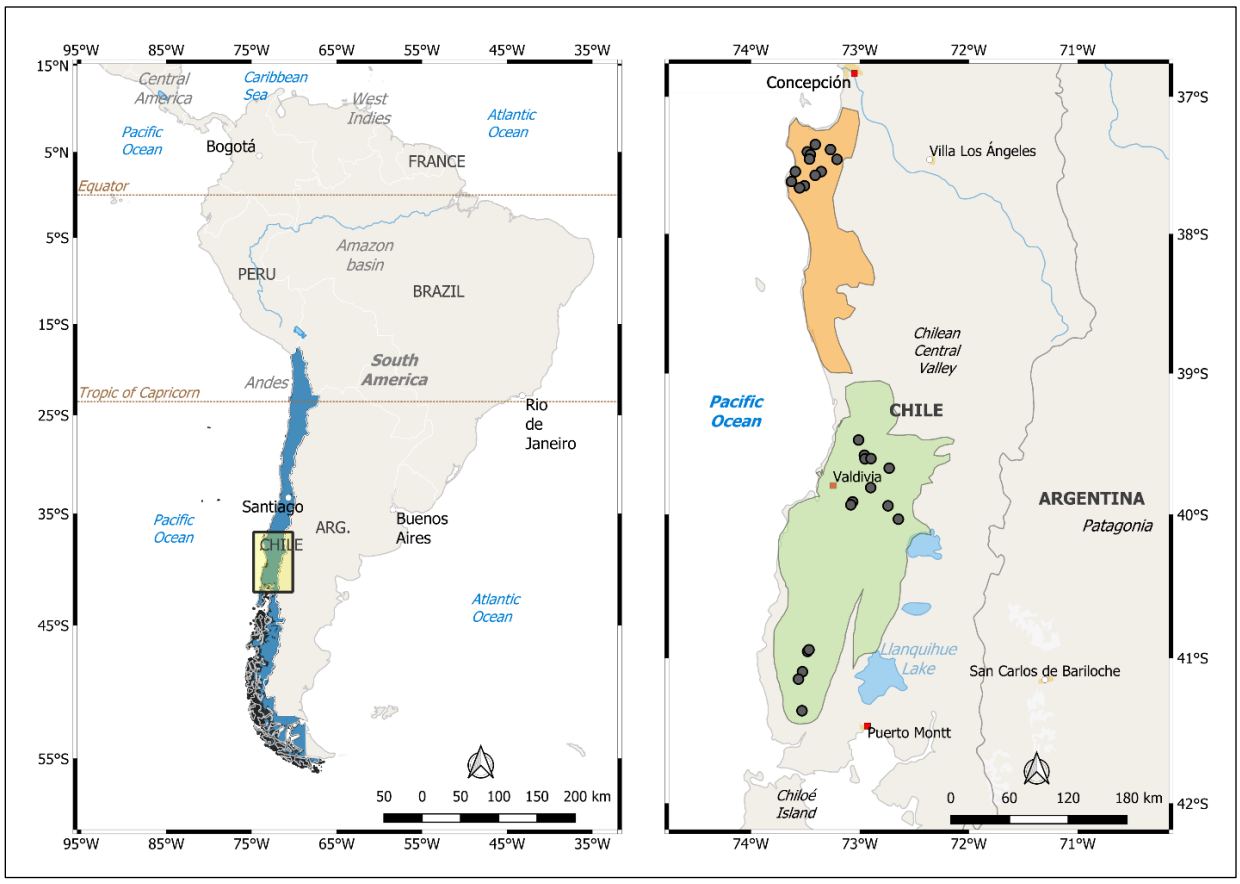


Figure 3.1. Map of *E. nitens x E. globulus* trial distribution. Left: global location of the trials, with the country region depicted in blue and the study location in yellow. Right: detailed location of the trials per breeding zone. Represented in orange is the Arauco zone, and in green the Valdivia zone. Each black dot represents the location of a field test.

The Arauco zone (represented in orange in Figure 3.1) has a Temperate Coastal climate, with oceanic influence near to the coast in the north and a Temperate Rainy Oceanic climate in the south of the zone, according to the regional macro descriptions of the country (BCN, 2021). Yearly seasonal changes influence the monthly mean temperature (MMT) and the monthly mean precipitation (MMP). Through the coldest months of winter, between June and July, the minimum mean monthly temperature is around 6.5 °C, while in the summer season, from December to January, the maximum mean monthly temperature is close to 21 °C. The total annual precipitation is approximately 1,360 mm, concentrated mainly during the winter season (see Figure 3.2 for more details). Soils in this zone are generally deep, derived principally from metamorphic rocks and marine sediments, and secondarily formed from ancient volcanic ashes.

The Valdivia zone (green in Figure 3.1) has a Temperate Rainy climate, with abundant precipitation and a low chance of dry periods during the summer (BCN, 2021). The minimum mean monthly temperature of the year is near to 3.6 °C during the winter season, and the maximum monthly mean temperature is roughly 22 °C in summer. The total annual precipitation is close to 1,900 mm, with a precipitation peak at the beginning of winter (Figure 3.2). Soils in this zone are deep, with good permeability, derived principally from ancient volcanic ashes deposited over a metamorphic rock complex and secondarily formed from newer volcanic ashes. In some places of this zone, the secondary parental material is marine sediments.

In terms of weather, the main difference between Arauco and Valdivia zones are mainly rainfall amount (mm) and mean temperature (°C). Arauco zone is on average warmer than Valdivia, whereas Valdivia is colder and with more precipitation than Arauco.

The minimum and maximum monthly mean temperature and the monthly mean precipitation were estimated with data of 30 years (1987-2017) from the Climatic Explorer of Chile (CR2, 2019) for each of the described zones, and it is depicted in Figure 3.2.

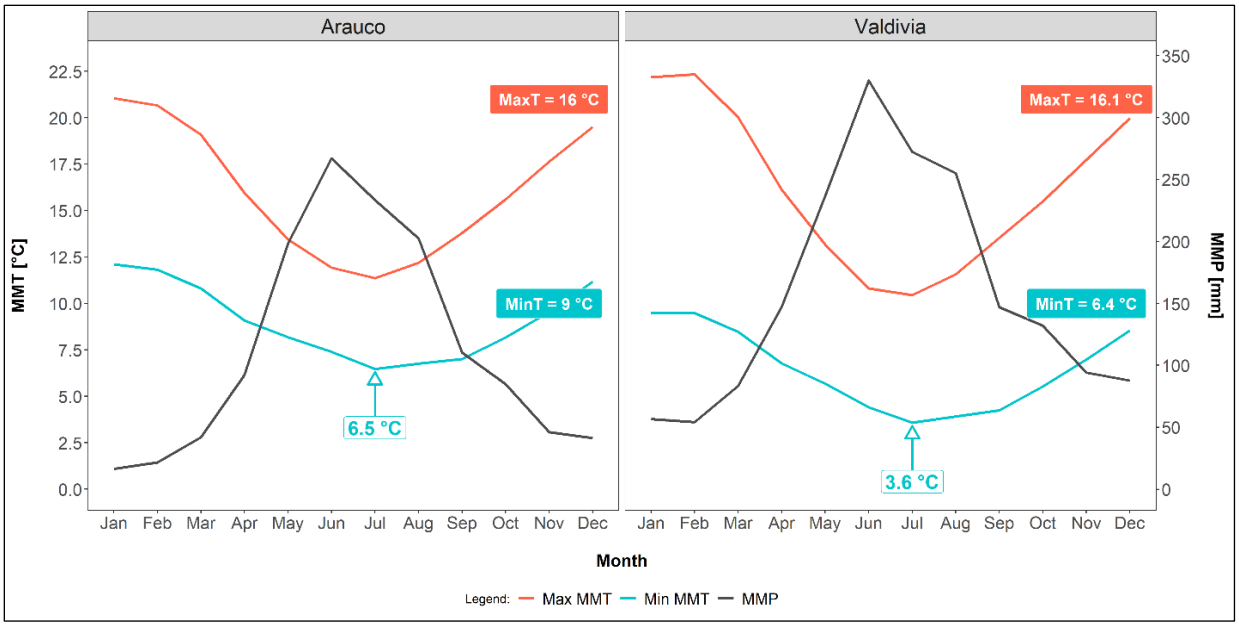


Figure 3.2. Yearly distribution of temperature and rainfall in Arauco and Valdivia breeding zones with data of 30 years (1987-2017). The primary Y-axis represents the Mean Monthly Temperature (MMT) in °C, and the secondary Y-axis represents the Mean Monthly Precipitation (MMP) in mm. The orange lines represent the maximum MMT and the blue line minimum MMT (°C). The gray line represents the monthly mean precipitation (mm). The orange label with the title MaxT represents the mean of the maximum MMT, and the blue label is the mean of the minimum MMT. The blue arrows in July represent the lowest minimum MMT of the year.

From the same dataset of CR2, the number of days under 0° C from May to August were counted to obtain the proportion of days with frost events during the winter months in both zones (Table 3.1). With the measurements over 30 years, it was found that the Valdivia zone had around 3-times more frost events than the Arauco zone. In total (during the months of May to August), Valdivia averaged 10.2 frost days per year, compared to 3.1 days in Arauco. This ratio also holds during the month of July, which is the coldest month of the year, with 13% of the total days of the month with frost events (Table 3.1).

Table 3.1. Average of 30 years period of observations of the proportion of days under 0 °C degrees on Arauco and Valdivia zones, summarized the months from May to August. Total frost days are estimated based on the proportion of days under 0°C and the number of days of each month from May to August.

Zone	Month				Total frost days
	May	Jun	July	August	
Arauco	1%	2%	4%	3%	3.1
Valdivia	6%	8%	13%	6%	10.2

3.2.2 Tests Design and Measurements

Field tests followed a randomized complete block design (RCBD), with 10-repetitions in each site, with the same experimental design in both breeding zones. Treatments (hybrid clones) were established in single-tree plots (STP) in each replication. Trees were planted with a spacing of 3 m x 2.5 m, which generates a planting density of approximately 1,667 trees per hectare (ha). The clonal population was derived from 28 full-sib crosses, among 12 *Eucalyptus nitens* (NIT) parents and 8 *Eucalyptus globulus* (GLO) parents. The total number of clones tested in Arauco zone was 1,260 across 12 sites. In the Valdivia zone, there were 1,214 clones tested across 17 sites. More or less, the same population of clones was tested in both zones, and the number of clones per full-sib family was on average 47 in Arauco and 53 in Valdivia.

For each trial, the following traits were measured: survival (%), diameter at breast high in cm (DBH), total height in m (HT), forking, the incidence of pests, and presence of broken tops. Growth traits (DBH and HT) were recorded for all trees at 4, 5, and 8 years across sites. The volume for each tree was estimated using the Ladrach (1986) formula for juvenile trees as follows:

$$Volume = 0.00003 * DBH^2 * HT \quad \text{Eq. 3.1}$$

The wood property traits Basic density (kg/m^3), Pulp yield (%), and Specific consumption (m^3/ADt) were estimated using Near-Infrared Reflectance Spectroscopy (NIRS), with a prediction model adjusted by the ARAUCO company. This model was calibrated with 365 samples and validated with 64 for each trait. The calibration fit statistic for this model was R_C^2=0.8 for all traits, with a RMSCV value of 24.4 kg/m^3 for BD, 1.18% for Pulp yield, and 0.25 m^3/ADt for Specific consumption, where R_C^2 is the coefficient of determination for calibration and RMSECV root mean square error of cross-validation. The coefficient of determination for prediction performance (R_P^2) was near to R_C^2 with prediction errors (root mean square error of prediction (RMSEP)) maintaining the same magnitude (Appendix E).

The non-destructive wood samples were taken for the best 20% genotypes in volume gain identified by earlier analysis (BLUP) within zones. Around 4 to 6 ramets were sampled across sites within zones at 6-years after tree establishment.

3.2.3 Individual Tree Volume Data Analysis and Cleaning

For the accurate estimation of individual tree volume (m^3/ha), trees with abnormalities were removed from the analysis, including trees with pests, broken tops, strong stem sinuosity, more than two main stems, and dead trees. Similarly, trees with an extreme DBH/HT ratio were removed (values > 3 or < 0.3), assuming that these trees had some type of measurement error or unreported damage, such as a broken top.

In each breeding zone, estimations of individual tree volume and survival rate were performed, followed by data cleaning and the calculation of descriptive site and group-age statistics of survival, growth, and wood traits through an R-software (Team R. Core., 2019) code written for that purpose.

3.2.4 Individual Tree Volume Standardization

The volume (m^3) of each tree was standardized prior to the estimation of genetic parameters to deal with scale effect due to Genotype × Environment (GxE) interaction variances, arising from differences in growth trait means, reflecting different productivity across the sites (Hodge & Dvorak, 2015). Commonly, sites with high productivity produce larger trees with higher phenotypic variances than sites with less productivity (Gezan et al., 2017). In forestry, growth traits frequently have a strong relationship between the mean of the trait (DBH, HT, or volume) and its phenotypic and genetic variances (Gapare & Musokonyi, 2002; Hodge & Dvorak, 2015; van den Berg et al., 2015). To correct for heterogeneous variance of volume across different test sites, the volume of each tree was standardized for each block within test, expressing the tree volume as a deviation from the block volume mean divided

by the standard deviation: $y - \bar{y}/\sigma$, where "*y*" is the observed tree volume, $\bar{y}$ is the mean volume within block and σ is the standard deviation of the volume within block.

Coefficients of variance (CV) for volume were calculated per block, and then the average CV across all sites was calculated (CV_y). Finally, the standardized volume for each tree (i.e., each ramet of each clone in each test) was estimated as:

$$stVol = \frac{y-\bar{y}}{\sigma} * (CV_y * 100) + 100 \quad \text{Eq. 3.2}$$

The standardized volume (stVol) is indicated in units of %, where the population's mean is centered on 100%, and the spread of the phenotypic data is expressed as CV_y*100 (White et al., 2007). Consequently, all variance components estimates and predicted breeding values could be interpreted in terms of gain (above or below 100%), without the necessity of rescaling (Hodge & Dvorak, 2015; van den Berg et al., 2015). Data standardization was conducted through an R script.

3.2.5 Estimation of Variance Components and Genetic Parameters

The phenotypic observations of wood properties and the standardized tree volume (stVol) were analyzed via restricted maximum likelihood (REML; Patterson & R., 1971 cited by Harville, 1977) using linear mixed models (LMM). Single- and multi-site analyses were conducted using ASReml-R (Butler, 2019; Butler et al., 2017; Team R. Core., 2019) by following an LMM, which includes the effect of both parental species. The single-site linear model used for the evaluation of each test is the following:

$$y_{ijklm} = u + B_i + GHA_{NIT_j} + GHA_{GLO_k} + SHA_{jk} + Clw_l + BxClw_{il} + e_{ijklm} \quad \text{Eq. 3.3}$$

Where:

y_{ijklm} is the m^{th} observation for the jk^{th} family for the l^{th} clone at the i^{th} block; *u* is the overall mean of the site; B_i is the fixed of the i^{th} block, GHA_{NIT_j} or GHA_{GLO_j} is the random General Hybridizing Ability (GHA) effect for the j^{th} female *E. nitens* or k^{th} male *E. globulus* parent; SHA_{jk} is the random Specific Hybridizing Ability (SHA) effect of the hybrid interaction

between the j^{th} female parent and the k^{th} male parent; Clw_l is the random clonal effect of the I^{th} hybrid clone; $BxClw_{il}$ is the random effect of the interaction between the i^{th} block and the l^{th} clone within hybrid family and e_{ijklm} is the random effect within plot error term.

With the variance components obtained from the model above [Eq.3.3], the phenotypic variance was estimated as:

$$\sigma^2_{phen} = \sigma^2_{GHA_{NIT}} + \sigma^2_{GHA_{GLO}} + \sigma^2_{SHA} + \sigma^2_{Clw} + \sigma^2_{BxClw} + \sigma^2_e \quad \text{Eq. 3.4}$$

Broad-sense heritability of single site (H^2_b) was estimated with formulas derived by Falconer (1996):

$$H^2_b = \frac{(\sigma^2_{GHA_{NIT}} + \sigma^2_{GHA_{GLO}} + \sigma^2_{SHA} + \sigma^2_{Clw})}{\sigma^2_{phen}} \quad \text{Eq. 3.5}$$

Previous studies have called the single-site heritability "biased heritability" (H^2_b) and called multi-environment or multiple-site heritability the "unbiased heritability" (H^2) (Dieters et al., 1995). These terms reflect the fact that single-site genetic variances are typically larger than multiple-site genetic variances due to the inability to separate genotype x environment variances; thus, these are "biased" upward.

After inspecting the estimates of H^2_b from single-sites, a combined-site analysis was performed. The statistical LMM model was similar to the single-site (Eq. 3.3), with the difference that terms related to the site were added, along with their interactions. The statistical model for combined-site analysis is presented below:

$$y_{ijklmn} = u + S_i + B_{j(i)} + GHA_{NIT_k} + GHA_{GLO_l} + SHA_{kl} + Clw_m + SxGHA_{NIT_{ik}} + SxGHA_{GLO_{il}} + SxSHA_{ikl} + SxClw_{im} + err_{ijklmn} \quad \text{Eq. 3.6}$$

Where:

y_{ijklmn} is the n^{th} observation of the j^{th} block for the kl^{th} family for the m^{th} clone at the i^{th} site; u is the overall mean; S_i is the fixed of the i^{th} site; $B_{j(i)}$ is fixed effect of the j^{th} replication within the i^{th} site; GHA_{NIT_k} or GHA_{GLO_l} is the random General Hybridizing Ability (GHA) effect for the k^{th} female of *E. nitens* parent or the l^{th} male of *E. globulus* parent; SHA_{kl} is the random

Specific Hybridization Ability (SHA) or full-sib family effect of the k^{th} and the l^{th} parents; Clw_m is the random effect of the m^{th} clone within hybrid family; $SxGHA_{NIT_{ik}}$ or $SxGHA_{GLO_{il}}$ is the random effect of the interaction between the i^{th} site and the k^{th} *E. nitens* female parent or the l^{th} *E. globulus* male parent; $SxSHA_{ikl}$ is the random effect of the interaction between the i^{th} site and k^{th} female with l^{th} male; $SxClw_{im}$ is the random effect of the interaction between the i^{th} site test and the m^{th} hybrid clone and err_{ijklmn} is the random effect within plot error term.

The effects estimated in models presented in Eq.3.3 and Eq.3.6 were assumed to be random and independently distributed, except for overall mean, site, and block, which are fixed effects (Madhibha et al., 2013; van den Berg et al., 2015).

The estimation of genetic parameters for the multi-site analysis was obtained following the derivation of Falconer & Mackay (1996) with the next formulae:

Phenotypic variance: $\sigma^2_{phen} = \sigma^2_{GHA_{NIT}} + \sigma^2_{GHA_{GLO}} + \sigma^2_{SHA} + \sigma^2_{Clw} + \sigma^2_{SxGHA_{NIT}} + \sigma^2_{SxGHA_{GLO}} + \sigma^2_{SxSHA} + \sigma^2_{SxClw} + \sigma^2_{err}$ Eq. 3.7

Genetic variance: $\sigma^2_G = \sigma^2_{GHA_{NIT}} + \sigma^2_{GHA_{GLO}} + \sigma^2_{SHA} + \sigma^2_{Clw}$ Eq. 3.8

Broad-sense heritability: $H^2 = \sigma^2_G / \sigma^2_{phen}$ Eq. 3.9

In pure species breeding, the narrow-sense heritability (h^2) is often estimated as 4 times the General Combining Ability (GCA) variance divided by phenotypic variance. Sometimes, authors working with hybrid populations will estimate a hybrid narrow-sense heritability for each parental species by multiplying the parental GHA variance by 4 and then dividing by the total phenotypic variance (Dieters et al., 1997; Madhibha et al., 2013; Mitchell et al., 2013; van den Berg et al., 2015; Zhu et al., 2017). In this research, just the broad-sense heritability will be reported (H^2), and total genetic variance is described as the GHA variances for the two parent species, SHA variance, and clone within family (Clw) variance.

Type-B genetic correlations were obtained for all genetic effects, clone within family, cross and parental level, estimated as:

All genetic effect: $rB_G = (\sigma_G^2)/(\sigma_G^2 + \sigma_{SxClw}^2 + \sigma_{SxGHA_{NIT}}^2 + \sigma_{SxGHA_{GLO}}^2 + \sigma_{SxSHA}^2)$ Eq. 3.10

Clone within family: $rB_{cl} = \sigma_{Clw}^2/(\sigma_{Clw}^2 + \sigma_{SxClw}^2)$ Eq. 3.11

Cross: $rB_{nxg} = \sigma_{SHA}^2/(\sigma_{SHA}^2 + \sigma_{SxSHA}^2)$ Eq. 3.12

NIT: $rB_{nit} = \sigma_{GHA_{NIT}}^2/(\sigma_{GHA_{NIT}}^2 + \sigma_{SxGHA_{NIT}}^2)$ Eq. 3.13

GLO: $rB_{glo} = \sigma_{GHA_{GLO}}^2/\left(\sigma_{GHA_{GLO}}^2 + \sigma_{SxGHA_{GLO}}^2\right)$ Eq. 3.14

The type-B genetic correlations, described by Burdon (1977), is a measure of the degree of Genotype × Environment interaction (GxE), when the same trait is measured in genotypes established in different environments (often called multi-environment trial analysis, MET). The magnitude of GxE interaction ranges from 0 to 1, where values near 0 indicate weak agreement, and close to 1 a near-perfect correlation between the performance of the genotypes for the trait of interest measured across the different environments.

Studies with cloned progeny from controlled crosses allow for the estimation of additive and non-additive genetic variances and the breakdown of the non-additive variance into an estimate of dominance and epistasis variances. Foster & Shaw (1988) showed that with full-sib clonal data, an approximation of the epistasis variance ($\hat{\sigma}_i^2$) can be calculated for pure species. Their methodology has been followed by many authors in studies of growth and wood properties, for example, in *Eucalyptus globulus* (Araújo et al., 2012; Costa E Silva et al., 2004, 2009) and in the estimation of the genetic parameters for rooting in loblolly pine (Baltunis et al., 2005). Adapting Foster's equation to the hybrid linear model used in this study, the epistasis effect was calculated for all trait in both zones as follow:

$\hat{\sigma}_i^2 = \hat{\sigma}_{Clw}^2 - (\hat{\sigma}_{GHA_{NIT}}^2 + \hat{\sigma}_{GHA_{GLO}}^2) - 3\hat{\sigma}_{SHA}^2$ Eq. 3.15

Where $\hat{\sigma}_{Clw}^2$ is the clone-within-family variance, $\hat{\sigma}_{GHA_{NIT}}^2$ and $\hat{\sigma}_{GHA_{GLO}}^2$ are the respective nitens and globulus GHA variances and $\hat{\sigma}_{SHA}^2$ is the SHA variance.

3.2.6 Consistency of Parents for Pure Species Progeny and Hybrid Progeny Performance

Forest tree breeders working with hybrids will often maintain breeding populations of both parental species, and periodically, new hybrid crosses are made between these populations. One method involves intra-species recurrent selection made in the parental populations, selecting based on pure-species progeny performance, and using the selected parents for interspecific species crosses, i.e., the production of new hybrids for commercial plantation establishment. This method is known as Recurrent Selection for General Combining Ability (RSGCA) and requires hybrid crosses only for deployment. However, it is necessary to know the correspondence between the GCA value of each parent with their GHA to see if there is a relationship between the parental ability to produce superior progeny in intra-specific and inter-specific crosses. A high correlation between these GCA and GHA parameters indicates that parents can be selected for interspecific crosses based on their pure species performance for a given trait.

When the GCA is not a good predictor of the GHA, perhaps due to high levels of dominance variance (SHA or SCA), epistasis variance, or other factors, Reciprocal Recurrent Selection (RRS) is another breeding strategy to develop new hybrid populations. This method is known for many breeders, and it was defined by Comstock et al. (1949) as a method of selecting inbred lines through SCA values in maize.

Both breeding strategies have been adopted successfully in forestry through the years, with interspecific crosses of species of the genus Populus, Eucalyptus, and Pinus, to mention some of them.

For the current *E. nitens* and *E. globulus* populations, elite parents' selections were made based on their General Combining Ability (GCA) estimated from pure species field tests established in the Arauco and Valdivia breeding zones by ARAUCO Company. The *E. nitens* parents were tested in several open-pollinated (OP) trials in both breeding zones. The *E. globulus* parents were also tested with control-pollinated (CP) and clonal field tests in both breeding zones.

Prediction of pure species GCA values for growth was done internally by the ARAUCO breeding team, using standardized volume from ages 5 to 12 years across multiple sites. For wood property traits, phenotypic data were obtained from tree samples in each breeding zone for roughly the top 20% of the genotypes for 6-year-old volume gain. The wood properties assessed were Basic density (kg/m^3), Specific Consumption (m^3/ADt), and Pulp Yield (%) with NIR spectroscopy, using similar techniques as were used for the hybrid progeny.

The parental GCA values of *E. nitens* and *E. globulus* were compared with their GHA value using Pearson correlations to evaluate the consistency of the parental performance evaluated as pure species parents and as hybrid parents. A high correlation would indicate that parental GCA could be a good indicator for GHA performance. The Pearson correlation and their significance level were determined by calculating the t value. Then, the corresponding p-value was determined using the t distribution table for df=n−2 through the native 'stats' package of R software for volume, basic density, specific consumption, and pulp yield traits.

3.3 Results

3.3.1 Growth

Average survival across tests in the Arauco zone was 93%, and in the Valdivia zone was 80%. In both zones, the survival decreased as the age of the trials increased, which is expected due to competition for light, water, and nutrients. The average volume per tree at 8-years was 0.265 m^3 in the Arauco zone and 0.237 m^3 in the Valdivia zone. This higher growth in Arauco than in Valdivia was also observed in earlier measurements. It was not detected a high dispersion from the mean in growth across the trials within age-zone, identified through lower SD for DBH (cm), HT (m), and volume (m^3), being inferred not big differences in growth rate across the trials within age-zone. A summary of the results can be seen in Table 3.2.

Table 3.2. Summary of field test growth measurement per age and breeding zone. HT = Total tree height (m), DBH measured at 1.3 meters above ground (cm). Each parameter was reported with its respective standard deviation (SD).

Zone	Age	Trial (N)	Survival ± SD (%)	HT ± SD	DBH ± SD	Volume ± SD
Arauco	4	4	99 ± 1	11.1 ± 1	11 ± 1.12	0.0459 ± 0.0127
Arauco	5	4	96 ± 2	14.23 ± 0.62	13.76 ± 0.52	0.0957 ± 0.0066
Arauco	8	4	86 ± 10	20.07 ± 1.35	19.05 ± 1.19	0.2646 ± 0.0529
Valdivia	4	4	87 ± 5	9.72 ± 0.92	10.67 ± 0.61	0.0384 ± 0.0052
Valdivia	5	7	80 ± 9	12.35 ± 2.24	13.03 ± 1.94	0.0791 ± 0.0315
Valdivia	8	4	74 ± 5	20.68 ± 1.26	17.75 ± 0.46	0.2371 ± 0.0129

The mean volume per tree and survival of the hybrid was compared with the pure species controls established in each site, measured at 8 years within zones. In this comparison, depicted in Figure 3.3.A, the hybrid survival was lower in both zones than parental species. In general, overall survival was higher in Arauco than Valdivia, and among varieties, NIT had the highest survival in both zones, followed by GLO and finally the hybrid.

The individual tree volume (Figure 3.3.B) of the hybrid (GloNi) was reasonably constant between zones, with a volume of 0.26 and 0.24 m^3/tree in Arauco and Valdivia, respectively, a difference of only 0.02 m^3/tree. The control species showed much bigger drops, with NIT decreasing 0.10 m^3 from Arauco to Valdivia (0.45 to 0.35) and GLO decreasing 0.06 m^3 (0.21 to 0.15) from Arauco to Valdivia.

NIT had better volume performance in both zones than GLO. It is interesting to see that in the Arauco zone, the hybrid grows better than GLO but much more similar to GLO than to NIT; in contrast, in Valdivia, the hybrid grows roughly similar to the mid-parent value.

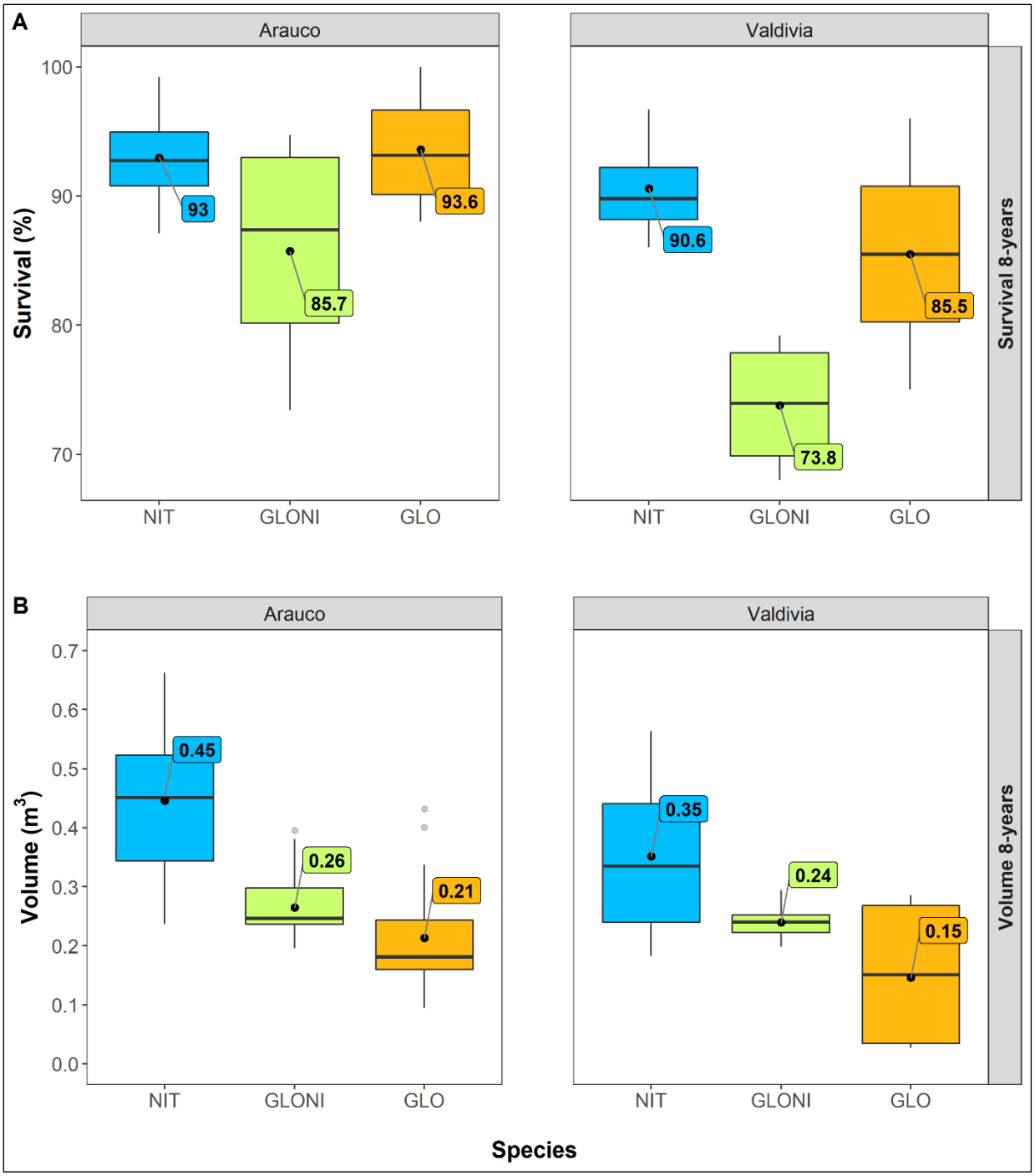


Figure 3.3. Parental and hybrid phenotypic values in survival (%) and individual tree volume (m^3), measured at 8-years on Arauco and Valdivia breeding zones. **A:** Survival plots. **B:** Volume plots. The abbreviation NIT is for *E. nitens* parent, GLO for *E. globulus* parent, and GLONI for the hybrid between these species. NIT and GLO survival and volume values were obtained from the pure species control installed of each hybrid field test. Black dots represent the mean of survival and volume, with their value in a label near to the dot for each species within zone.

3.3.2 Wood Properties

The number of wood samples analyzed was about 6 ramets per clone for 385 selected clones in Arauco and about 5 ramets per clone for 325 clones in Valdivia. The results indicate that hybrid wood properties for pulp production were slightly better in the Arauco zone than in Valdivia, with higher pulp yield (PY) and lower specific consumption (SC) of the clones (Table 3.3). Basic density (BD) was the only trait where the hybrid had a better value in Valdivia than Arauco.

Table 3.3. Summary of wood properties measurements for a clonal population of *E. nitens x E. globulus* at 6-years. BD=Basic density (kg/m^3), PY= Pulp Yield (%) and SC=Specific Consumption (m^3/ADt), N Clones= Number of clones evaluated.

Zone	BD ± SD	PY ± SD	SC ± SD	N ramets sampled	N Clones
Arauco	448.8 ± 20.8	50.5 ± 1.1	4.4 ± 0.2	6.0	385
Valdivia	451.7 ± 21.9	49.5 ± 1.2	4.5 ± 0.2	4.6	325

The mean of the wood property traits for the GloNi hybrid and the two parental species are compared in Figure 3.4. First, as was seen with the hybrids, the pure species has slightly superior wood properties in the Arauco zone than in Valdivia, with higher values for BD and PY and lower values for SC in both *E. nitens* and *E. globulus*. Secondly, for all three wood property traits in both zones, the GloNi hybrid clones had a mean near the mid-parent value of the two parent species. In the Valdivia zone, the mean BD of the hybrid was a bit closer to the *E. globulus* mean, and the mean PY was a bit closer to the *E. nitens* mean. Hence, the functional product of the two traits, SC for the hybrid, was almost exactly intermediate to the parent species.

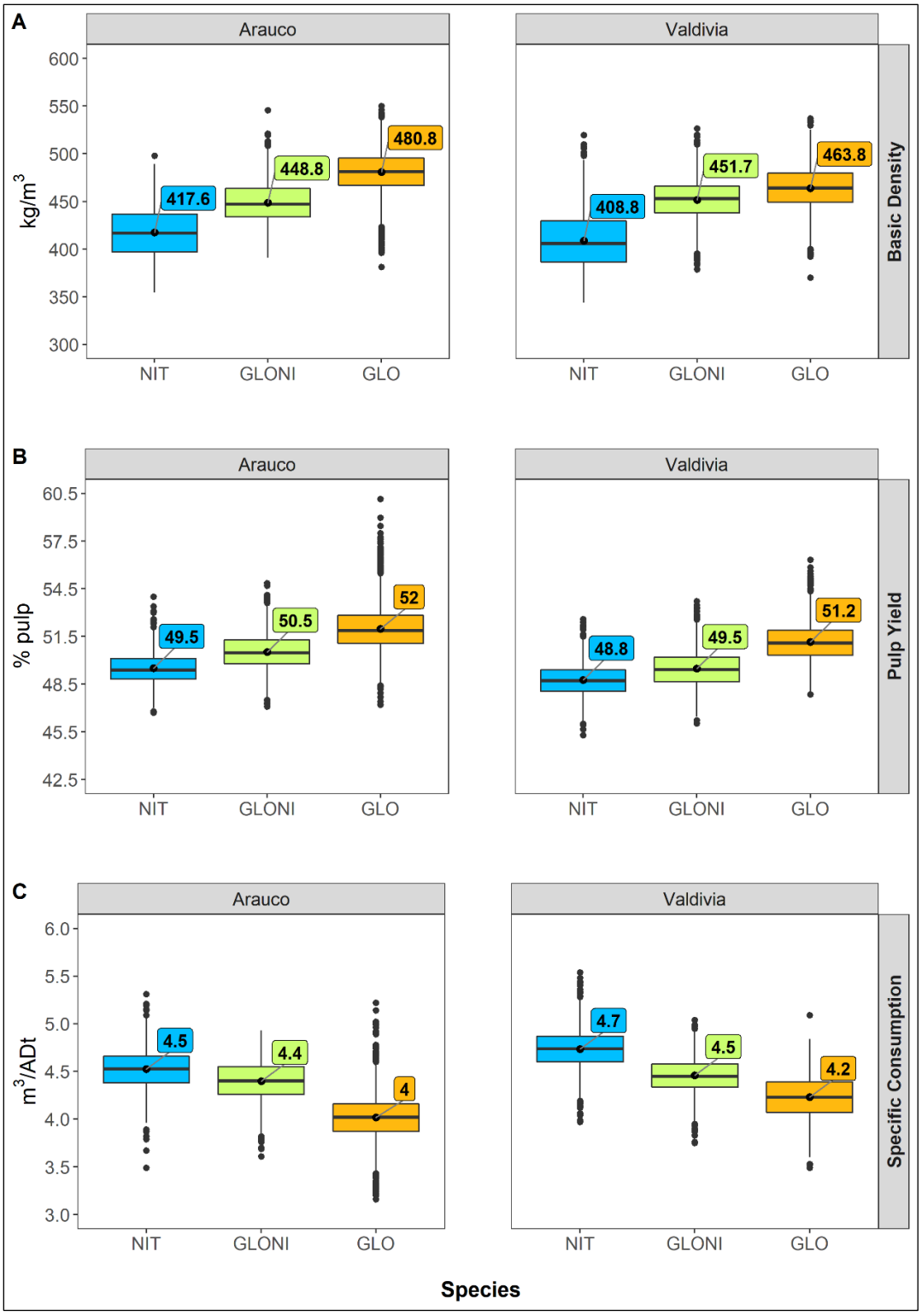


Figure 3.4. Parental and hybrid wood properties phenotypic values in Arauco and Valdivia breeding zones. **A:** Basic density (kg/m^3). **B:** Pulp yield (%) plot. **C:** Specific consumption (m^3/ADt). The abbreviation NIT is for *E. nitens* parent, GLO for *E. globulus* parent, and GLONI for the hybrid between these species. NIT and GLO wood property values were obtained from the company's current pure species breeding programs (around 5-8 tests for each zone).

3.3.3 Genetic Parameters for Individual Tree Volume

The multi-site analysis revealed a high broad-sense heritability for the standardized tree volume (stVol) for the GloNi hybrid, with an estimated $H^2 > 0.50$ in both zones (Table 3.4). This high H^2 value was associated with low levels of Genotype × Environment (GxE) variance for all levels of genetic effects, as indicated by high estimates of type B genetic correlations. Type B correlations for GHA_{NIT}, GHA_{GLO}, SHA, clone within family, and total genetic variance ranged from rB = 0.89 to 0.96.

Table 3.4. Genetic parameters of combined site analysis for standardized volume (stVol), Basic density (BD), Specific Consumption (SC), and Pulp Yield (PY) in Arauco and Valdivia zone with their respective SE. GHA_{NIT} or GHA_{GLO} is the General Hybridizing Ability variance due to *E. nitens* female or *E. globulus* male. SHA is the Specific Hybridizing Ability variance. Clw is the clonal variance within family. G is the total genetic variance. All these genetic variances were expressed in sigma values (σ). H^2 is the broad-sense heritability. rB_{NIT}, rB_{GLO}, rB_{SHA}, $rB_{Clw,}$ and rB_G are the type-B genetic correlations for nitens, globulus, SHA, Clw, and G x site interaction, respectively.

Zone	Traits	σ values ± SE					Genetics Parameters ± SE					
		GHA_{NIT}	GHA_{GLO}	SHA	Clw	G	H^2	rB_{NIT}	rB_{GLO}	rB_{SHA}	rB_{Clw}	rB_G
Arauco	stvol	21.06 *± 5.07*	-	-	39.89 *± 0.96*	45.1 *± 2.51*	0.54 *± 0.03*	0.92 ± 0.05	-	-	0.92 *± 0.01*	0.92 *± 0.01*
	BD	4.5 *± 1.74*	6.76 *± 2.7*	-	7.91 *± 0.57*	11.34 *± 1.79*	0.36 *± 0.07*	0.80 *± 0.22*	1 *± 0*	-	0.93 *± 0.08*	0.93 *± 0.05*
	SC	0.04 *± 0.02*	0.08 *± 0.03*	0.04 *± 0.02*	0.10 *± 0.01*	0.14 *± 0.02*	0.53 *± 0.07*	1 *± 0.18*	0.99 *± 0.03*	1 *± 0*	0.99 *± 0.04*	0.99 *± 0.03*
	PY	0.29 *± 0.16*	0.53 *± 0.24*	0.32 *± 0.13*	0.59 *± 0.03*	0.90 *± 0.15*	0.60 *± 0.08*	1 *± 0*	0.93 *± 0.09*	1 *± 0*	0.88 *± 0.05*	0.92 *± 0.04*
Valdivia	stvol	27.37 *± 6.87*	-	-	42.75 *± 1.06*	50.76 *± 3.81*	0.51 *± 0.04*	0.96 *± 0.03*	-	-	0.89 *± 0.01*	0.89 *± 0.02*
	BD	3.59 *± 1.5*	3.56 *± 1.97*	1.81 *± 1.68*	5.91 *± 0.65*	7.99 *± 1.17*	0.22 *± 0.05*	1 *± 0*	1 *± 0*	0.38 *± 0.48*	0.87 *± 0.16*	0.86 *± 0.1*
	SC	0.05 *± 0.02*	0.04 *± 0.02*	0.02 *± 0.02*	0.07 *± 0.01*	0.1 *± 0.01*	0.36 *± 0.06*	0.86 *± 0.18*	1 *± 0*	0.64 *± 0.73*	0.84 *± 0.09*	0.86 *± 0.06*
	PY	0.43 *± 0.13*	-	0.14 *± 0.12*	0.54 *± 0.03*	0.7 *± 0.08*	0.55 *± 0.06*	0.95 *± 0.07*	-	0.81 *± 0.44*	0.94 *± 0.04*	0.94 *± 0.03*

Table 3.4 presents measurements of genetic variation in genetic standard deviations, i.e., sigma values ($\sigma = \sqrt{\sigma^2}$), to express the range of the genetic effects present in the population in the measurement units for each trait evaluated (stVol, BD, SC, PY).

In both zones, the estimated GHA_{GLO} effect for stVol was zero (Table 3.4). The estimate of $\hat{\sigma}_{GHA_{GLO}}$ = 0 means that the *E. globulus* fathers had no consistent and detectable effect on the volume gain of the hybrid clones. In contrast, there was an enormous impact of the *E. nitens* mothers on the volume gain of the hybrid clones, with $\hat{\sigma}_{GHA_{NIT}}$ = 21.06% in the Arauco zone, and 27.37% in Valdivia, implying that if a large population of *E. nitens* females was tested as hybrid parents, the range of GHA_{NIT} values would be roughly from ±42% in Arauco, and ±55% in Valdivia. For growth, SHA variation was also zero in both zones. However, clonal variation within full-sib hybrid families was very significant, with $\hat{\sigma}_{Clw}$= 39.89% in the Arauco zone, and 42.75% in Valdivia, indicating the possibility to find clones within a full-sib family with a genetic worth up to 80% above the family mean. Considering the two genetic effects with non-zero variation ($\hat{\sigma}_{GHA_{NIT}}$ and $\hat{\sigma}_{Clw}$), total genetic variation for stVol among hybrid clones appears to be very large, with $\hat{\sigma}_{G}$ = 45.1% in the Arauco zone, and 50.8% in Valdivia.

3.3.4 Genetic Parameters for Wood Properties

Moderate to high broad-sense heritability values were observed for all wood properties in both zones, with H^2 ranging from 0.22 to 0.60 (Table 3.4). The highest H^2 values were obtained for the trait of Pulp Yield (PY), with an H^2 = 0.60 in Arauco and H^2 = 0.55 in Valdivia. For specific consumption (SC) trait, estimated heritability was H^2 = 0.53 in Arauco, and H^2 = 0.36 in Valdivia. Basic density showed the lowest level of genetic control among the wood traits, with H^2 = 0.36 and 0.22 in Arauco and Valdivia, respectively; these heritabilities were even lower than the heritability observed for volume in both zones. Also of interest was the fact that the H^2 of all wood traits was higher in the Arauco zone than the Valdivia zone, though this difference was larger for BD and SC than for PY. Similarly, the clone within family ($\hat{\sigma}_{Clw}$) and total genetic effect ($\hat{\sigma}_{G}$) have higher level of variation in Arauco than Valdivia.

In all wood traits, there is a valuable GHA variance from both parent species, with the only exception of the trait PY in Valdivia, where the GHA variance of *E. globulus* was $\hat{\sigma}_{GHA}$= 0. For all other wood traits in both regions, $\hat{\sigma}_{GHA_{GLO}}$ was important, in contrast to volume,

where $\hat{\sigma}_{GHA_{GLO}}$ was zero in both regions. There was also an interesting relationship of the GHA variance for wood traits in the two zones. In the Arauco zone, the $\hat{\sigma}_{GHA_{GLO}}$ was more important than $\hat{\sigma}_{GHA_{NIT}}$ for all wood traits, while in the Valdivia zone, the opposite was observed and $\hat{\sigma}_{GHA_{NIT}}$ was higher than $\hat{\sigma}_{GHA_{GLO}}$. For example, for BD in Arauco, $\hat{\sigma}_{GHA_{GLO}}$ = 6.76 and $\hat{\sigma}_{GHA_{NIT}}$ = 4.50, a difference of 2.25 kg/m^3 of globulus over nitens. In contrast, for BD in Valdivia, the GHA variance of the two species was almost the same, with $\hat{\sigma}_{GHA_{NIT}}$ = 3.59 and $\hat{\sigma}_{GHA_{GLO}}$ = 3.56 kg/m^3 (Table 3.4).

There appeared to be SHA variation for most wood traits in both zones. In the Arauco zone, there was a zero estimate for $\hat{\sigma}_{GHA_{GLO}}$ in BD, and a low SHA variation compared with the GHA variation for SC, while for PY, SHA variation was substantial ($\hat{\sigma}_{SHA}$ = 0.14 vs $\hat{\sigma}_{GHA_{NIT}}$ = 0.43 and $\hat{\sigma}_{GHA_{GLO}}$ = 0). In Valdivia, SHA variance appeared to be substantially less important than GHA variation for the two species for all wood traits (BD, SC, and PY).

The clone-within-family effect ($\hat{\sigma}_{Clw}$) was the most important source of genetic variation for all wood traits in both zones. In every case, $\hat{\sigma}_{Clw}$ was greater than $\hat{\sigma}_{GHA_{NIT}}$, $\hat{\sigma}_{GHA_{GLO}}$, and $\hat{\sigma}_{SHA}$ (Table 3.4).

There was little evidence of any genotype x environment interaction for wood traits in either zone. Aggregating across all sources of genetic variation, the type-B genetic correlations ranged from rBg = 0.92 to 0.99 in in Arauco, and rBg = 0.86 to 0.94 in Valdivia.

3.3.5 Epistasis

There was substantial clone-within-family variation ($\hat{\sigma}_{Clw}$) for volume gain observed in both zones, implying a significant amount of epistasis for this trait. Appling Foster's equation (Eq. 3.15) an estimate of epistasis variance was calculated for all traits in both zones (Table 3.5).

Table 3.5. Variance components of multi-site analysis for standardized volume (stVol), Basic density (BD), Specific Consumption (SC), and Pulp Yield (PY) in Arauco and Valdivia zone. $\hat{\sigma}^2_{GHA_{NIT}}$ or $\hat{\sigma}^2_{GHA_{GLO}}$ is the General Hybridizing Ability variance due to *E. nitens* female or *E. globulus* male, $\hat{\sigma}^2_{SHA}$ is the Specific Hybridizing Ability variance. $\hat{\sigma}^2_{Clw}$ is the clone-within-family variance. $\hat{\sigma}^2_G$ is the total genetic variance, $\hat{\sigma}^2_i$ is the estimate of epistasis variance.

Zone	Trait	$\hat{\sigma}^2_{GHA_{NIT}}$	$\hat{\sigma}^2_{GHA_{GLO}}$	$\hat{\sigma}^2_{SHA}$	$\hat{\sigma}^2_{Clw}$	$\hat{\sigma}^2_G$	$\hat{\sigma}^2_i$	$\hat{\sigma}^2_i/\sigma^2_G$
Arauco	stVol	443.3 ± 213.4	-	-	1590.9 ± 76	2034.2 ± 226.5	1147.6 ± 226.5	0.56 ± 0.16
	BD	20.3 ± 15.6	45.7 ± 36.5	-	62.6 ± 9.0	128.6 ± 40.5	-3.4 ± 41	-0.03 ± 0.31
	SC	0.002 ± 0.002	0.006 ± 0.005	0.002 ± 0.002	0.009 ± 0.001	0.019 ± 0.006	-0.003 ± 0.007	-0.17 ± 0.33
	PY	0.09 ± 0.09	0.28 ± 0.25	0.1 ± 0.09	0.34 ± 0.04	0.81 ± 0.26	-0.32 ± 0.33	-0.4 ± 0.31
Valdivia	stVol	748.7 ± 375.8	-	-	1827.3 ± 90.2	2576 ± 386.4	1078.5 ± 386.5	0.42 ± 0.21
	BD	12.9 ± 10.8	12.6 ± 14	3.3 ± 6.1	35 ± 7.7	63.8 ± 18.7	-0.4 ± 22.9	-0.01 ± 0.36
	SC	0.003 ± 0.002	0.001 ± 0.002	0 ± 0.001	0.005 ± 0.001	0.009 ± 0.002	0 ± 0.003	-0.01 ± 0.31
	PY	0.18 ± 0.11	-	0.02 ± 0.04	0.29 ± 0.03	0.5 ± 0.11	0.05 ± 0.15	0.09 ± 0.31

For the three wood traits in both zones, the estimates of epistasis variances ($\hat{\sigma}^2_i$) were negative or near, which can be interpreted as a null or minimal epistasis effect since these values are not significantly different from zero. In contrast, the estimate of epistasis variance for volume is relatively high in both zones, and the ratio of $\hat{\sigma}^2_i / \hat{\sigma}^2_G$ is 0.419 in Arauco and 0.564 in Valdivia, with both of these values significantly different from zero. In other words, epistasis variance accounts for 41% to 56% of the total genetic variance for volume growth among GloNi clones.

3.3.6 Hybrid GHA vs. Pure Species GCA for Volume Gain

Comparisons of hybrid GHA value and pure species GCA value for volume gain were only possible for the *E. nitens* parents since there was no variation for GHA_{GLO} in both breeding zones (Table 3.4). The relationship between *E. nitens* GCA (x-axis) and GHA (y-axis) was plotted and is shown in Figure 3.5. In Figure 3.5.A, a scatter plot is shown, with the GHA value for each hybrid mother plotted against pure species GCA. In Figure 3.5.B, a box-and-

whisker plot shows the range of total genetic values for all clones from a given hybrid mother, plotted against pure species GCA.

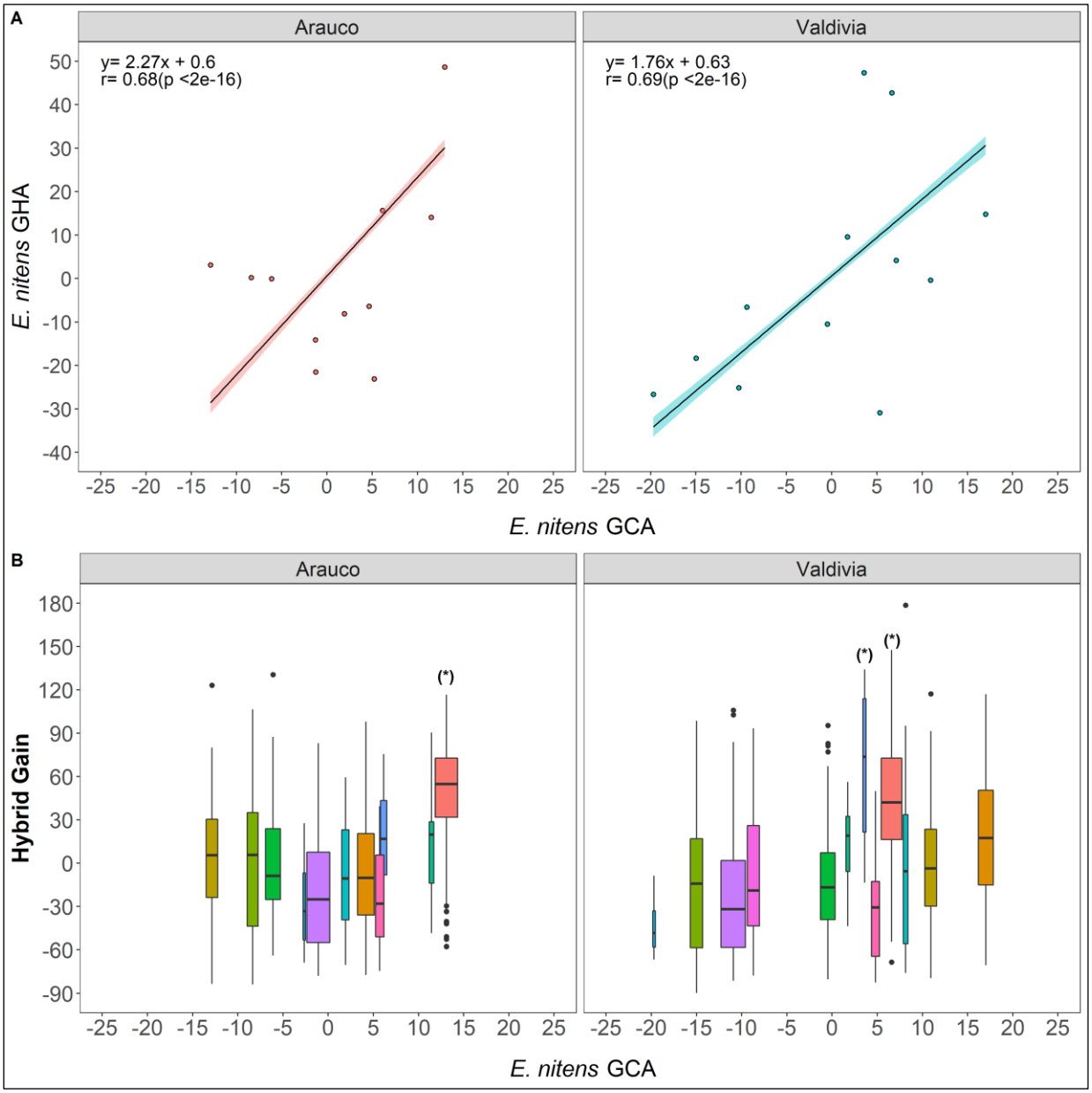


Figure 3.5. Relationship between *E. nitens* GCA and their effect in Hybrid Gain and *E. nitens* GHA for standardized volume (stVol) in Arauco and Valdivia zones. **A:** GHA vs. GCA of nitens parents, where each dot represents a nitens mother, and the significance of the correlations are in parenthesis (p-value). **B:** Hybrid Gain vs. *E. nitens* GCA, where each nitens parent was represented with a unique color for both breeding zones Identified with an asterisk (*) *E. nitens* mothers with clones that on average obtained 50% gain relative to the hybrid population mean (zero).

Comparing the GHA (y-axis) to the parental GCA (x-axis) in Figure 3.5.A, there is a clear positive and moderately strong relationship (r_{HP}= GHA-GCA correlation), with correlations of r_{HP} = 0.68 in the Arauco zone and r_{HP} = 0.69 in Valdivia, with both values significant and different from zero ($p < 2e\text{-}16$). These high correlations suggest that the parental GCA values are good estimators of the hybrid GHA for volume in both breeding zones. The pure species GCA values were calculated with an approach similar to the one used for the hybrids, and the GCA values are expressed in units of percent gain above the pure species population mean. The regression coefficients, i.e., the slope of GHA vs. GCA, were greater than 1 in both zones, indicating that the growth gains observed in pure species *E. nitens* will, in general, be multiplied when extended to the hybrid. For example, in Arauco, the slope was 2.27, so an *E. nitens* parent with a GCA of 10% would be expected to have a GHA value of around 22% in the Arauco zone. Similarly, in Valdivia, the slope was 1.76, so an *E. nitens* parent with a GCA of 10% in the Valdivia zone would be expected to have a GHA value of around 18%.

Examining the box-and-whisker plots in Figure 3.5.B, a positive association can be seen between *E. nitens* GCA and Hybrid Gain, where higher GCA values for *E. nitens* correspond to higher values of the hybrid gain in volume due to the important GHA variance of this parent in both zones. It is clear that some *E. nitens* mothers produced better clones than others, where some mothers produced clones that, on average, performed around 50% better than the population mean in both zones (indicated with an asterisk symbol in Figure 3.5.B). Furthermore, each *E. nitens* mother showed high clonal variability in volume (standardized volume), consistent with the large clone within family variation that was observed. The data indicate that, in general, we can expect to find clones ranging from ±40% gain in volume relative to the mean of each *E. nitens* mother.

3.3.7 Hybrid GHA vs. Pure Species GCA for Wood Properties

The GHA for the wood properties BD, SC, and PY were compared with the parental GCA of both species, *E. nitens* and *E. globulus,* with a series of scatterplots and regressions (Figure 3.6).

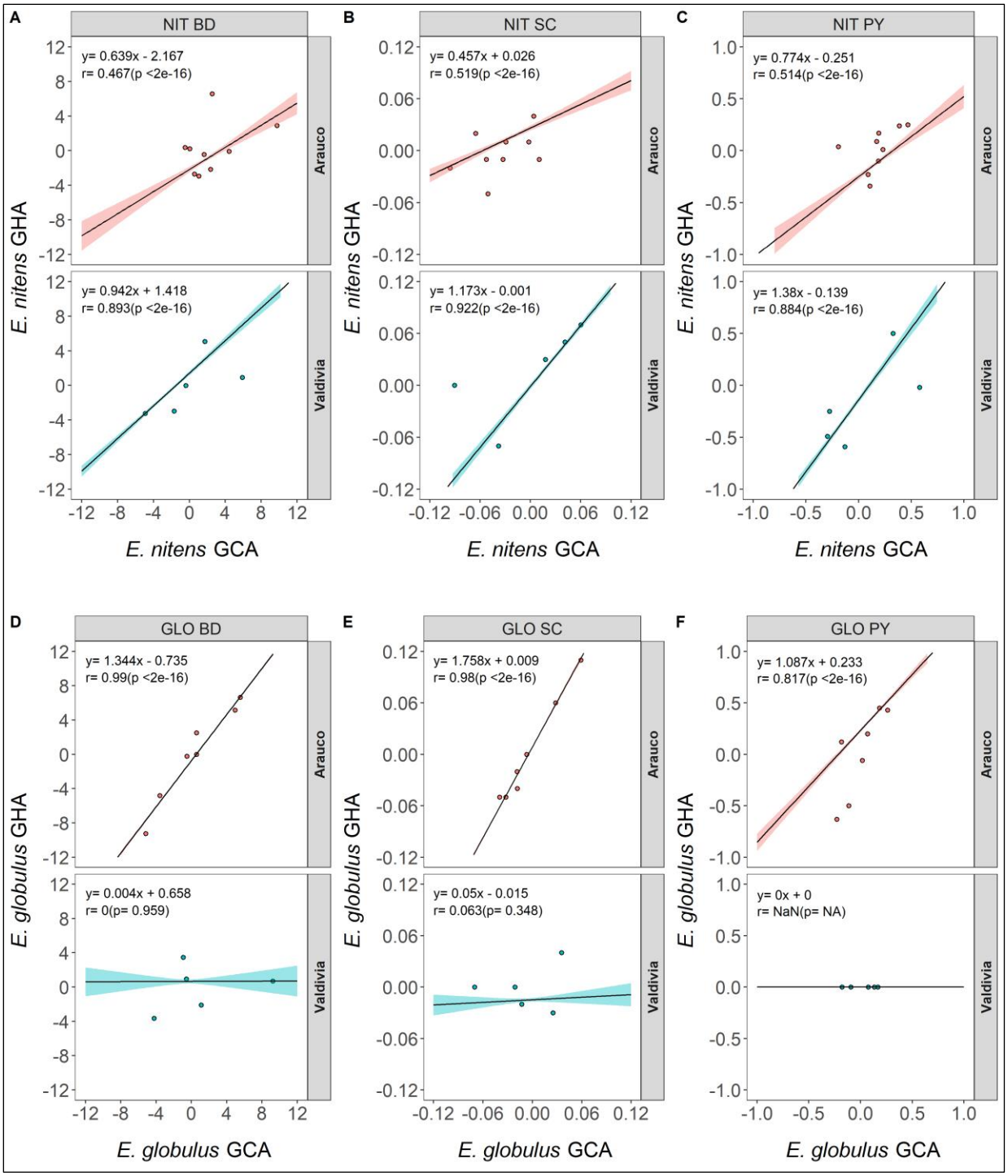


Figure 3.6. GHA vs GCA on wood properties of *E. nitens* and *E. globulus* parents in Arauco and Valdivia breeding zones. Each plot has the species and trait analyzed in the title, being NIT=*E. nitens* and GLO= *E. globulus* parent, BD=Basic Density, SC= Specific Consumption, and PY= Pulp Yield traits. In parenthesis, the p-value of each correlation per trait-species. The globulus PY relationship between GHA and GHA in Valdivia was not possible to plotted, since the globulus GHA variance of PY was zero in this zone.

For *E. nitens* parents, the correlations between parental GCA and GHA (r_{HP}= correlation GHA-GCA) values were very high in Valdivia, across all the wood traits: r_{HP} = 0.89 for BD, r_{HP} = 0.92 for SC, and r_{HP} = 0.88 for PY (Figure 3.6. A, B and C), and all correlations were strongly significant. In contrast, the correlations in the Arauco breeding zone were positive, but with moderate size: r_{HP} = 0.47 for BD, r_{HP} = 0.52 for SC and r_{HP} = 0.51 for PY.

Despite the lower values, all the correlations values were also statistically significant, with p < 2e-16. Looking at correlations between GHA and GCA for the *E. globulus* parents, there were strong correlations between GHA_{GLO} with GCA_{GLO} for all wood properties in the Arauco zone, with r_{HP} = 0.99 for BD, r_{HP} = 0.98 for SC, and r_{HP} = 0.82 for PY, with all values significantly different from zero at p < 2e-16 (Figure 3.6). In Valdivia, the correlations between GHA_{GLO} and GCA_{GLO} were essentially zero for BD and SC. For the trait PY, there was no observed variation for GHA_{GLO}, so it was not possible to calculate a correlation with GCA_{GLO}.

3.4 Discussion

3.4.1 Survival and Hybrid Performance in Arauco and Valdivia

Although the annual temperature profiles of Arauco and Valdivia (Figure 3.2) are reasonably similar, Valdivia is cooler, with a minimum mean monthly temperature (MMT) around 2.6 °C lower across the year and a difference near 3 °C in the coldest month of the year, July. Moreover, Valdivia is a more frost-prone region with 10.2 frost days in winter per year, compared to 3.1 in Arauco (see Table 3.1). The temperature difference between zones is probably a primary factor in the differences in survival observed in the two zones. Age 8-year survival of both species and the hybrid were lower in Valdivia than Arauco; however, the decrease was larger for the hybrid than for the pure species: 8.1% lower survival for *E. globulus* and 2.4% lower for *E. nitens,* compared to 11.9% lower for GloNi (Figure 3.3). Tibbits et al. (1991) reviewed and studied the performance of a number of eucalyptus hybrids and concluded that, on average, F1 hybrids tend to be intermediate to the parent species for frost tolerance with a slight tendency toward the more frost susceptible species.

These authors also observed this to be true for the *E. nitens x E. globulus* hybrid in particular, with sporadic statistically significant deviation toward the less tolerant *E. globulus*, depending on the trait and the time of year. Potts et al. (2000) comment on high levels of abnormal phenotypes (dwarfs) and mortality in *E. nitens x E. globulus* nursery and field experiments in Australia. In the current clonal trials in Chile, it seems likely that many abnormal phenotypes would be culled during the rooting phase of the selection process. However, the lower survival at 8-years could still result from some level of incompatibility for this hybrid that becomes more apparent with age and under stress from competition, frost, and other factors. Notwithstanding that, the best 20% clones in volume gain (BLUP) had an average survival of 89% and 90% in Arauco and Valdivia, respectively; this is higher than the overall average survival of the hybrid observed at 8 years in Figure 3.3, especially in Valdivia, where the difference is around a 16% more survival from the selected clones than the average survival in the zone.

This climatic difference also has an impact on the individual tree volume of the pure species and the hybrid in the two zones. At 8 years of age, all three varieties have larger individual tree volume in the Arauco zone than in Valdivia; however, the magnitude of the difference varies considerably. Comparing individual tree volume at 8 years between Arauco and Valdivia (Figure 3.3), *E. nitens* grows some 22.2% less in Valdivia than Arauco. The *E. globulus* grows 29.6% less volume in Valdivia and shows much more variation in growth rate, perhaps indicating a slightly lower degree of adaptation to the zone. However, the GloNi hybrid also grows less in Valdivia but has a decrease of only 7.7% compared to the growth in Arauco. This phenomenon is partly due to the higher mortality (lower survival) observed in Valdivia. However, even accounting for this, it seems that the hybrid tree volume is less affected by Valdivia's environment than Arauco in general.

In this study, the GloNi hybrid had wood properties intermediate to the parent species in both zones, being consistent with the conclusions of various authors indicating that on average, *Eucalyptus* hybrids tend to have intermediate values for wood density and other wood property traits (de Assis, 2000; Potts et al., 2000; Zobel & Jett, 2012).

3.4.2 Genetic parameters

The genetic variances in these clonal populations of GloNi were, in general, high for volume, pulp yield, specific consumption, and moderate for basic density. Broad sense heritabilities ranged from H^2 = 0.22 to 0.60 for wood traits (Basic density, Pulp yield, and Specific consumption) and were above 0.50 for volume (Table 3.4). All traits in the Arauco breeding zone had higher H^2 estimates than in Valdivia, and this was also reflected in a lower level of GxE observed in Arauco. At the clonal level, rB_G values for all traits ranged from rB_G = 0.92 to 0.99 in Arauco and were typically lower in Valdivia, ranging from rB_G = 0.86 to 0.94 in Valdivia (Table 3.4). In general, the low levels of GxE suggest that in the future, more clones can be tested for growth across fewer sites, and for wood properties, samples could also be taken in fewer trails.

Volker *et al*. (2008), working with *E. nitens x E. globulus* trials in Australia, estimated narrow-sense heritability for the growth traits DBH, and the wood property trait Pilodyn penetration. Pilodyn is a useful indirect measurement of wood basic density (Greaves et al., 1996; Wu et al., 2010). Volker *et al*. (2008) reported a DBH heritability of h^2 = 0.42 and a Pilodyn heritability of h^2 = 0.20, roughly comparable to the estimate of $H^2 \approx 0.50$ for growth and H^2 ranging from 0.22 to 0.36 for basic density observed in this study. The estimated H^2 for basic density in the hybrid population in this study was lower than might have been expected, based on published heritability estimates for the parental species. Raymond (2002) reported a summary of wood genetic parameters for the species *E. globulus* and *E. nitens* based on a decade of publications and reported a mean h^2 estimate for basic density around 0.60 for *E. nitens* (based on 12 publications) and a mean h^2 estimate around 0.70 for *E. globulus* (based on 7 publications). For pulp yield, the mean h^2 estimates for both *E. nitens* and *E. globulus* was around 0.40 (based on 7 and 5 publications, respectively), and these values correspond more closely to the H^2 = 0.55 and 0.60 in the two zones in this study.

For the trait of volume, there was a strong effect of the *E. nitens* parent on the growth performance of the hybrid in both breeding zones. In addition, there was a substantial clonal variation found within full-sib GloNi families, and the data suggests that it should be

possible to find clones ranging from ±80% in volume gain, relative to the mean of each family. There was no SHA variance found for these interspecific crosses for volume, and there was no GHA variance found for volume among *E. globulus* parents.

It is important to remember that there were only 8 *E. globulus* fathers evaluated in this study. Simulation results suggest a low (but non-zero) chance of estimating zero variance for a random effect when the number of parents is low and the true underlying genetic variance is low (see Chapter 2). So even if true GHA_{GLO} variance ($\hat{\sigma}^2_{GHA_{GLO}}$) is not zero, it is likely that $\hat{\sigma}^2_{GHA_{GLO}}$ is relatively low and much less important than the other effects for hybrid tree volume. Other authors have estimated zero GHA variance in *Eucalyptus* hybrids. In a study with a large hybrid population of *Eucalyptus grandis x E. urophylla* progeny seedling propagation, van den Berg et al. (2015) reported a broad-sense heritability of $H^2 = 0.37$, but found that the GHA variance for the 30 *E. grandis* parents to be zero. There was substantial GHA variance for the 27 *E. urophylla* parents, but there was also a considerable non-additive SHA variance, indicating a high amount of dominance genetic variance. In the current study, there was no SHA variance detected for volume in either zone. However, there does appear to be a substantial amount of non-additive epistasis variation observed in both zones.

The estimates of epistasis ($\hat{\sigma}^2_i$) for volume was large in both zones, making up 42% to 50% of the total genetic variation. In contrast, there was no evidence of epistasis for any of the wood traits in either zone. These results correspond to the ones reported by Tan et al. (2018) in an extensive study of the progeny of 476 full-sib hybrid families of *E. urophylla x E. grandis*. In that study, hybrid families were derived from 86 *E. urophylla* and 95 *E. grandis* parents and represented by 35 individuals each. The hybrids were tested in a randomized complete block design in single-tree plots and 35 replicates in one trial. Using 41,304 SNP markers, genomic models were evaluated that accounted for additive, dominance, and first-order epistatic interactions for two growth traits (Circumference at Breast Height (CBH) and Height) and two wood traits (basic density and pulp yield) evaluated at 3 and 6 years-old. The study results showed significant epistasis variation in height and CBH at 3-years, with the epistasis variance comprising 91% and 65% of the total genetic variance,

respectively. In the measurement at 6-years, the epistasis variance was zero for height but still accounted for 36% of the total genetic variance in CBH. Similar to the current GloNi study results, Tan et al. found no epistasis variation for any wood traits at either age.

By applying the Foster and Shaw's epistasis variance estimation on the variances components reported for van den Berg et al. (2015) for a sizeable clonal population of *Eucalyptus grandis x E. urophylla*, discussed above, for tree volume, the epistasis variance was estimated to be 40% of the total genetic variation, a value roughly similar to the estimates found in this study for the same trait. It seems that the direction of the cross for *E. urophylla x E. grandis* does not affect the amount of epistatic variation, comparing the result of van den Berg et al. with Tan et al., both showing high epistasis variation for growth in these hybrid populations.

At the pure species level, Costa E Silva et al. (2004) estimated the effect of epistasis in a study with full-sib families and clonally-replicated progeny of *E. globulus* in Portugal. Epistasis variation was estimated following the derivation of Foster & Shaw (1988) for DBH growth and Pilodyn penetration measured at 4 years of age. That study reported a very low amount of epistasis variance for DBH, accounting for only 3% of the total genetic variation. In contrast, for Pilodyn penetration, a substantial epistasis effect was reported, corresponding to 24% of the total genetic variance.

In another study of *E. globulus* in Portugal, Araújo et al. (2012) partitioned additive and non-additive variation for DBH growth in a clonal population, testing more than 4,200 genotypes in 40 sites. These authors reported a small amount of epistasis variance $\hat{\sigma}_i^2$ = -0.02, not significantly different than zero (using the approach of Foster & Shaw (1988)). They did report some important non-additive genetic variation in this eucalyptus clonal population, but this was dominance variation ($\hat{\sigma}_d^2$ = 0.096) of a similar size to the additive variance ($\hat{\sigma}_a^2$ = 0.096), followed for the clonal within family variation ($\hat{\sigma}_{Clw}^2$= 0.055)

Overall, the studies of Costa E Silva et al. (2004) and Araújo et al. (2012) are in accord with a low to zero epistasis variance for the DBH growth trait in *E. globulus*. For *E. nitens*, there were no studies that characterize epistasis.

Isik et al. (2003), studying clonally replicated progeny tests with loblolly pine (*Pinus taeda*), composed of 9-full sib families, partitioned genetic variance into additive, dominance, and epistatic components, and found a negative epistasis variance for growth traits (Total height, DBH, and volume), which was interpreted as zero variance. These authors also mention that the epistasis variance estimates obtained with Eq. 3.15 is an approximation of the real value and could be underestimated under the Foster and Shaw's methods since their model assumes that the epistasis came mostly from a high-level loci interaction. Although, if low-level loci interaction occurs, the additive variance could be slightly overestimated, and similarly, dominance variance could be slightly overestimated. However, the Foster and Shaw method should generally give a very good and straightforward approximation of the epistasis variation.

3.4.3 Pure species GCA – Hybrid GHA correlations

Positive correlations of moderate size were found for *E. nitens* pure species GCA and hybrid GHA (r_{HP}) for volume in both the Arauco and Valdivia zones, with correlation values of r_{HP} = 0.68 and 0.69, respectively. These correlations were significantly different from zero (Figure 3.5), and this suggests that pure species genetic value for growth traits will be a good indicator of genetic worth as a hybrid parent. Since there was zero GHA variance found for *E. globulus*, r_{HP} for this species is undefined.

There was a relationship between *E. nitens* GCA and GHA in both the Arauco and Valdivia zones for wood property traits, but with much higher correlations found in Valdivia (r_{HP} = 0.88 to 0.92) than in Arauco (r_{HP} = 0.47 to 0.52). For *E. globulus*, there were quite strong correlations between GCA and GHA for all three wood traits in the Arauco zone, ranging from r_{HP} = 0.81 to 0.99. However, in Valdivia, the correlations were near zero for BD and SC, and were non-estimable for PY since the *E. globulus* GHA was zero for this trait.

Volker et al. (2008), mentioned previously, stated that there was *"no reliable quantitative genetic method of predicting which parents should be used"* to produce the best hybrid families of *E. nitens x E. globulus*, which would seem to imply that there are low GCA-GHA

correlations. However, for *E. nitens*, they reported a correlation for GCA-GHA of r_{HP} = 0.67 for 6-year-old DBH and r_{HP} = 0.65 for 10-year-old DBH. In contrast, for *E. globulus*, they reported a GCA-GHA correlation of r_{HP} = 0.16 for 6-year-old DBH and a negative correlation for 10-year-old DBH. For Pilodyn penetration (analogous to basic density (BD) in the current study), GCA-GHA correlations of r_{HP} = 0.60 and 0.65 were found for *E. globulus* and *E. nitens*, respectively. In all cases, the standard errors of the correlation estimates were high, but in general, those results correspond well to the current study results. Thus, it appears that *E. nitens* GCA values are moderate predictors of the hybrid GHA for both growth and wood properties, while for *E. globulus*, only for wood traits are GCA values related to GHA values.

For a different hybrid, *E. grandis x E. urophylla*, van den Berg et al. (2017) found a statistically significant correlation of r_{HP} = 0.58 between GCA and GHA for *E. urophylla* for DBH. As there was very little GHA variance for *E. grandis*, the GCA-GHA correlation for *E. grandis* was not reported. These authors concluded that individual tree breeding values for growth traits would be relatively good indicators of GHA. However, they also noted a large amount of non-additive genetic variation for DBH in this hybrid. Nevertheless, there would be some value in selecting the best pure species *E. urophylla* parents for growth to test as hybrid parents in a hybrid breeding program, similar to the case for *E. nitens* parents and the GloNi hybrid variety in Chile.

Finally, it is crucial to highlight that the current population of GloNi does not have many NIT and GLO parents to date, and relatively few crosses per parents (approximately 2 per *E. nitens*, and 4 per *E. globulus*), so the GCA-GHA correlations have to be viewed with some caution. Nevertheless, for tree volume, the data suggest a clear tendency with *E. nitens* for high GCA values to be associated with high GHA in both breeding zones (Arauco and Valdivia). For wood properties in Arauco, *E. globulus* GCA in Arauco is an excellent predictor of the GHA, and *E. nitens* GCA is a moderate GHA predictor. For wood properties in Valdivia, E. nitens GCA is an excellent GHA predictor.

3.4.4 Impact of Environment on Hybrid Genetic Architecture

An interesting pattern emerged in this study where there was a clear relationship between the environment (i.e., the Arauco and Valdivia zones) and the expression of genetic variances related to the *E. nitens* or *E. globulus* parentage of the GloNi hybrid. Comparing the two parental pure species, it is clear that *E. nitens* should be better adapted to the Valdivia zone than Arauco, and the reverse is true for *E. globulus*. The importance of the *E. nitens*-related genetic parameters was more important in Valdivia than in Arauco, while the opposite was true for the *E. globulus*-related parameters.

Regarding *E. nitens*, the GHA variance was higher in Valdivia than in Arauco for volume and the wood traits SC and BD. The correlation between *E. nitens* GCA and GHA for volume was almost exactly the same in Valdivia and Arauco (r_{HP} = 0.69 vs. 0.68). However, for the three wood traits, this correlation was much higher in Valdivia than in Arauco (r_{HP} = 0.88 to 0.92 in Valdivia, and r_{HP} = 0.46 to 0.51 in Arauco). Concerning *E. globulus*, GHA variance was higher in Arauco than in Valdivia for all three wood traits. The correlation between *E. globulus* GCA and GHA for wood traits was very high in Arauco (r_{HP} = 0.82 to 0.99) and near-zero in Valdivia (r_{HP} = 0.00 to 0.06).

It is conceivable that in a hybrid tree variety, where one species brings adaptability to specific environmental conditions and stresses, more of the genetic variation in hybrid performance could derive from that species relative to the other parental species. He et al. (2012) examined genetic variation in *E. urophylla x E. tereticornis* in a cool frost-prone environment where *E. tereticornis* would be expected to bring frost tolerance to the hybrid. These authors found higher (and statistically significant) GHA variance for 4-year volume in the *E. tereticornis* parents than the *E. urophylla* parents. Also, they found that genetic variation in "cold hardiness" (i.e., field assessed cold and frost damage) derived only from the *E. tereticornis* parents.

Trials of *E. grandis x E. tereticornis* and *E. grandis x E. camaldulensis* hybrids were planted on four sites in Zimbabwe that were considered marginal for *E. grandis* due to low rainfall (Madhibha et al., 2013), and where the *E. tereticornis* and *E. camaldulensis* parents were

intended to bring drought tolerance to the hybrid. There was no GHA variance for 43-month height and DBH among the *E. grandis* parents in either hybrid combination in these four tests. In contrast, there was significant GHA variance among the *E. tereticornis* parents for both height and DBH, and among the *E. camaldulensis* parents for DBH.

The above examples involve growth traits and seem consistent with the pattern observed in this study for volume, where *E. nitens* (GHA) contributes more genetic variation for volume in the cooler zone Valdivia than in the warmer zone of Arauco. However, the current results appear to be the first observation of this kind of pattern with wood traits, where *E. globulus* parents explain more of the hybrid performance in the warmer zone (Arauco), and *E. nitens* parents explain more of the hybrid performance in the cooler zone (Valdivia).

This last observation is somewhat surprising, as generally, there is a low level of GxE reported for wood properties in forest trees, and specifically, this has been found to be true for pure species *E. globulus* and *E. nitens*. For *E. globulus,* no GxE effect was found for wood properties analyzed in a study conducted by Raymond et al. (2001) in Tasmania, Australia, for basic density and pulp yield, or in a recent study by Nickolas et al. (2020) for basic density and Kraft pulp yield also in Tasmania. In *E. nitens*, there was no significant GxE in wood properties in multiple studies in different environments in Victoria and Tasmania, Australia. In those studies, Greaves et al. (1996) evaluated Pilodyn trait for indirect measurement of wood density, Blackburn et al. (2014) used acoustic wave velocity for indirect selection of trees related to MOE, and Hamilton et al. (2009) evaluated wood density and cellulose content, among other traits. As pure species, both *E. nitens* and *E. globulus* present very stable behavior across sites in wood properties evaluations.

3.4.5 Implications for Crossing Strategy for F1 GloNi Clone Production

This study provides some guidance for the formulation of crossing strategies to identify new GloNi clones. All traits examined show considerable genetic variation for clones within family, and volume appears to have a large amount of epistatic variation. These results first suggest that it is essential to test large numbers of clones, as this will be the only way to

capture these potential genetic gains. However, the selection of specific *E. nitens* and *E. globulus* parents can also provide some gains.

First, an important strategy to improve volume in both breeding zones would be to increase the number of *E. nitens* parents used in the crossing design, as there is important GHA variation due to *E. nitens* in both zones. Thus, a sizable increment of gain could be achieved by testing more *E. nitens* females in hybrid crosses, and *E. nitens* parents could be selected based on their performance as pure species due to the moderately high correlation observed between GCA_{NIT} and GHA_{NIT} in both breeding zones. Moreover, it appears that increments of GCA_{NIT} might result in larger increments of GHA_{NIT} in the hybrid, as the regression coefficient suggests a multiplier of 2.3 and 1.8 for Arauco and Valdivia, respectively. For example, an *E. nitens* parent with a GCA = 10% could be expected to have a GHA = 23% in Arauco and 18% in Valdivia.

GHA variation for wood properties was also due to *E. nitens* in both zones, and GCA-GHA correlations were moderate (Arauco) to high (Valdivia), so some emphasis could also be placed on wood properties when selecting *E. nitens* parents.

There was substantial GHA variation for wood properties due to *E. globulus* in both zones and a very high GCA-GHA correlation in the Arauco zone. Similarly to the discussion above, there may be some scale effect where increments of GCA_{GLO} for wood traits might result in larger increments of GHA_{GLO} in the hybrid: the regression coefficients suggest a multiplier of 1.34 for BD and 1.76 for SC in Arauco (Figure 3.6 D and E, respectively). Since *E. globulus* does not contribute GHA variation to hybrid performance in volume, it seems likely that breeders can ignore pure species volume and focus only on wood properties when selecting *E. globulus* parents for hybrid crosses. Possibly a breeder would want to be cautious with the interpretation of these results since a very small number of parents were tested, and therefore would not want to ignore pure species volume GCA completely. Nevertheless, even selecting the top half of the population for volume would allow substantial selection intensity for wood traits.

Finally, the low amount of GxE observed within hybrid zones (almost all type-B genetic correlations ranging from 0.80 to 1.00) indicates that clones should perform in essentially the same way across all sites within the Arauco and Valdivia breeding zones. In other words, the top 10 clones selected on one or a small number of sites should be excellent performers on any other site within the zone. Therefore, clones could be tested on relatively few sites within zones, allowing more hybrid families and clones to be included in the testing program.

In summary, to obtain gain in growth and wood properties for the Arauco zone, it is proposed to select the best *E. nitens* females with high performance in growth and the best *E. globulus* parents with good performance for wood properties, using the results of the parent-tested as pure species or hybrid crosses. In Valdivia, parental selection should focus more on the performance of *E. nitens* females for volume and wood properties.

3.5 Conclusions

Significant clonal differences were found among GloNi hybrids in volume gain in both breeding zones, indicating that huge genetic gains can be obtained. The NIT parents demonstrated a considerable impact on the volume of the hybrids, which makes it important to test more NIT parents in future interspecific crosses in Arauco and Valdivia zones. For unknown reasons, maybe ascribable to the low number of parents tested, the GHA of GLO parents were zero in both breeding zones. Under this result should be preferable to use fewer GLO fathers in future interspecific crosses or select these parents for their merits in other traits as wood properties or rooting abilities, something important in the clonal propagation of hybrid progeny. This study did not find an SHA effect of the crosses for volume, maybe related to the low number of parents tested.

The selection of NIT parents to improve volume gain in future interspecific crosses could be made based on pure species results or as hybrid parents. The positive relationship in tree volume for GHA and GCA in both breeding zones supports this strategy. However, this result

should be considered with some cautions. The small number of *E. nitens* and *E. globulus* parents tested (12 and 8 respectively) used to date in this hybrid program is expected to count with more parents tested in the future.

The very low genotype by environment interaction in both zones is a good indicator that selected clones within zones will perform similarly across sites, without ranking changes.

A strong environmental effect between zones was found in the expression of the wood properties in GloNi hybrids, indicating that neither *E. nitens* nor *E. globulus* parents have a consistent effect on the wood properties trait of the hybrid progeny among the zones. However, a significant improvement on wood properties can be made in both breeding zones. In Arauco, most of the selection should be made based on *E. globulus* performance. Parent selection for all wood traits can be made based on parent performance either in pure species or hybrid breeding programs. In Valdivia, the selection should be made based on the performance of both species in wood properties. Only *E. nitens* selection could be made based on GCA or GHA estimates, but in *E. globulus* should be made just as a hybrid parent.

In order to obtain gain in volume and wood properties, Arauco looks like a more complex scenario, where the parents' performance was different between these traits, which was not the case on Valdivia, where the best performance was in all traits for *E. nitens* parents. With the result obtained to date, in Arauco, it may be preferable to select the best NIT parents based on the growth performance and the best GLO parents based on their performance in the wood traits analyzed. In Valdivia, the strategy should be focused mainly on NIT performance by selecting known parents with good performance in growth and wood properties and *E. globulus* parents with good performance in Basic density and Specific Consumption.

3.6 References

Araújo, J. A., Borralho, N. M. G., & Dehon, G. (2012). The importance and type of non-additive genetic effects for growth in Eucalyptus globulus. Tree Genetics and Genomes, 8(2), 327–337. https://doi.org/10.1007/s11295-011-0443-x

Baltunis, B. S., Huber, D. A., White, T. L., Goldfarb, B., & Stelzer, H. E. (2005). Genetic effects of rooting loblolly pine stem cuttings from a partial diallel mating design. Canadian Journal of Forest Research, 35(5), 1098–1108. https://doi.org/10.1139/x05-038

BCN. (2021, March 10). Clima y Vegetacion Region del Bio Bio. Biblioteca Del Congreso Nacional de Chile / BCN. www.bcn.cl/siit/nuestropais/region8/clima.htm

Blackburn, D., Hamilton, M., Williams, D., Harwood, C., & Potts, B. (2014). Acoustic Wave Velocity as a Selection Trait in Eucalyptus nitens. Forests, 5(4), 744–762. https://doi.org/10.3390/f5040744

Butler, D. (2019). asreml: Fits the Linear Mixed Model. In R package version 4.1.0.110. www.vsni.co.uk

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2017). ASReml-R Reference Manual Version 4. ASReml-R Reference Manual, 176. http://www.homepages.ed.ac.uk/iwhite/asreml/uop.

Burdon, R. D. (1977). Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genet.*, *26*(5–6), 168–175.

Chen, S., Weng, Q., Li, F., Li, M., Zhou, C., & Gan, S. (2018). Genetic parameters for growth and wood chemical properties in Eucalyptus urophylla × E. tereticornis hybrids. Annals of Forest Science, 75(1), 16. https://doi.org/10.1007/s13595-018-0694-x

Costa E Silva, J., Borralho, N. M. G., Araújo, J. A., Vaillancourt, R. E., & Potts, B. M. (2009). Genetic parameters for growth, wood density and pulp yield in Eucalyptus globulus. Tree Genetics and Genomes, 5(2), 291–305. https://doi.org/10.1007/s11295-008-0174-9

Comstock, R. E., Robinson, H. F., & Harvey, P. H. (1949). A Breeding Procedure Designed To Make Maximum Use of Both General and Specific Combining Ability 1 . *Agronomy Journal*, *41*(8), 360–367. https://doi.org/10.2134/agronj1949.00021962004100080006x

Costa E Silva, J., Borralho, N. M. G., & Potts, B. M. (2004). Additive and non-additive genetic parameters from clonally replicated and seedling progenies of Eucalyptus globulus. Theoretical and Applied Genetics, 108(6), 1113–1119. https://doi.org/10.1007/s00122-003-1524-5

CR2. (2019). Climatic Explorer. Center for Climate and Resilience Research CR2. https://explorador.cr2.cl/

de Assis, T. F. (2000). Production and Use of Eucalyptus Hybrids for Industrial Purposes. In Proceedings of QFRI/CRC-SPF Symposium, 9-14 April 2000, Noosa, Queensland, Australia (Issue April 2000).

Dieters, M. J., Nikles, D. G., Toon, P. G., & Pomroy, P. (1997). Genetic parameters for F1 hybrids of Pinus caribaea var. hondurensis with both Pinus oocarpa and Pinus tecunumanii. Canadian Journal of Forest Research, 27(7), 1024–1031. https://doi.org/10.1139/x97-053

Dieters, M. J., White, T. L., Littell, R. C., & Hodge, G. R. (1995). Application of approximate variances of variance components and their ratios in genetic tests. Theoretical and Applied Genetics, 91(1), 15–24. https://doi.org/10.1007/BF00220853

Falconer, D. . S., & Mackay, T. F. . (1996). Introduction to quantitative genetics (4th edn). Pearson. https://doi.org/10.1002/bimj.19620040211

Foster, G. S., & Shaw, D. V. (1988). Using clonal replicates to explore genetic variation in a perennial plant species. Theoretical and Applied Genetics, 76(5), 788–794. https://doi.org/10.1007/BF00303527

Gapare, W. J., & Musokonyi, C. (2002). Provenance performance and genetic parameter estimates for Pinus caribaea var. hondurensis planted at three sites in Zimbabwe.

Forest Genetics, 9(3), 183–189.

Gezan, S. A., de Carvalho, M. P., & Sherrill, J. (2017). Statistical methods to explore genotype-by-environment interaction for loblolly pine clonal trials. Tree Genetics & Genomes, 13(1), 1. https://doi.org/10.1007/s11295-016-1081-0

Greaves, B. ., Borralho, N. M. ., Raymond, C. A., & Farrington, A. (1996). Use of a Pilodyn for the indirect selection of basic density in Eucalyptus nitens. Canadian Journal of Forest Research, 26(9), 1643–1650.

Griffin, A. R., Harbard, J. L., Centurion, C., & Santini, P. (2000). Breeding Eucalyptus grandis x globulus and other inter-specific hybrids with high inviability - problem analysis and experience with Shell Forestry projects in Uruguay and Chile. Hybrid Breeding and Genetics of Forest Trees, April, 1–13. ftp://ftp.aphis.usda.gov/foia/FOLDER_10/AR00035535 Griffin 2000 QFRICRC_SPF Symposium.pdf

Hamilton, M. G., Raymond, C. A., Harwood, C. E., & Potts, B. M. (2009). Genetic variation in Eucalyptus nitens pulpwood and wood shrinkage traits. Tree Genetics & Genomes, 5(2), 307–316. https://doi.org/10.1007/s11295-008-0179-4

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72(358), 320–338. https://doi.org/10.1080/01621459.1977.10480998

Hodge, G. R., & Dvorak, W. S. (2015). Provenance variation and within-provenance genetic parameters in Eucalyptus urophylla across 125 test sites in Brazil, Colombia, Mexico, South Africa and Venezuela. Tree Genetics and Genomes, 11(3). https://doi.org/10.1007/s11295-015-0889-3

Humphreys, J. R., O'Reilly-Wapstra, J. M., Harbard, J. L., Davies, N. W., Griffin, A. R., Jordan, G. J., & Potts, B. M. (2008). Discrimination between seedlings of Eucalyptus globulus, E. nitens and their F1hybrid using near-infrared reflectance spectroscopy and foliar oil content. Silvae Genetica, 57(4–5), 262–269. https://doi.org/10.1515/sg-2008-0040

INFOR. (2020). CHILEAN STATISTICAL YEARBOOK OF FORESTRY 2020. Statistical Bulletin N° 174. Instituto Forestal, Gobierno de Chile, Santiago, Chile, 274.

Isik, F., Li, B., & Frampton, J. (2003). Estimates of additive, dominance and epistatic genetic variances from a clonally replicated test of loblolly pine. Forest Science, 49(1), 77–88. https://doi.org/10.1093/forestscience/49.1.77

Ladrach, W. E. (1986). Comparaciones entre procedencias de siete coníferas en la Zona Andina al finalizar los ochos años.

Luan, Q. F., Li, Y. J., Jiang, J. M., Yue, H. F., & Diao, S. F. (2013). Discrimination of Pinus taeda × P. caribaea var. hondurensis between its allele-species and hybrids using near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 21(3), 203–211. https://doi.org/10.1255/jnirs.1051

Madhibha, T., Murepa, R., Musokonyi, C., & Gapare, W. (2013). Genetic parameter estimates for interspecific Eucalyptus hybrids and implications for hybrid breeding strategy. New Forests, 44(1), 63–84. https://doi.org/10.1007/s11056-011-9302-8

Mitchell, R. G., Wingfield, M. J., Hodge, G. R., Steenkamp, E. T., & Coutinho, T. A. (2013). The tolerance of Pinus patula × Pinus tecunumanii, and other pine hybrids, to Fusarium circinatum in greenhouse trials. New Forests, 44(3), 443–456. https://doi.org/10.1007/s11056-012-9355-3

Nickolas, H., Williams, D., Downes, G., Tilyard, P., Harrison, P. A., Vaillancourt, R. E., & Potts, B. (2020). Genetic correlations among pulpwood and solid-wood selection traits in Eucalyptus globulus. New Forests, 51(1), 137–158. https://doi.org/10.1007/s11056-019-09721-0

Patterson, H. D. and, & Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. Biometrika, 58(3), 545–554. https://www.jstor.org/stable/2334389

Potts, B. M., & Dungey, H. S. (2004). Interspecific hybridization of Eucalyptus : key issues for breeders and geneticists. Forest Research, 27(2), 115–138.

https://doi.org/10.1023/A:1025021324564

Potts, B. M., Volker, P. W., Tilyard, P. a, & Joyce, K. (2000). The genetics of hybridisation in the temperate Eucalyptus. In: Dungey H.S.D.,M.J. and Nikles D.G. (Eds), Hybrid Breeding and Genetics of Forest Trees. Proceedings of QFRI/CRC-SPF Symposium, 9–14th April 2000 Noosa, Queensland, Australia. Department of Primary Industries, Brisbane, April, 200–211.

Raymond, C. A. (2002). Genetics of Eucalyptus wood properties. Annals of Forest Science, 59(5–6), 525–531. https://doi.org/10.1051/forest:2002037

Raymond, C. A., Schimleck, L. R., Muneri, A., & Michell, A. J. (2001). Genetic parameters and genotype-by-environment interactions for pulp yield predicted using near infrared reflectance analysis and pulp productivity in Eucalyptus globulus. Forest Genetics, 8(3), 213–224.

Salas, C., Donoso, P. J., Vargas, R., Arriagada, C. A., Pedraza, R., & Soto, D. P. (2016). The Forest Sector in Chile: An Overview and Current Challenges. Journal of Forestry, 114(5), 562–571. https://doi.org/10.5849/jof.14-062

Stockwell, P., & Richter, F. I. (1947). Hybrid Forest Trees. Yearb. Agr. 1943– 1947, 465–472.

Tan, B., Grattapaglia, D., Wu, H. X., & Ingvarsson, P. K. (2018). Genomic relationships reveal significant dominance effects for growth in hybrid Eucalyptus. BioRxiv, 267(August 2017), 84–93. https://doi.org/10.1016/j.plantsci.2017.11.011

Team R. Core. (2019). R: A language and environment for statistical computing. In R Foundation for Statistical Computing. https://doi.org/10.1108/eb003648

Tibbits, W. N., Potts, B. M., & Savva, M. H. (1991). Inheritance of freezing resistance in interspecific F1 hybrids of Eucalyptus. Theoretical and Applied Genetics, 83(1), 126–135. https://doi.org/10.1007/BF00229235

van den Berg, G. J., Verryn, S. D., Chirwa, P. W., & Deventer, F. Van. (2015). Genetic Parameters of Interspecific Hybrids of Eucalyptus grandis and E. urophylla Seedlings

and Cuttings. Silvae Genetica, 64(1–6), 291–308. https://doi.org/10.1515/sg-2015-0027

Volker, P. W., Potts, B. M., & Borralho, N. M. G. (2008). Genetic parameters of intra- and inter-specific hybrids of Eucalyptus globulus and E. nitens. Tree Genetics and Genomes, 4(3), 445–460. https://doi.org/10.1007/s11295-007-0122-0

White, T. L., Adams, W. T., & Neale, D. B. (2007). Forest Genetics. In CABI Publishing. CABI Publishing. https://doi.org/10.1038/167764a0

Wu, S. jun, Xu, J. min, Li, G. you, Risto, V., Lu, Z. hua, Li, B. qi, & Wang, W. (2010). Use of the pilodyn for assessing wood properties in standing trees of Eucalyptus clones. Journal of Forestry Research, 21(1), 68–72. https://doi.org/10.1007/s11676-010-0011-5

Zhu, Y., Wu, S. J., Xu, J., Lu, Z., Li, G., Hu, Y., Yang, X., & Bush, D. (2017). Genetic parameters for growth traits and stem-straightness in Eucalyptus urophylla × E. camaldulensis hybrids from a reciprocal mating design. Euphytica, 213(7), 1–16. https://doi.org/10.1007/s10681-017-1923-3

Zobel, B. J., & Jett, J. B. (2012). Genetics of Wood Production. Springer Berlin Heidelberg. https://books.google.com/books?id=FRHpCAAAQBAJ

Chapter 4:

Quantitative Genetics in Hybrid Populations of *Pinus patula* x *Pinus tecunumanii*. Estimation of Genetic Parameters and Implication for Breeding Strategies

4.1 Introduction

Through the years, the development of new hybrid plant varieties has been common to increase the range of crop adaptability, increase resistance to diseases, and combine traits from parental species. An example in forest tree breeding might be the combination of a high growth rate from one parent and excellent wood properties from the other. In exceptional occasions, heterosis may be observed, which is a phenomenon where the hybrid progeny can perform better than the average of the parents (mid-parent heterosis), or performing even better than the best parent (high-parent heterosis) (White et al., 2007), and this would be very desirable in the development of a new commercial variety of trees.

In South Africa, *Pinus patula* is the most important pine species established, first introduced as early as 1907 and widely distributed across the country (Kanzler et al., 2014; Mitchell et al., 2013; Nöjd & Isango, 2003). *Pinus patula* is a temperate/subtropical species native to southern Mexico. The species is found primarily on deep and well-drained soils in its native environment, between altitudes of 1,490 and 3,100 m, and it exhibits a high growth rate and excellent wood quality, being cold tolerant and adapted to a range of climatic conditions changes (Dvorak, Hodge, Kietzka, et al., 2000; Hodge & Dvorak, 2012; Hongwane et al., 2017). In South Africa, in the mist belt regions at an elevation above 1,000 m, *P. patula* exhibits remarkable growth. In the south of the country, the species is usually planted in zones with mean annual temperatures below 17 °C (Kanzler et al., 2014). A limiting factor for its continued use in the country is its susceptibility to pitch canker fungus (*Fusarium circinatum*), which has become an established pathogen over the past two decades. South African breeders have tried to improve the tolerance of *P. patula* to *F. circinatum* via

greenhouse inoculation studies; however, these studies have not provided a level of tolerance that can increase field plantation survival (Hodge & Dvorak, 2012; Mitchell et al., 2013). Considering this issue, new *Pinus* species have been introduced in the last few decades, some of which offer an opportunity to improve tolerance to *F. circinatum* (Hongwane et al., 2017), including the species *Pinus tecunumanii*.

Pinus tecunumanii is a tropical/subtropical species found in Mexico and Central America. There are two recognized subpopulations: 1) high elevation (HE), where *P. tecunumanii* occurs above 1,500 masl (meters above sea level), and 2) low elevation (LE) *P. tecunumanii* that occurs below 1,500 masl. These two groups differ in several aspects. For example, *P. tecunumanii* (HE) is less tolerant to *F. circinatum*, has a lower wood density, and is more tolerant to cold environments. In South Africa, *P. tecunumanii* is replacing *P. patula* on some sites due to its rapid nursery growth, better drought tolerance, higher growth rate, wood uniformity, high wood density, and high tolerance to *F. circinatum* (Hodge & Dvorak, 2012; Kanzler et al., 2014).

The interspecific hybridization of *P. patula* with *P. tecunumanii* offers the potential to combine the well-known advantages of these species. Both *P. patula* and *P. tecunumanii* are closed-cone pines and compatible for the generation of a hybrid variety that could be better adapted to the environment in South Africa (Kanzler et al., 2012). The hybridization of these two species provides an opportunity to replace the commercial plantations of *P. patula* at the low and mid-elevations in order to decrease the high mortality caused by *F. circinatum*. In addition, considering that South Africa has very limited options to expand the commercial pine plantations due to limited rainfall in some areas, the hybrid between *P. patula* and *P. tecunumanii* could help to increase wood productivity in this finite growing area (Lopez et al., 2018).

In response to the issue of high mortality due to *F. circinatum* and limited land options for commercial plantations due to limited rainfall (Lopez et al., 2018), forestry companies in South Africa started to introduce new pine species, and they began hybridizations programs with interspecific crosses with *P. patula* in the 1990s (Kanzler et al., 2014). The goal was to improve the survival due to improved *Fusarium* tolerance of these hybrid varieties and

possibly improve growth as well. Two important forestry companies, South African Forestry Company (SAFCOL) and Sappi, initiated substantial hybrid programs for *P. patula x P. tecunumanii,* using both high-elevation and low-elevation varieties.

In SAFCOL, full-sib hybrid families were made between *P. patula* and the high- and low-elevation subpopulations of *P. tecunumanii* and established several seedling progeny tests. On the other hand, Sappi focused primarily on the hybrid of *P. patula* and low-elevation *P. tecunumanii* and included a clonal component to the testing program. In this manuscript, the interspecific crosses of *P. patula x P. tecunumanii* (high-elevation) will be referred to as PATxTECH, and crosses between *P. patula x P. tecunumanii* (low-elevation) will be referred to as PATxTECL.

The hybrid was established in field tests located in South Africa and Eswatini, with the primary goal of identifying the best hybrid parents for future commercial hybrid production, with selection planned to focus on important economic traits, including volume growth and wood properties. Modulus of Elasticity (MOE) is a measurement of wood stiffness (Gapare et al., 2009) and is used as an indicator of the wood strength. In this study, growth data and MOE measurements were used to estimate important genetic parameters using quantitative genetic methods and to predict genetic values, including General Hybridizing Ability (GHA) for both parent species and Specific Hybridizing Ability (SHA) for the specific full-sib crosses. The obtained GHA values were compared with analogous genetic values from pure species genetic trials within each company to evaluate if there is a relationship with the parental performance as pure species and evaluate possible implications for the hybrid breeding strategies of these varieties.

4.2 Material and Methods

4.2.1 Field Test Locations and Site Descriptions

Field tests of *P. patula x P. tecunumanii* hybrids were established in the north-east of South Africa and Eswatini (Figure 4.1), with the main objective of evaluating hybrid performance in growth, MOE, and survival for *a posteriori* selection of the best hybrid parents for future crosses for commercial hybrid production in SAFCOL and Sappi companies. SAFCOL established 5-field tests in the province of Mpumalanga, with several full-sib hybrid families. In Sappi, two field tests were established, one in South Africa, in the province of Mpumalanga, and the other in Eswatini, in the province of Manzini, with a full-sib hybrid progeny with clonal replication.

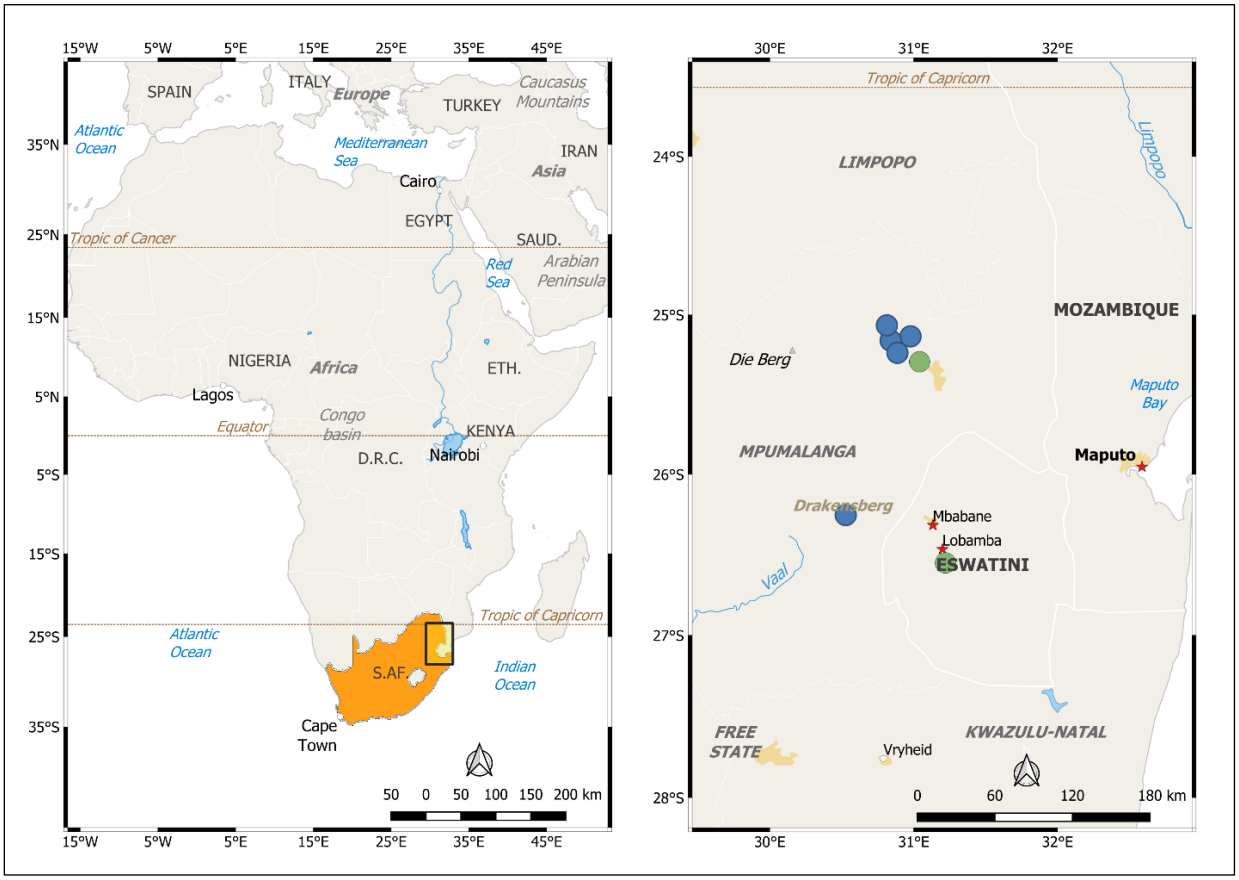


Figure 4.1. Map of *P. patula x P. tecunumanii* field tests locations. Left: global locations of the trials, with the country region depicted in orange and the study location in yellow. Right: specific location of each trial, illustrated in blue dots SAFCOL trials, green dots Sappi trials.

The trial locations are presented in Figure 4.1, where SAFCOL trials are depicted in blue dots and Sappi trials in greed dots.

Field tests at SAFCOL followed a randomized completed block design, with 6-replications per site, with an experimental plot of 1x6 trees per family, and with a spacing between plants of 3 m x 3 m, which corresponds to 1,666 trees/ha. These full-sib hybrid families were created using two sources of *P. tecunumanii*, low and high elevation, which came for different provenances, described in Appendix G. In this study, the two sources were differentiated by naming them PATxTECH (*P. patula x P. tecunumanii* high elevation) and PATxTECL (*P. patula x P. tecunumanii* low elevation). In each field test, the number of families per hybrid and the number of trees varied; details are shown in Table 4.1.

Table 4.1. Environmental parameters description of each trial in SAFCOL and Sappi, with the number of clones, families, and trees per hybrid variety.

Company	Location	Trial	Species	N Families	N Clone	N Trees	MAT °C	MAP (mm)	Altitude (masl)
SAFCOL	S. Africa Mpumalanga	SAF 1	PATxTECH	39	-	1,308	16 - 18	1,050-1,300	1,470
			PATxTECL	80	-	2,590			
	S. Africa Mpumalanga	SAF 2	PATxTECH	26	-	879	16 - 18	1,050-1,300	1,265
			PATxTECL	70	-	2,320			
	S. Africa Mpumalanga	SAF 3	PATxTECH	31	-	887	16 - 18	1,050-1,300	1,050
			PATxTECL	72	-	1,980			
	S. Africa Mpumalanga	SAF 4	PATxTECH	-	-	-	14 - 16	850-1,050	1,050
			PATxTECL	22	-	475			
	S. Africa Mpumalanga	SAF 5	PATxTECH	19	-	492	14 - 16	850-1,050	1,724
			PATxTECL	-	-	-			
Sappi	S. Africa Mpumalanga	SAP 1	PATxTECL	24	81	936	16.8	1,154	1,237
	Eswatini Manzini	SAP 2	PATxTECL	26	212	2,556	15.8	1,154	1,080

The SAFCOL sites SAF 2, SAF 3, and SAF 4 were established in 2008, site SAF 1 was established in 2002, and site SAF 5 in 2007.

The trials established by Sappi followed an Alpha-lattice experimental design (incomplete block design), with 3-replications per site and plots of 1x4 trees and an average of 3 treatments (clones) per replication. In those field tests, a total of 271 clones of PATxTECL were tested, representing 30 full-sib families, created by crossing 5 P. *patula* and 11 *P. tecunumanii* low parents. There was an average of 127 clones per full-sib family, with an

average of 12 ramets per clone. The two Sappi sites (SAP 1 and SAP 2) were established in April 2008.

4.2.2 Measurements of Growth and MOE

The measurements obtained in each trial were survival (percentage), diameter at breast high in cm (DBH), total height in meters (HT), forking presence in the main stem, incidence of pests, and presence of broken tops. Growth traits (DBH and HT) were recorded for all trees at 8-years of age in SAFCOL trials at 5-years of age in Sappi. The volume per tree was estimated using Ladrach's formula for juvenile trees (Ladrach, 1986; Ladrach & Mazuera, 1978) as follows:

$$Volume = 0.00003 * DBH^2 * HT \quad \text{Eq. 4.1}$$

where volume is expressed in m^3 per tree.

The acoustic velocity of the standing trees was measured in SAFCOL trials at 8 years of age, using the TreeSonic tool (Fakopp TreeSonic Timer, Hungary) in sites SAF 1 to SAF 4. This measurement was used to estimate the wood strength property Dynamic Modulus of Elasticity (MOE_d). The velocity data (measured in microseconds, μsec) was transformed to MOE_d values (measured in GPa) accordingly with the formula derived by Wielinga, Raymond, James, & Matheson (2009):

$$MOE_d = \delta V^2 x 10^{-3} \quad \text{Eq. 4.2}$$

Where V is the velocity of sound through the wood (km/s), δ is the wood density (kg/m^3) assumed as a constant value of 1000 kg/m^3 (Hongwane et al., 2017; Wielinga et al., 2009).

4.2.3 Volume Data Cleaning and Descriptive Parameters

For the accurate estimation of individual tree volume (m^3/ha), trees with abnormalities were removed from the analysis, including trees with the occurrence of pests, broken tops, strong stem sinuosity, more than two main stems, and dead trees. Additionally, trees with

an extreme DBH/HT relationship were removed (with values over 3 and below 0.3), assuming that these trees could not have reported broken tops or measurement errors.

For each company, the mean growth, survival rate, and wood properties for each hybrid species were estimated. Field tests with a survival rate lower than 70% were not included in multi-site analyses to avoid introducing additional error variation in the genetic parameter estimations. The data cleaning process was carried out using custom R scripts, and descriptive analysis per trait was carried out using the 'rstatix' package (Kassambara, 2021; Team R. Core., 2019).

4.2.4 Tree Volume Data Standardization

The volume (m^3) of each tree was standardized prior to estimation of genetic parameters to deal with scale-effect Genotype × Environment (GxE) interaction variance arising from differences in growth trait means reflecting different productivity across the sites (Hodge & Dvorak, 2015). Commonly, sites with high productivity produce larger trees with higher phenotypic variances than sites with less productivity (Gezan et al., 2017). In forestry, growth traits frequently have a strong relationship between the mean of the trait (DBH, HT, or volume) and its phenotypic and genetic variances (Gapare & Musokonyi, 2002; Hodge & Dvorak, 2015; van den Berg et al., 2015). To correct for a heterogeneous variance for volume across different test sites, the volume of each tree was standardized for each block within test, expressing the tree volume as a deviation from the block volume mean divided by the standard deviation: $y - \bar{y}/\sigma$, where "*y*" is the observed tree volume, $\bar{y}$ is the mean volume within block and σ is the standard deviation of the volume within block.

Coefficients of variance (CV) for volume were calculated per block, and then the average CV across all sites was calculated (CV_y). Finally, the standardized volume for each tree (i.e., each ramet of each clone in each test) was estimated as:

$$stVol = \frac{y-\bar{y}}{\sigma} * (CV_y * 100) + 100 \qquad \text{Eq. 4.3}$$

The standardized volume (stVol) is indicated in units of %, where the population's mean is centered on 100%, and the spread of the phenotypic data is expressed as CV_y*100. As a consequence, all variance component estimates are in units of percent and predicted genetic values could be interpreted in terms of gain (above or below 100%), without the necessity of rescaling (Hodge & Dvorak, 2015; van den Berg et al., 2015). Data standardization was conducted through an R script.

4.2.5 Estimation of Variance Components and Genetic Parameters

The phenotypic observations of MOE and standardized tree volume (stVol) were analyzed through restricted maximum likelihood (REML; Patterson & R., 1971 cited by Harville, 1977) using linear mixed models (LMM) for single-site and multi-site analysis. The single- and multi-site analyses were conducted using ASReml-R (Butler, 2019; Butler et al., 2017; Team R. Core., 2019) by following an LMM, which includes the effect of both parental species. The single-site linear model used for the evaluation of each clonal and seedling test was the following:

Seedling test:

$$y_{ijkl} = u + B_i + GHA_{PAT_j} + GHA_{TEC_k} + SHA_{jk} + BxSHA_{ijk} + e_{ijkl} \quad \text{Eq. 4.4}$$

Where:

y_{ijkl} is the l^{th} observation of the jk^{th} family in the i^{th} block; μ is the overall mean for the site; B_i is the fixed for the i^{th} block; GHA_{PAT_j} or GHA_{TEC_j} is the random General Hybridizing Ability (GHA) effect of the j^{th} female *P. patula* or k^{th} male *P. tecunumanii* parent; SHA_{jk} is the random Specific Hybridizing Ability (SHA) effect of the hybrid interaction between j^{th} female parent and the k^{th} male parent; $BxSHA_{ijk}$ is the random effect of the interaction between the i^{th} block, the j^{th} female and k^{th} male parent and e_{ijkl} is the random effect within the plot error term.

Clonal test:

$$y_{ijklm} = u + B_i + GHA_{PAT_j} + GHA_{TEC_k} + SHA_{jk} + Clw_l + BxClw_{il} + e_{ijklm} \quad \text{Eq. 4.5}$$

Where:

y_{ijklm} the m^{th} observation for the jk^{th} family for the l^{th} clone at the i^{th} block; Clw_l is the random clonal effect of the l^{th} hybrid clone; $BxClw_{il}$ is the random effect of the interaction between the i^{th} block and the l^{th} clone within a hybrid family and e_{ijklm} is the random effect within the plot error term. The other terms were described in the previous model (Eq. 4.4)

The term TEC_k in the GHA parameter in the models presented in equations 4.4 and 4.5 were changed to $TECH_l$ to specify *P. tecunumanii* high parent and $TECL_l$ to specify the *P. tecunumanii* low parent.

With the components of variance obtained from models described in Eq. 4.4 and Eq. 4.5, the phenotypic variance in the seedling tests ($\sigma^2_{phen\_s}$) was calculated in this way:

$$\sigma^2_{phen\_s} = \sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}} + \sigma^2_{SHA} + \sigma^2_{BxSHA} + \sigma^2_e \quad \text{Eq. 4.6}$$

The phenotypic variance in the clonal tests ($\sigma^2_{phen\_cl}$) was calculated as follows:

$$\sigma^2_{phen\_cl} = \sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}} + \sigma^2_{SHA} + \sigma^2_{Clw} + \sigma^2_{BxClw} + \sigma^2_e \quad \text{Eq. 4.7}$$

Single site broad-sense heritability (H^2_b) was estimated with formulas derived by Falconer (1996). For seedling trials:

$$H^2_b = \frac{2(\sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}}) + 4\sigma^2_{SHA}}{\sigma^2_{phen\_s}} \quad \text{Eq. 4.8}$$

and for clonal trials:

$$H^2_b = \frac{\sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}} + \sigma^2_{SHA} + \sigma^2_{Clw}}{\sigma^2_{phen\_cl}} \quad \text{Eq. 4.9}$$

Previous studies have sometimes called the single-site heritability a “biased heritability” (H^2_b), and considered the estimates from a combined multi-environment test analysis (H^2) an “unbiased heritability” (Dieters et al., 1995). These terms reflect the fact that single-site

genetic variances are typically larger than multiple-site genetic variances due to the inability to separate genotype and genotype x environment variances; therefore, these are "biased" upward.

After inspecting the estimates of H_b^2 from single sites, a multi-site analysis was performed. The statistical LMM model was similar to the single site (Eq. 4.4 and 4.5), with the difference that terms related to the site were added, along with their interactions. The statistical models for multi-site analysis for seedling and clonal trials are presented below:

Seedling test:

$$y_{ijklm} = u + S_i + B_{j(i)} + GHA_{PAT_k} + GHA_{TEC_l} + SHA_{kl} + SxGHA_{PAT_{ik}} + SxGHA_{TEC_{il}} + SxSHA_{ikl} + e_{ijklm} \quad \text{Eq. 4.10}$$

Where:

y_{ijklm} is the m^{th} observation for the kl^{th} family, for the j^{th} block at the i^{th} test site; u is the overall mean; S_i is the fixed of the i^{th} site; $B_{j(i)}$ is fixed effect of the j^{th} replication within the i^{th} site; GHA_{PAT_k} or GHA_{TEC_k} is the random General Hybridizing Ability (GHA) effect for the k^{th} female of *P. patula* or the l^{th} male of *P. tecunumanii* parent; SHA_{kl} is the random Specific Hybridization Ability (SHA) or full-sib family effect of the k^{th} and the l^{th} parents; $SxGHA_{PAT_{ik}}$ or $SxGHA_{TEC_{il}}$ is the random effect of the interaction between the i^{th} site with the k^{th} *P. patula* female parent or with the l^{th} *P. tecunumanii* male parent; $SxSHA_{ikl}$ is the random effect of the interaction between the i^{th} site and k^{th} female with l^{th} male and e_{ijklm} is the random effect within the plot error term.

Clonal test:

$$y_{ijklmn} = u + S_i + B_{j(i)} + GHA_{PAT_k} + GHA_{TEC_l} + SHA_{kl} + Clw_m + SxGHA_{PAT_{ik}} + SxGHA_{TEC_{il}} + SxSHA_{ikl} + SxClw_{im} + e_{ijklmn} \quad \text{Eq. 4.11}$$

Where:

y_{ijklmn} is the n^{th} observation of the j^{th} block for the kl^{th} family for the m^{th} clone at the i^{th} site; Clw_m is the random effect of the m^{th} clone within the hybrid family; $SxClw_{im}$ is the random

effect of the interaction between the *i*th site test and the *m*th hybrid clone and e_{ijklmn} is the random effect within the plot error term. The other terms were described previously in Eq. 4.10.

Equivalently, the terms TEC_k in the models presented on Eq. 4.10 and 4.11 were changed to $TECH_l$ or $TECL_l$ accordingly to *P. tecunumanii* low or high altitude.

The estimated effects in the models [Eq. 4.4, 4.5, 4.10, and 4.11] were assumed to be random and independently distributed, except for overall mean, site, and block, which were considered fixed effects (Madhibha et al., 2013; van den Berg et al., 2015).

The estimation of genetic parameters for the multi-site analysis was obtained following the derivation of Falconer & Mackay (1996) with the following formula:

Seedling phenotypic variance:

$$\sigma^2_{phen\_s} = \sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}} + \sigma^2_{SHA} + \sigma^2_{SxGHA_{PAT}} + \sigma^2_{SxGHA_{TEC}} + \sigma^2_{SxSHA} + \sigma^2_e \quad \text{Eq. 4.12}$$

Seedling total genetic variance: $\sigma^2_{Gs} = 2(\sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}}) + 4\sigma^2_{SHA}$ Eq. 4.13

Clonal phenotypic variance:

$$\sigma^2_{phen\_cl} = \sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}} + \sigma^2_{SHA} + \sigma^2_{Clw} + \sigma^2_{SxGHA_{PAT}} + \sigma^2_{SxGHA_{TEC}} + \sigma^2_{SxSHA} + \sigma^2_{Sxcl} + \sigma^2_e \quad \text{Eq. 4.14}$$

Total genetic variance: $\sigma^2_G = \sigma^2_{GHA_{PAT}} + \sigma^2_{GHA_{TEC}} + \sigma^2_{SHA} + \sigma^2_{Clw}$ Eq. 4.15

Broad-sense heritability: $H^2 = \sigma^2_G / \sigma^2_{phen}$ Eq. 4.16

where the term σ^2_G and σ^2_{phen} were changed for σ^2_{Gs} and $\sigma^2_{phen\_s}$ or $\sigma^2_{phen\_cl}$ accordingly.

In pure species breeding, the narrow-sense heritability (h^2) is often estimated as 4-times the General Combining Ability (GCA) variance divided by phenotypic variance. Sometimes, authors working with hybrid populations will estimate a "hybrid narrow-sense heritability" for each parental species by multiplying the parental GHA variance by 4 and then dividing by the total phenotypic variance (Dieters et al., 1997; Madhibha et al., 2013; Mitchell et al., 2013; van den Berg et al., 2015; Zhu et al., 2017). In this research, just the broad-sense

heritability will be reported (H^2), and the contribution of each parent to the total genetic variance can be seen as the GHA variance of each species compared with SHA variance and clone within family (Clw) variance where appropriate.

Type-B genetic correlations were obtained depending on the type of field test for clone, family, cross and parental level, estimated as:

Seedling total genetic: $rB_{Gs} = \frac{\sigma^2_{GHA_{PAT}}+\sigma^2_{GHA_{TEC}}+\sigma^2_{SHA}}{\sigma^2_{GHA_{PAT}}+\sigma^2_{GHA_{TEC}}+\sigma^2_{SHA}+\sigma^2_{SxGHA_{PAT}}+\sigma^2_{SxGHA_{TEC}}+\sigma^2_{SxSHA}}$ Eq. 4.17

Clone within family: $rB_{clw} = (\sigma^2_{Clw})/(\sigma^2_{Clw} + \sigma^2_{SxClw})$ Eq. 4.18

Total genetic : $rB_G = \frac{\sigma^2_{GHA_{PAT}}+\sigma^2_{GHA_{TEC}}+\sigma^2_{SHA}+\sigma^2_{Clw}}{\sigma^2_{GHA_{PAT}}+\sigma^2_{GHA_{TEC}}+\sigma^2_{SHA}+\sigma^2_{Clw}+\sigma^2_{SxGHA_{PAT}}+\sigma^2_{SxGHA_{TEC}}+\sigma^2_{SxSHA}+\sigma^2_{SxClw}}$ Eq. 4.19

Pat : $rB_{pat} = \sigma^2_{GHA_{PAT}}/(\sigma^2_{GHA_{PAT}} + \sigma^2_{SxGHA_{PAT}})$ Eq. 4.20

Tec : $rB_{tec} = \sigma^2_{GHA_{TEC}}/(\sigma^2_{GHA_{TEC}} + \sigma^2_{SxGHA_{TEC}})$ Eq. 4.21

SHA : $rB_{SHA} = \sigma^2_{SHA}/(\sigma^2_{SHA} + \sigma^2_{SxSHA})$ Eq. 4.22

Type-B genetic correlations, described by Burdon (1977), are measures of the degree of Genotype × Environment (GxE) interaction when the same trait is measured in trials established in different environments (often called multi-environment trial analysis, MET). The magnitude of GxE interaction ranges from 0 to 1, where values near 0 indicate weak agreement, and close to 1 a near-perfect correlation between the performance of our trait of interest of the genotypes across the different environments.

The H^2 is one of the essential genetic parameters estimates and expresses the degree to which the traits are under genetic control for the hybrid family or hybrid clone. Similarly, the rB_G in clonal tests and rB_{Gs} in seedling tests are important parameters to understand how consistently a hybrid clone or hybrid full-sib family will perform across environments.

Studies with clonal progeny from controlled crosses allow the estimation of additive and non-additive genetic variances and the breakdown of the non-additive variance into an estimate of dominance and epistasis variances. Foster & Shaw (1988) showed that with full-

sib clonal data, an approximation of the epistasis variance ($\hat{\sigma}_i^2$) can be calculated for pure species. Their methodology has been followed by many authors in studies of growth and wood properties in *Eucalyptus globulus* (Araújo et al., 2012; Costa E Silva et al., 2004, 2009) and in the estimation of the genetic parameters for rooting in loblolly pine (Baltunis et al., 2005). Adapting Foster's equation to the hybrid linear model used in this study, the epistasis effect was calculated for all trait in both zones as follow:

$$\hat{\sigma}_i^2 = \hat{\sigma}_{Clw}^2 - (\hat{\sigma}_{GHA_{PAT}}^2 + \hat{\sigma}_{GHA_{TECL}}^2) - 3\hat{\sigma}_{SHA}^2 \qquad \text{Eq. 4.23}$$

Where $\hat{\sigma}_{Clw}^2$ is the clone-within-family variance, $\hat{\sigma}_{GHA_{PAT}}^2$ and $\hat{\sigma}_{GHA_{TECL}}^2$ are the respective *P. patula* and *P. tecunumanii* low GHA variances and $\hat{\sigma}_{SHA}^2$ is the SHA variance.

4.2.6 Correlation of Parental Value for Pure Species Progeny and Hybrid Progeny Performance

The ability to pass favorable genes from a parent or characters to their progeny in the process of fertilization is referred to as *combining ability* (Bison et al., 2007; Kabir et al., 2014). The General Combining Ability (GCA) quantifies the average performance of a progeny related to a single parent compared to the population mean. GCA corresponds to half of a parent breeding value (BV) since a parent contributes, on average, one-half of its total additive genetic value to its progeny. (Fukatsu et al., 2014; Lynch & Walsh, 1998; White et al., 2007).

GCA information was available for some species and parents for volume and MOE from the pure species breeding programs of the two companies. From SAFCOL, GCA values were estimated for standardized volume with measurements taken at 8-years from pure species field tests of *P. patula* and high- and low-elevation *P. tecunumanii*. GCA values for MOE were estimated from acoustic velocity data (TreeSonic), with measurements ranging from 5 to 8 years, but only for *P. patula* populations. All quantitative analyses were a joint effort done by breeders from SAFCOL and Camcore (NC State University) (personal communication, G.R. Hodge). From Sappi, only from *P. patula* volume GCA information

was available, estimated from progeny tests of the company. The GCA values were predicted using standardized volume for measurements taken at 5-years, and this was done internally by the quantitative genetics team of Sappi.

All the GCA estimations in both traits and companies were performed through BLUP/REML methodology using a corresponding Linear Mixed Model.

The parental GCA values were compared with the GHA values using Pearson correlations to evaluate the consistency of the parental performance tested as pure species parents and as hybrid parents. A high correlation would indicate that parental GCA could be a good indicator for GHA performance. First, the Pearson correlation and their significance level were determined by calculating the t value. Then, the corresponding p-value was determined using the t distribution table for df=n−2 through the native 'stats' package of R software (Team R. Core., 2019) for volume and MOE.

4.3 Results

4.3.1 Survival and Single-Site Heritabilities

On average, the observed survival in SAFCOL field tests was 84% for PATxTECH and 82% for PATxTECL. In Sappi, the clonal tests of PATxTECL had a survival of 87%. Consequently, all the field tests were used in our combined-site analysis. Within-sites, differences in survival percentage were not observed in SAFCOL sites SAF 1, SAF 2, and SAF 3, in which PATxTECH and PATxTECL were tested together. However, PATxTECH exhibited a slightly high survival rate than PATxTECL. Hybrid survival percentage across trials can be inspected in Figure 4.2.

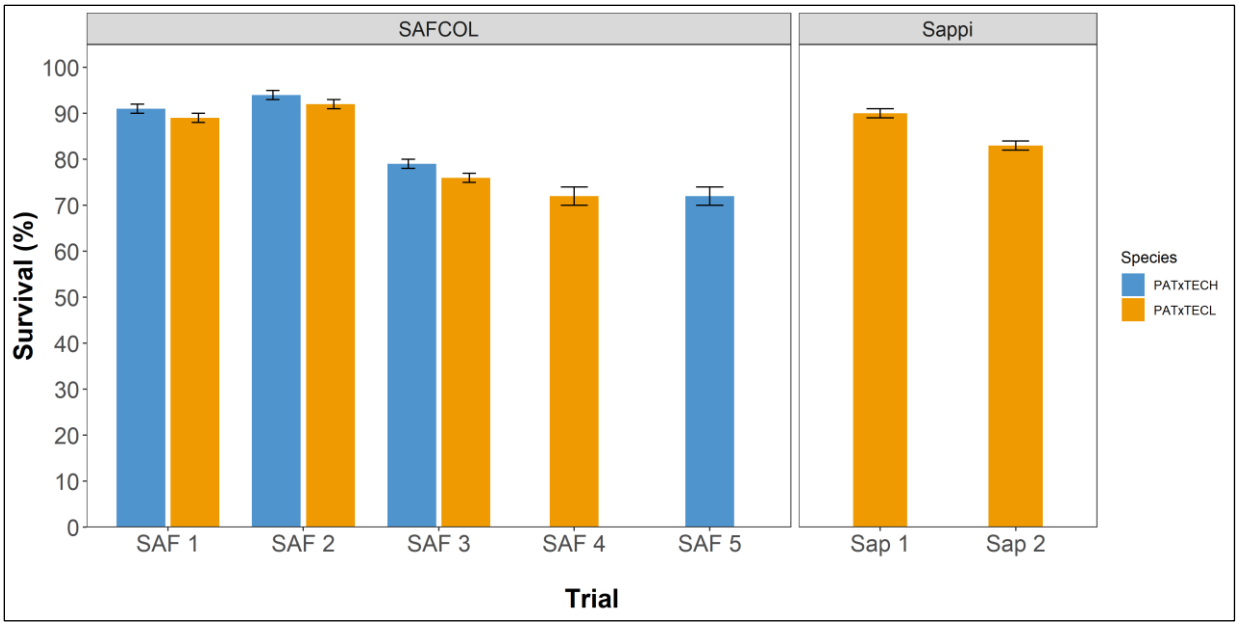


Figure 4.2. Survival rate with theirs SE (black bars) in each trial and company, separated by species. Depicted in blue, *P. patula x tecunumanii* high elevation (PATxTECH), and in orange *P. patula x tecunumanii* low elevation (PATxTECL).

The single-site heritability (H_b^2) for tree volume in SAFCOL trials ranged from H_b^2 = 0.24 to 0.36 in PATxTECL and from H_b^2 = 0.36 to 0.62 in PATxTECH. In Sappi, the H_b^2 of PATxTECL for the same trait ranged between H_b^2 = 0.24 and 0.27 (Figure 4.3.A), being more and less similar to the results in the SAFCOL trials. Measurements for MOE were available for four SAFCOL trials, and for this trait, single-site heritability ranged between from H_b^2 = 0.08 and 0.62 for PATxTECH and between H_b^2 = 0.11 and 0.46 for PATxTECL. It is interesting to see in both Figure 4.2 and Figure 4.3 that PATxTECH (blue bars) showed, in general, higher survival and higher tree volume heritability than PATxTECL.

There was some indication that H_b^2 for MOE was higher in PATxTECL than in PATxTECH hybrid, with higher heritability in 2 out of the 3-sites evaluated, and a very high value (H_b^2 = 0.70) observed in site SAF 4.

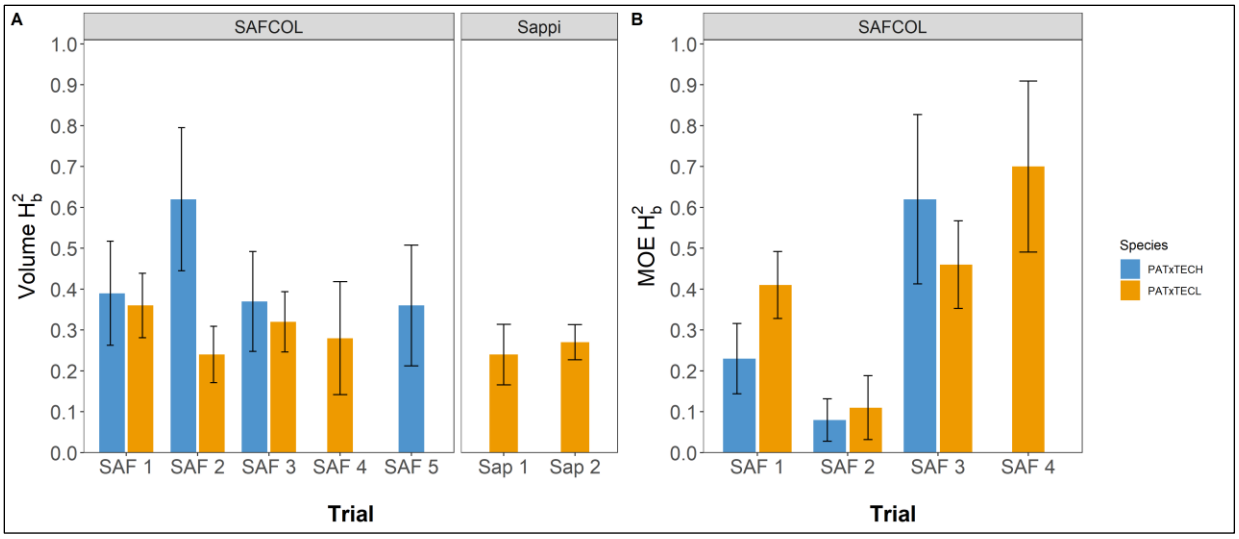


Figure 4.3. Volume and MOE broad-sense heritabilities (H^2_b) within sites, depicted in different colors the PATxTECH and PATxTECL hybrid (blue and orange, respectively). **A:** Volume H^2_b of SAFCOL and Sappi trials. **B:** MOE H^2_b in SAFCOL field tests. Black bars represent SE of broad-sense heritabilities of each species trait.

4.3.2 Volume and MOE

The individual tree volume (m^3) was consistently higher for PATxTECL than for PATxTECH across the seedling tests of SAFCOL (orange boxes, Figure 4.4.A). In those tests where both species were tested together (SAF 1, SAF 2, and SAF 3), pairwise comparisons using Tukey honest significance test (HSD) indicated statistically significant differences between that PATxTECH and PATxTECL ($p<0.001$), with PATxTECL yielding a higher volume. Site SAF 1 showed the lowest tree volume for both hybrid species, but still, the trait stVol had a H_b^2 estimate over 0.30. Site SAF 4 has the higher individual tree volume, perhaps somewhat related to the lower survival observed on that site (72% compared with a range of 79% to 93% on the other sites, see Figure 4.2).

In the Sappi clonal field tests, the volume of PATxTECL was significantly different between the sites ($p<0.001$), with a better performance in the site SAP 1. Observed tree volume in Sappi field tests was lower than the observed volume in SAFCOL. However, it is necessary to note that measurement ages differ between companies, with tree volume measured at 5-years of age for Sappi and 8-years for SAFCOL.

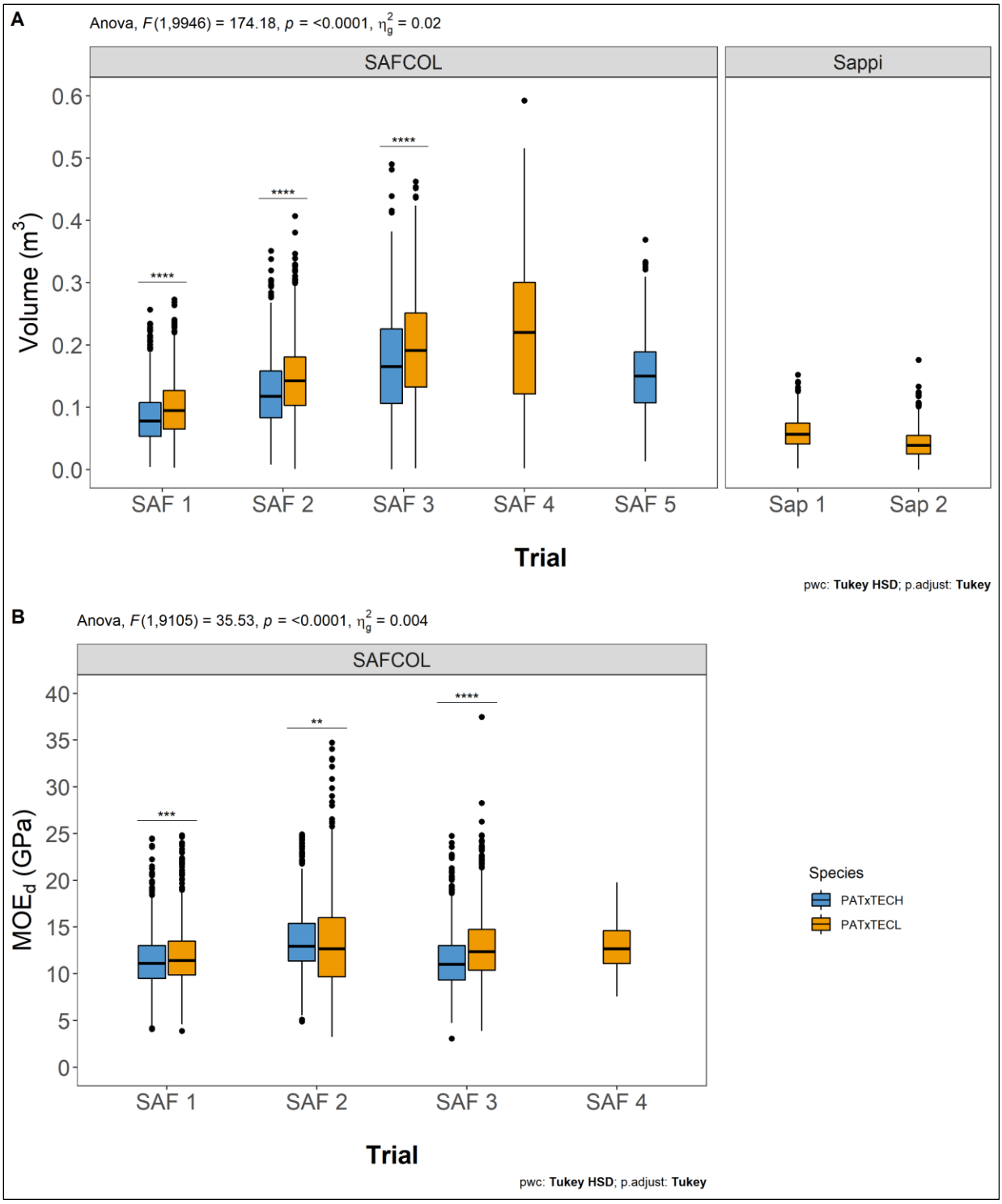


Figure 4.4. Field test description of Volume (m^3) and MOE (Gpa) in each test of SAFCOL and Sappi with measurements. **A:** Volume (m^3) per hybrid variety in each trial per company. **B:** MOE (GPa) of each hybrid variety in SAFCOL trials. The Tukey HSD significance test in Volume (m^3) and MOE_d was performed only in the sites which have the two-hybrid variety growing together ($p < 0.0001$). Significance: 0 '****'; 0.0001 '***'; 0.001 '**'.

The MOE_d was measured in four out of the five SAFCOL sites. The Tukey HSD test indicated that PATxTECL has a higher MOE_d value than PATxTECH in the sites SAF 1 and SAF 3, and the difference is statistically significant ($p<0.001$). In site SAF 2, the observed MOE_d value for PATxTECH was slightly higher than the MOE_d values observed for PATxTECL; similarly, this difference was statistically significant ($p<0.001$).

4.3.3 Genetic Parameters for Volume and MOE

All single-site analyses showed important genetic variation in tree volume, with H_b^2 ranging from 0.36 to 0.62 for PATxTECH, and from 0.24 to 0.36 for PATxTECL. For MOE, the H_b^2 values ranged from 0.08 to 0.62 for PATxTECH, and from 0.11 to 0.7 for PATxTECL. Single-sites results are presented in Appendix E.

In the first multi-site analysis, a very low Type-B genetic correlation (rB_{Gs}) was found at the family level for the PATxTECH hybrid variety, with an rB_{Gs} = 0.41, indicating a moderately high-level genotype x environment interaction. In order to elucidate the possible causes of this interaction, paired-site genetic correlations were calculated using the CORGH structure in Asreml-R (Butler et al., 2017). Thus, several structures of variance-covariances components for the LMM (Eq. 4.10) were parameterized differently, replacing the variance-covariance matrix's off-diagonal for correlations with heterogeneous variance structure (a different variance for volume for each site).

Table 4.2. Site-to-site genetic correlations for standardized volume in SAFCOL field tests for the PATxTECH hybrid. Values are rBg (s.e.) and were calculated using ASREML-R and a CORGH structure to treat volume growth on each site as distinct traits. PATxTECH is not represented in test SAF 4.

	SAF 2	SAF 3	SAF 5
SAF 1	0.55 *(0.16)*	0.69 *(0.13)*	0.13 *(0.25)*
SAF 2		0.79 *(0.12)*	0.10 *(0.26)*
SAF 3			0.19 *(0.30)*

The site-to-site genetic correlations presented in Table 4.2, where site SAF 4 was not included since PATxTECH is not represented on this test, revealed that the SAF 5 has a very low correlation with the other sites (rB_{Gs} = 0.10 to 0.19). Despite the moderately high broad-

sense heritability of site SAF 5, with a H_b^2=0.39, SAF 5 appears to be a different environment, and therefore should not be included in the multi-site analysis with the other three sites. A possible reason for this low correlation could be attributed to the elevation of this test; SAF 5 is located above 1,700 masl, whereas sites SAF 1, SAF 2, and SAF 3 are located at 1,470, 1,265, and 1,050 masl, respectively. SAF 5 has a higher probability of frost events, and it is exposed to lower temperatures, which could cause a higher interaction with the environment and be responsible for the low pairwise site-to-site genetic correlations.

4.3.4 Multiple-Site Analysis for PATxTECH

The multiple-site analysis for PATxTECH showed moderate broad-sense heritability estimates (H^2) for tree volume and MOE. The broad-sense heritability estimate was H^2 = 0.31 for volume and H^2 = 0.20 for MOE. The variance components presented in Table 4.3 were specified in terms of genetic standard deviations ($\sigma = \sqrt{\sigma^2}$), to express the range of the genetic effects in the measurement units of the analyzed trait (i.e., % volume gain and GPa for MOE), and to facilitate the discussion of the results.

Table 4.3. Genetic parameters of multi-site analysis for standardized tree volume (stVol) and MOE with their respective standard errors (SE) in SAFCOL field tests for *P. patula x P. tecunumanii* (HE) variety. GHA_{PAT} or GHA_{TECH} is the General Hybridizing Ability variance due to *P. patula* or *P. tecunumanii* (HE). SHA is the Specific Hybridizing Ability variance. G is the total genetic variance. All genetic variances were expressed in sigma values (σ). H^2 is the broad-sense heritability. rB_{PAT}, rB_{TECH}, $rB_{SHA,}$ and rB_G are the type-B genetic correlations for *P. patula, P. tecunumanii* (HE), SHA, and total genetic variance in the seedling tests (G), respectively.

Traits	σ values ± SE				Genetic Parameters ± SE				
	GHA_{PAT}	GHA_{TECH}	SHA	G_s	H^2	rB_{PAT}	rB_{TECH}	rB_{SHA}	rB_{Gs}
stVol	8.71 *± 3.67*	-	11.15 *± 2.54*	25.48 *± 4.2*	0.31 *± 0.09*	0.76 *± 0.31*	-	0.69 *± 0.19*	0.71 *± 0.11*
MOE	0.66 *± 0.19*	0.34 *± 0.24*	0.43 *± 0.20*	1.36 *± 0.257*	0.2 *± 0.07*	0.98 *± 0.22*	0.71 *± 0.52*	0.52 *± 0.39*	0.76 ± *0.13*

For individual tree volume at 8-years of age, there was an important effect of PAT parentage, with $\hat{\sigma}_{GHA_{PAT}}$ = 8.71. On the TECH side, there was no observed variation due to GHA_{TECH} ($\hat{\sigma}_{GHA_{TEC}}$ = 0), which indicates that TECH parents had no consistent effect on the tree volume of the hybrid progeny. This result was also observed for the single-site analyses (Appendix

E). There seems to be important SHA variation, with $\hat{\sigma}_{SHA}$ = 11.15, indicating that a specific full-sib cross might range from ±22% (representing 95% of the population) above or below the expected family mean based on the GHA_{PAT} and GHA_{TECH} of the parents (Table 4.3). The seedling type-B genetic correlation (rB_{Gs}) estimate was moderate, with a value of rB_{Gs}= 0.71. This correlation was substantially larger than the value from the multi-site analysis mentioned above, including SAF 5 (rB_{Gs}= 0.41). Thus, we can infer that the performance of the families should be relatively stable across the sites SAF 1, SAF 2, and SAF 3.

Both parents contributed to the genetic variation of the MOE trait, with $\hat{\sigma}_{GHA_{PAT}}$ = 0.66 and $\hat{\sigma}_{GHA_{TEC}}$ = 0.34. There was important SHA variation, with $\hat{\sigma}_{SHA}$ = 0.43. In general, there appears to be substantial total seedling genetic variation for this trait, with $\hat{\sigma}_{GS}$ = 1.36 GPa (Table 4.3). This parameter indicates that it should be possible to identify full-sib hybrid families with MOE values substantially above the population mean; a gain of one $\hat{\sigma}_{GS}$ would equal +1.36 GPa, and some families with gains of $2\hat{\sigma}_{GS}$ = +2.72 GPa (or higher) exist in the population and might be identified with substantial selection intensity. There was a moderately low level of GxE interaction found for MOE (rB_{Gs} = 0.76), which indicates a stable performance of families across sites for this trait.

4.3.5 Multi-Site analysis of PATxTECL

Broad-sense heritability estimates for tree volume for PATxTECL were moderate in both datasets (seedling and clonal trials), with H^2 = 0.18 for SAFCOL trials and H^2 = 0.24 for Sappi trials (Tables 4-4 and 4-5). For MOE (measured only at SAFCOL), the estimate of heritability was also moderate, with an H^2 = 0.22 (Table 4.4). The genetic parameters obtained for SAFCOL and Sappi field tests are expressed in sigma values. These values indicate that most of the genetic variation for PATxTECL in both traits analyzed (tree volume and MOE) is due to GHA variation for PAT and TECL parents, with a small amount attributed to the specific cross effect (SHA). The partition of this genetic variation is presented in Table 4.4 and Table 4.5, in which comparisons between the magnitude of $\hat{\sigma}_{GHA_{PAT}}$, $\hat{\sigma}_{GHA_{TEC}}$ and $\hat{\sigma}_{SHA}$ can be observed. For example, in the SAFCOL seedling trials, for volume $\hat{\sigma}_{GHA_{PAT}}$= 7.15 and

$\hat{\sigma}_{GHA_{TECL}}$ = 9.32, while SHA variation was substantially less, with $\hat{\sigma}_{SHA}$ = 5 (Table 4.4). Similar behavior was observed for the partition of the genetic variances in the Sappi trials, $\hat{\sigma}_{GHA_{PAT}}$ = 7.36 and $\hat{\sigma}_{GHA_{TECL}}$ = 11.12, and $\hat{\sigma}_{SHA}$ = 1.30 (Table 4.5). Also, for Sappi tests, there were large clonal differences within full-sib families of the PATxTECL hybrid variety, with $\hat{\sigma}_{Clw}$ = 19.35, indicating that it might be possible to find clones within family that range as much as ±40% around the family mean.

Table 4.4. Genetic parameters of multi-site analysis for standardized volume (stVol) and MOE with their respective SE in SAFCOL field tests for *P. patula x P. tecunumanii* (LE) variety. GHA_{PAT} or GHA_{TECL} is the General Hybridizing Ability variance due to *P. patula* or *P. tecunumanii* (LE). SHA is the Specific Hybridizing Ability variance. Gs is the total seedling genetic variance. All genetic variances were expressed in sigma values (σ). H^2 is the broad-sense heritability. rB_{PAT}, rB_{TECL}, rB_{SHA}, $rB_{Clw,}$ and rB_G are the type-B genetic correlations for *P. patula*, *P. tecunumanii* (LE), SHA, and G x site interaction, respectively.

Traits	σ values ± SE				Genetic Parameters ± SE				
	GHA_{PAT}	GHA_{TECL}	SHA	Gs	H^2	rB_{PAT}	rB_{TECL}	rB_{SHA}	rB_{Gs}
stVol	7.15 *± 1.89*	9.32 *± 2.3*	5.00 *± 1.56*	19.38 *± 2.86*	0.18 *± 0.05*	1.00 *± 0.19*	1.00 *± 0*	0.28 *± 0.16*	0.72 *± 0.09*
MOE	0.64 *± 0.17*	0.37 *± 0.16*	0.61 *± 0.11*	1.6 *± 0.20*	0.22 *± 0.05*	0.91 *± 0.13*	0.78 *± 0.32*	0.56 *± 0.16*	0.71 *± 0.10*

Table 4.5. Genetic parameters of multi-site analysis for standardized volume (stVol) with their respective SE in Sappi clonal field tests for *P. patula x P. tecunumanii* (LE) variety. GHA_{PAT} or GHA_{TECL} is the General Hybridizing Ability variance due to *P. patula* female or *P. tecunumanii* (LE) male. SHA is the Specific Hybridizing Ability variance. Clw is the clonal variance within family. G is the total genetic variance. All genetic variances were expressed in sigma values (σ). H^2 is the broad-sense heritability. rB_{PAT}, rB_{TECL}, rB_{SHA}, rB_{Clw} and rB_{Gcl} are the type-B genetic correlations for patula, tecunumanii (LE), SHA, Clw and G x site interaction respectively.

Trait	σ values ± SE					Genetic Parameters ± SE					
	GHA_{PAT}	GHA_{TECL}	SHA	Clw	G	H^2	rB_{PAT}	rB_{TECL}	rB_{SHA}	rB_{Clw}	rB_G
stVol	7.36 *± 3.61*	11.12 *± 4.37*	1.3 *± 3.97*	19.35 *± 2.17*	23.75 *± 2.87*	0.24 *± 0.05*	1.00 *± 0*	1.00 *± 0*	1.00 *± 0*	0.88 *± 0.15*	0.91 *± 0.10*

The amount of GxE for tree volume, measured through type-B genetic correlations, seems moderate to low for both companies, with rB_{Gs}= 0.72 in the SAFCOL trials and a total genetic type-B correlation rB_G = 0.91 in the Sappi trials. As with the PATxTECH, these parameters indicate that full-sib families and clones of PATxTECL will generally display stable performance across most environments where this variety is likely to be planted.

Nevertheless, this is a preliminary result based on 4-tests for SAFCOL and 2-tests for Sappi (4 pairs of environments), and it would be desirable to evaluate a more significant number of tests to assess if this genotype performance stability across sites is maintained.

MOE genetic parameters for the PATxTECL variety indicated that the genetic variation is controlled by both PAT and TECL parents and the specific full-sib cross. It can be observed in Table 4.4 that $\hat{\sigma}_{GHA_{PAT}}$= 0.64, $\hat{\sigma}_{GHA_{TECL}}$ = 0.37, and $\hat{\sigma}_{SHA}$ = 0.61, with a total full-sib family variation of $\hat{\sigma}_{Gs}$ = 1.6. This last parameter indicates that it should be possible to find full-sib families with MOE as much as +3.2 GPa above the population mean value. The amount of GxE for this trait was relatively small at family level, with an rB_{Gs}= 0.71, a similar value to what was observed for individual tree volume.

4.3.6 Epistasis Variance for Tree Volume

For the Sappi clonal population of the PATxTECL variety, there was substantial clone-within-family variation ($\hat{\sigma}_{Clw}$) for volume gain, and this could imply a significant amount of epistasis variance for this trait. Applying Foster's equation (Eq. 4.23), an estimate of epistasis variance was calculated for all traits in both zones (Table 4.6).

Table 4.6. Variance components of multi-site analysis for standardized volume (stVol) for Sappi PATxTECL clonal population. $\hat{\sigma}^2_{GHA_{PAT}}$ and $\hat{\sigma}^2_{GHA_{TEC}}$ is the General Hybridizing Ability variance due to P. patula or P. tecunumanii low, $\hat{\sigma}^2_{SHA}$ is the Specific Hybridizing Ability variance. $\hat{\sigma}^2_{Clw}$ is the clone within family variance. $\hat{\sigma}^2_G$ is the total genetic variance and $\hat{\sigma}^2_i$ is the estimate of epistasis variance.

Trait	$\hat{\sigma}^2_{GHA_{PAT}}$	$\hat{\sigma}^2_{GHA_{TEC}}$	$\hat{\sigma}^2_{SHA}$	$\hat{\sigma}^2_{Clw}$	$\hat{\sigma}^2_G$	$\hat{\sigma}^2_i$	$\hat{\sigma}^2_i/\hat{\sigma}^2_G$
stVol	54.12 *± 53.05*	123.75 *± 97.27*	11.75 *± 27.2*	374.32 *± 84.01*	563.95 *± 136.21*	161.21 *± 160.04*	0.29 *± 0.31*

The estimate of epistasis variance for volume is relatively high, and the ratio of $\hat{\sigma}^2_i/\hat{\sigma}^2_G$ = 0.29. However, these estimates were not statistically different from zero.

4.3.7 Stability of parent performance: Comparison of GHA vs. GCA

The comparison of the parent genetic value as a pure species parent (GCA) and as a hybrid parent (GHA) was evaluated by a series of plots and regressions for the traits of tree volume and MOE for both the PATxTECL and PATxTECH varieties.

4.3.7.1 Volume

There were only five *P. patula* parents used in the Sappi crosses. Pure species GCA for those five parents were available from Sappi tests, which were calculated using a similar REML/BLUP approach used in this study for the GHA values (Andre Nel, personal communication). Figure 4.5.A displays the distribution of hybrid clone genetic values in each of the five *P. patula* hybrid families.

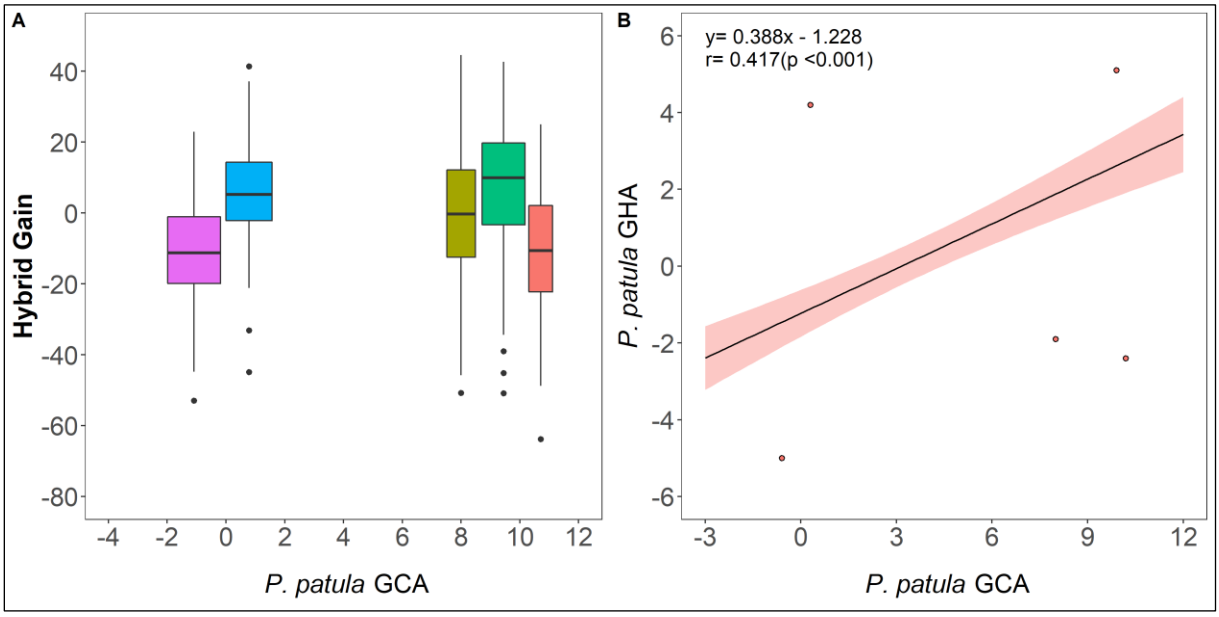


Figure 4.5. Effect of *P. patula* GCA in the clonal hybrid gain and the *P. patula* GHA in Sappi. **A:** Hybrid gain of *P. patula x P. tecunumanii* (LE) *vs. P. patula* GCA. **B:** *P. patula* GHA vs. GCA for *P. patula x P. tecunumanii* (LE) interspecific crosses.

It is clear that there is substantial variability of hybrid gain in all families. For example, a *P. patula* parent with a pure species GCA = +8% can produce hybrid clones ranging from -45% to 45% in volume gain relative to the population mean. Even though the 5 *P. patula mothers*

plotted had an average GCA near 6 ($\overline{GCA}_{PAT}$ = 5.56), the best hybrid clones from each mother surpass +20% of hybrid gain.

Figure 4.5.B shows a weak relationship between the *P. patula* GCA and GHA estimates for Sappi tests, with a GCA-GHA correlation r_{HP} = 0.42 (p<0.001), but with only 5 *P. patula* parents, this correlation must be interpreted with some caution.

Looking at the same GCA-GHA relationship in the SAFCOL tests, *P. patula* GCA-GHA correlation values for tree volume were positive and moderate, with an r_{HP} = 0.65 for the PATxTECL and an r_{HP} = 0.69 for the PATxTECH variety. In both cases, there is some evidence of an unfavorable scale effect; that is, a GCA_{PAT} = +10% corresponds to an expected GHA_{PAT} of approximately +5% in tree volume (lower GHA predictions for PATxTECH variety). Analogous results were observed by examining *P. tecunumanii* low-elevation GHA vs. GCA estimates, with a correlation of r_{HP} = 0.68.

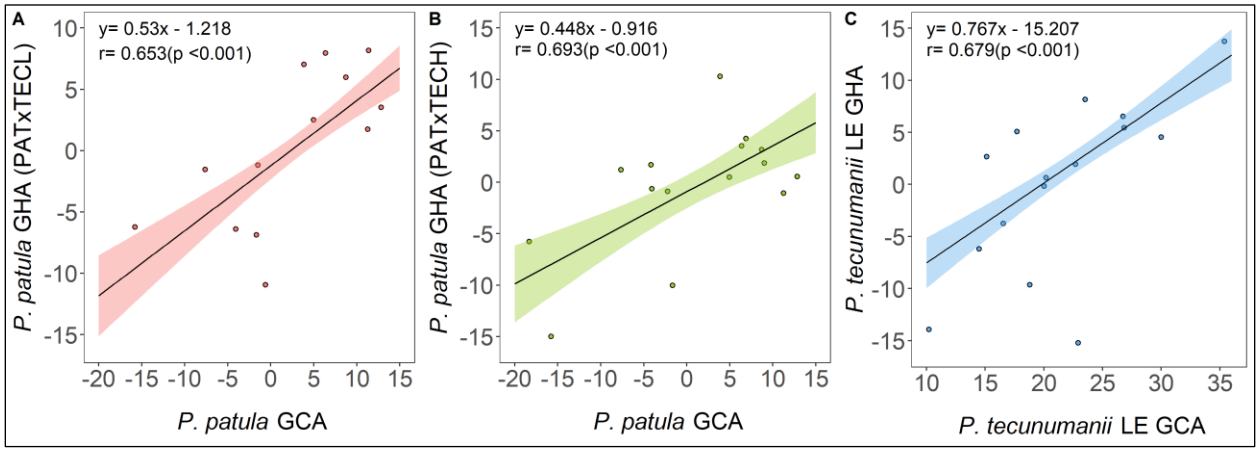


Figure 4.6. GHA vs GCA relationship of standardized volume in SAFCOL family hybrid population of *P. patula x P. tecunumanii*. **A:** *P. patula* GHA vs GCA on *P. patula x P. tecunumanii* (LE). **B:** *P. patula* GHA vs GCA on *P. patula x P. tecunumanii* (LE). **C:** *P. tecunumanii* GHA vs GCA on *P. patula x P. tecunumanii* (LE). The relationship of *P. patula x P. tecunumanii* (HE) was not plotted since $\hat{\sigma}^2_{GHA_{PAT}}$=0

Similarly, for a given GCA_{TECL} value, the predicted GHA_{TECL} value was lower. It was not possible to compare GCA_{TECH} and GHA_{TECH} for high-elevation *P. tecunumanii* due to a zero estimate of GHA_{TECH} variance estimates ($\hat{\sigma}_{GHA_{PAT}}$= 0, Table 4.3). All the correlation values mentioned between GHA and GCA were significantly different from zero (p<0.001), and these relationships can be inspected in Figure 4.6.

4.3.7.2 MOE

For the seedling SAFCOL tests, the relationship between *P. patula* GCA and GHA values for MOE was moderately high, with r_{HP} = 0.72 for PATxTECL (Figure 4.7.A), and a lower but still favorable correlation for PATxTECH, with an r_{HP} = 0.56 (Figure 4.7.C). The GHA-GCA relationship for *P. tecunumanii* HE was non-existent for PATxTECH (Figure 4.7.D), and it is very weak for *P. tecunumanii* low variety, with a correlation value of r_{HP} = 0.32 for PATxTECL (Figure 4.7.B).

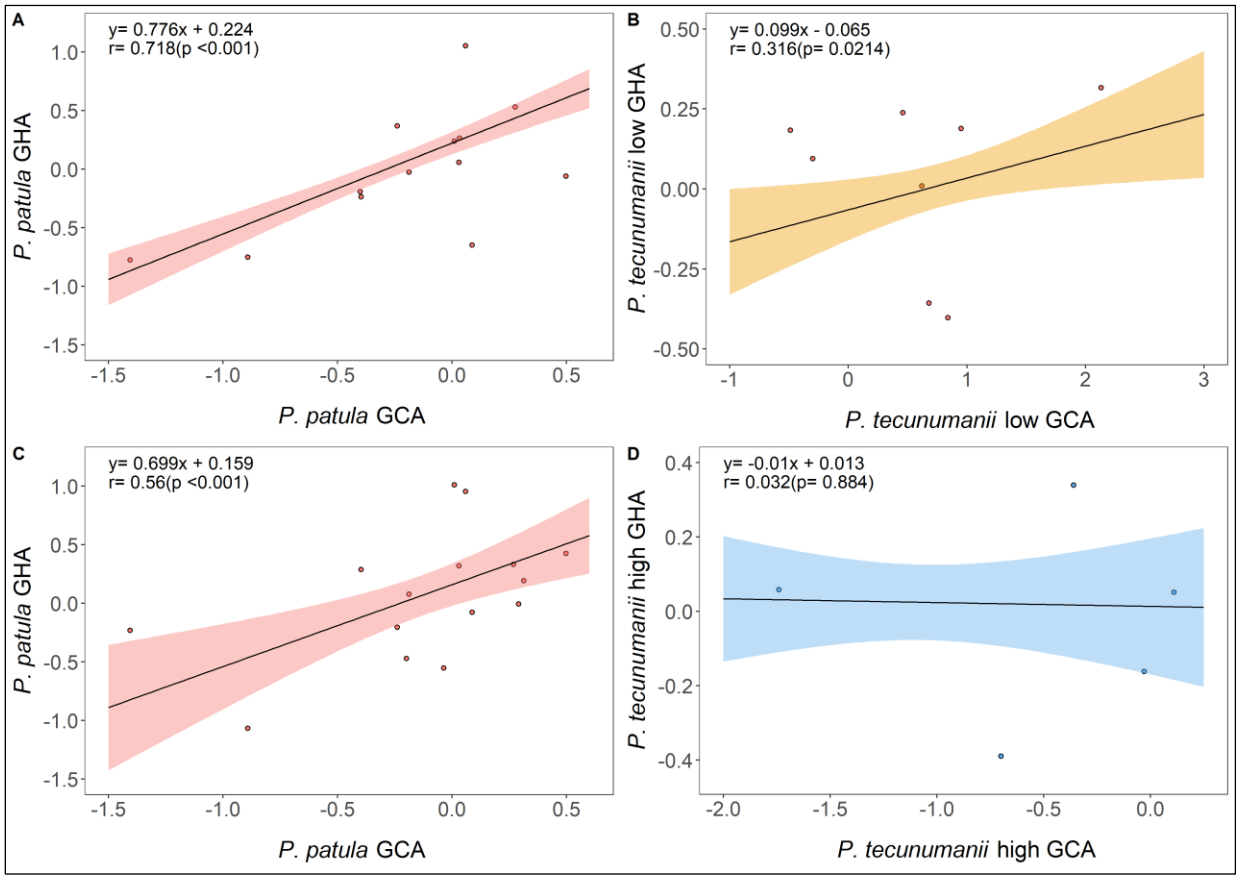


Figure 4.7. GHA vs. GCA relationship on MOE in SAFCOL seedling hybrid population of *P. patula x P. tecunumanii* Low and High Elevation. **A:** *P. patula* relationship on *P. patula x P. tecunumanii* LE variety. **B:** *P. tecunumanii* LE relationship. **C:** *P. patula* relationship on *P. patula x P. tecunumanii* HE variety. **D:** *P. tecunumanii* HE relationship. P-value indicates the significance of the GHA-GCA correlation (p<0.005).

4.4 Discussion

4.4.1 Analysis of Survival and Traits

The survival of the field tests was, on average, just slightly higher for PATxTECH variety than for PATxTECL (84% vs. 82%, respectively). This small difference in survival is similar to the results reported by Hongwane *et al*. (2017) in a study comparing volume growth and MOE at 5-years of several *Pinus* taxa growing in Mpumalanga, South Africa, including hybrids of *P. patula x tecunumanii* var high and low. On sites where these hybrid varieties were together, Hongwane et al. (2017) reported graphically in several bar plots a higher survival rate of PATxTECH than PATxTECL. The same study described higher volume growth at 5-years of PATxTECL compared to PATxTECH, but those differences were not statistically significant. The current study also found differences in volume, but the differences were statistically significant ($p < 0.001$), perhaps since this evaluation was made with measurements taken at 8-years. In a different study of several Pine hybrids tested in trials across South Africa, Hongwane et al. (2018) reported that the productivity of PATxTECL was always higher than the variety PATxTECH, across all different climate types where the tests were established, broken out as Mediterranean, Cool-Temperate, Warm-Temperate, Warm-Temperate/Tropical. These differences in volume productivity (m^3) were statistically significant in 5 of the 11 tests reported.

For MOE, in the current study, PATxTECL also performs better than PATxTECH, with statistically significant differences ($p<0.001$). The same result was reported by Hongwane et al. (2017), with statistically significant differences between these varieties. Thus, to summarize the performance of the two hybrid varieties, PATxTECH has displayed slightly better survival, but PATxTECL has shown better volume productivity and higher wood MOE.

4.4.2 Genetic Parameters for Individual Tree Volume

For the two parental species, estimates of heritability (at the within-provenance level) have been reported for individual tree volume with estimates of $h^2 = 0.25$ for *P. patula* (volume measured at 8 years) and $h^2 = 0.16$ for *P. tecunumanii* (for volume measured at 5 to 8 years)

(Dvorak, Hodge, Gutierrez, et al., 2000; Dvorak, Hodge, Kietzka, et al., 2000). These reported heritabilities are a little lower than the estimates of H^2 = 0.18 to 0.31 for the hybrid populations in the current study.

Type-B genetic correlations across sites of SAFCOL family hybrids of PATxTECL and PATxTECH were similar to what was described for the parental species. For example, estimates of rB_{Gs}= 0.75 for *P. patula* parent in pure species tests in South Africa (Dvorak, Hodge, Kietzka, et al., 2000), and an rB_{Gs}= 0.81 in *P. tecunumanii* progeny tests (Dvorak, Hodge, Gutierrez, et al., 2000) have been reported, very similar to the estimates in this study (see Tables 4.3, 4.4 and 4.5). The similarity in the genetic parameters of the parental species with the hybrid gives more confidence in the credibility of the hybrid genetic parameter estimates.

It is notable that in the genetic analysis of individual tree volume for PATxTECL, the partition of GHA variance among the two species was similar for the independent datasets from the two companies. In both cases, the variance associated with GHA_{TECL} was higher than the variance for GHA_{PAT}. In other words, more of the variation in volume among hybrid progeny came from the *P. tecunumanii* parentage than the *P. patula* parentage. This result suggests that a breeding strategy to maximize the volume gain in PATxTECL variety could focus on increasing the selection intensity on *P. tecunumanii* parents over *P. patula*. In addition, although SHA variance is not zero in the PATxTECL hybrid, it appears to be substantially less important than either $\hat{\sigma}^2_{GHA_{PAT}}$ or $\hat{\sigma}^2_{GHA_{TEC}}$, suggesting that for this trait, the parents do not need to be tested in a large number of crosses.

The PATxTECH hybrid showed a very different pattern of genetic variance for volume. There was substantial variance for GHA_{PAT}, but zero variance for GHA_{TECH}, and the largest contributor to the genetic variance was SHA. So even if true GHA_{TECH} variance ($\hat{\sigma}^2_{TECH}$) is not zero, it is likely that $\hat{\sigma}^2_{TECH}$ is relatively low and much less important than the other effects for hybrid tree volume.

This study appears to be the first quantitative genetic analysis of a growth trait for the PATxTEC hybrid; however, there are some genetic studies for the hybrid for *Fusarium* (pitch canker) resistance and for frost tolerance.

A study of resistance to *Fusarium circinatum*, conducted by Mitchell et al. (2013), stated that the survival of *P. patula* in South Africa is frequently unacceptable from a commercial standpoint due to the susceptibility of that species to pitch canker. With hybridization with a tolerant species (i.e., *P. tecunumanii*), the survival is improved. Mitchell conducted a study composed of 25 full-sib hybrid families of PATxTECL (from 13 PAT x 12 TECL parents) and 24 full-sib hybrid families of PATxTECH (from 10 PAT x 7 TECH parents). In that study, seedlings were artificially inoculated with the pathogen, and the resulting lesion length was measured. For pure *P. patula,* the average lesion length was high (approx. 17 mm to 39 mm), followed by PATxTECH with medium size of lesion length (from 5 to 25 mm), and finally PATxTECL, with a small size of the lesion length (from 3 to 15 mm). These phenotypic observations are in concordance with what was found in the genetic parameters for the hybrids: for PATxTECH, the parameter estimates were: h^2_{PAT} = 0.25, h^2_{TECH} = 0.08, and $d^2_{PATxTECH}$ = 0.39; for PATxTECL, the parameter estimates were: h^2_{PAT} = 0.03, h^2_{TECL} = 0.07, and $d^2_{PATxTECL}$ = 0.17. In summary, PATxTECL is more resistant and more genetically uniform than PATxTECH, which shows improved tolerance for pure *P. patula*, and substantial genetic variability. In PATxTECH, more variation comes from the *P. patula* than the TECH, while in PATxTECL, more variation comes from the TECL than the *P. patula*. In both varieties, SHA contributes substantially to the genetic variation.

In frost resistance, an artificial freeze testing using the electrolyte leakage method was conducted by Cerda (2012) on families of PATxTECL (from 5 PAT x 4 TECL parents), measuring the damage on the needles under 3 different temperatures (-3, -14, and -21 °C). A significant variation among the PAT parents was found, but for TECL, the variation for these traits was almost zero. It appears that for frost resistance, the *P. patula* parentage brings most of the genetic variation for the hybrid progeny.

These results suggest that *P. patula* and *P. tecunumanii* may bring different amounts of genetic variance to the hybrid depending on the trait. It seems possible that this would be

the case for growth traits as well. Some authors working with hybrids reported a combined or aggregate GHA variance for the two-parent species in the literature as a single combined additive effect (e.g., Belaber et al., 2018; Gwaze et al., 2000), while other authors separate the GHA variance estimates of the parent species. Authors examining separate GHA variances for the parent species have often found quite different GHA variances from different parent species (and sometimes zero GHA variance).

For the *Pinus elliottii x P. caribaea* var. *hondurensis* hybrid (PEExPCH), Kain (2003) reported 11-year growth trait results from a 12 x 12 mating design and concluded that greater gain is possible from the selection of PCH parents for growth than from selecting PEE parents for growth. For the trait of under-bark volume with 10% parent selection, the expected gain was 5.0% from PEE parent selection and 8.4% from PCH selection, which corresponds to approximately 3-times higher GHA variance from PCH than from PEE.

Data from populations of two different subtropical pine hybrids tested in Australia and Zimbabwe are reported by Dieters *et al*. (1997) and Mutete *et al*. (2015), respectively. A set of 29 *Pinus caribaea* var. *hondurensis* × *Pinus tecunumanii* (PCH×TECL) families derived from 11 PCH parents and 6 PTEC parents, and another set of 26 *P. caribaea* var. *hondurensis* × *Pinus oocarpa* (PCH×OOC) families from the same 11 PCH parents and 6 POOC parents. The hybrid families were produced by the Queensland Forest Research Institute, Australia, with *P. tecunumanii* material initially collected by Oxford Forestry Institute, UK, from provenances from Belize and Nicaragua countries, collected at an altitude lower than 900 masl (Dieters *et al*., 1997). Therefore, it can be inferred that those *P. tecunumanii* parents correspond to the low elevation variety (TECL). Dieters *et al*., 1997 report results for 5-year DBH (DBH5), and Mutete *et al*. (2015) report results for 8-year height and DBH (HT8 and DBH8). Both authors calculated GHA variance separately for the parent species and reported the results as hybrid heritabilities, e.g., h^2_{PCH} and h^2_{TECL}.

For the PCHxTECL hybrid, in Australia, more variation for DBH5 came from the PCH parents than the TECL parents, which had almost zero GHA variance (h^2_{PCH} = 0.17, h^2_{TECL} = 0.03), indicating a GHA variance due to PCH that was 5.6 times bigger than from TECL. In

Zimbabwe, the opposite pattern was observed, with more variation reported from the TECL parents than the PCH parents for HT and DBH measured at 8 years (HT8 and DBH8, respectively: for HT8, $h^2_{PCH} = 0.00$, $h^2_{TECL} = 0.16$; for DBH8, $h^2_{PCH} = 0.17$, $h^2_{TECL} = 0.38$). For the PCHxOOC hybrid, there was also an opposite pattern observed in Australia and Zimbabwe. In Australia, more GHA variance came from *P. oocarpa* than PCH ($h^2_{PCH} = 0.18$, $h^2_{OOC} = 0.30$), while in Zimbabwe, very large GHA variance came from PCH and almost none from OOC (for HT8, $h^2_{PCH} = 0.50$, $h^2_{OOC} = 0.07$; for DBH8, $h^2_{PCH} = 0.31$, $h^2_{OOC} = 0.04$).

The differences in genetic parameters from the Australian and Zimbabwe experiments might be attributed to sampling variation of different progeny that composed these populations. However, the fact that the same parents were used for both experiments and tested in different continents also suggests the possibility of an environmental effect on the expression of GHA variances.

The results obtained by Mutete *et al.* (2015) roughly correspond to the current study results, where a high amount of GHA variance for growth traits was found for TECL relative to the other parent species in the hybrid combination (*P. patula* or PCH). In addition, a GHA variance of zero was found for one of the parents (PCH for HT8), similar to the GHA variance of zero found in this study for TECH (for tree volume). Perhaps this similarity in the results is due to some resemblance in climate and environment for Zimbabwe and South Africa tests.

Regarding PATxTECH, in the first attempt of SAFCOL seedling analysis for standardized volume, a low type-B genetic correlation was obtained, with an $rB_{Gs} = 0.41$ when site SAF 5 was included, comparing with the $rB_{Gs} = 0.71$ obtained without this site. This result indicated a possible zoning effect in the SAFCOL landbase for this hybrid variety, suggesting that it might be necessary to establish breeding zones based on altitude, MAP, MAT, or all these variables in conjunction and then select the best-adapted genotypes for each zone (higher or lower than 1,500 masl).

4.4.3 Genetic Parameters for MOE

For the wood trait MOE, the estimates of the genetic parameters were very similar for the two-hybrid varieties evaluated by SAFCOL (Table 4.3 and 4.4). In both hybrids, variances associated with GHA_{PAT} and SHA were larger than the variance of $GHA_{TEC.}$ Broad-sense heritability was slightly higher for PATxTECL than for PATxTECH, and this is due primarily to a marginally higher SHA variance observed for PATxTECL. Since the SHA variance for MOE was considerable in both hybrid varieties, this would imply that a moderate number of crosses per parent would be required to capture this potential genetic gain.

There is a low to moderately low amount of GxE for MOE in both hybrid varieties, so it is expected to find families that perform very well across a wide range of sites. In the SAFCOL studies, MOE was measured in standing trees with high moisture content in the wood. This "green MOE" value is assumed to be determined primarily by cellulose microfibril angle (MFA) (Gapare et al., 2009) and is a good indicator of timber strength.

For the PEExPCH hybrid, Kain (2003) reported roughly equal amounts of genetic gain possible from parental selection for wood density (5.7% from PEE, and 5.2% from PCH), indicating similar amounts of GHA genetic variance from both species. Kain further reports almost zero SHA variance for density, a very different finding than for MOE in the PATxTEC hybrids. Therefore, for PATxTEC breeding, if increased MOE is a desirable breeding objective, this would favor increasing the numbers of crosses per parent to capture SHA gains.

4.4.4 Epistasis

From the Sappi population trials, it seems that there is some epistasis genetic variance for tree volume in PATxTECL. Estimated epistasis variance was $\hat{\sigma}_i^2$= 161.2, which corresponded to 29% of the total genetic variation. However, this result must be viewed with some caution since the SE of the estimates was very high (Table 4.6), and the variance estimate was not statistically different from zero.

There are estimates of epistasis variance for growth traits available from clonal studies made for pure pine and conifer species. For example, in a study of a clonal population of *Pinus radiata* in New Zealand, based on 52 full-sib families, with a total of 664 clones, Baltunis et al. (2009) estimated epistasis variance using the Foster and Shaw (1988) approach. In that study, they reported a significant epistasis variance for DBH measured at 5-years, with a value of $\hat{\sigma}_i^2$ = 0.10 ± 0.05, which corresponded to 35.7% of the total genetic variation. This epistasis value was also not significantly different from zero, but it is of interest that it was of a very similar size to the estimated epistasis variance in the current study.

In *Pinus taeda*, Isik et al. (2003) estimated the additive, dominance, and epistatic effect for 9 full-sib families and clones of the same families for growth traits and fusiform rust incidents. These authors reported negative values (near zero) for epistasis variance for the growth traits total height measured at 1, 2, 4, and 6 years, DBH at 4 and 6 years, and volume at 4 and 6 years. These negative variance values were interpreted as a non-significant effect for epistasis in growth traits.

In a full-sib clonal population of *Picea abies,* composed of 32 full-sib families with a total of 1430 clones, growth traits and Pilodyn penetration were evaluated by Chen et al. (2020). A significant amount of epistasis for volume was reported at 12 years, corresponding to 32% of the total genetic variance in that study. Similar to the results mentioned above, the epistasis variance was estimated with a high standard error and was not significantly different from zero but was of a similar size to the result in the current study.

In each of the studies where non-zero epistasis variation for growth traits was found for pine or conifer populations, the estimates were statistically non-significant. However, the size of epistasis variance estimates was very consistent, ranging from 29% to 36% of the total genetic variance in 3 of the 4 Pine populations described above.

4.4.5 Hybrid GHA vs. Pure Species GCA

In both varieties, positive correlations of moderate size were found for *P. patula* pure species GCA and hybrid GHA for volume (r_{HP} = correlation of hybrid GHA and pure species GCA). For PATxTECL, the correlation of *P. patula* GCA and GHA was r_{HP} = 0.65, and for PATxTECH the correlation was r_{HP} = 0.69. There was also a moderate correlation of *P. tecunumanii* low pure species GCA and hybrid GHA, with r_{HP} = 0.68. All of these correlations were significantly different from zero. They suggested that pure species PAT and TECL GCA values for volume will be good indicators of genetic worth as a hybrid parent. Since there was zero GHA variance found for TECH, GCA-GHA correlations for this species are undefined.

For the wood trait MOE, there was a relationship between *P. patula* GCA and GHA in both varieties (PATxTECL and PATxTECH), but with higher correlations found for PATxTECL (r_{HP} = 0.72) than for PATxTECH (r_{HP} = 0.56). In contrast, for *P. tecunumanii* low and high varieties, there were weak to almost zero correlations between GCA and GHA, with r_{HP} = 0.32 for PATxTECL, and r_{HP} = 0.03 for PATxTECH. Thus, it appears that in general, *P. patula* GCA values are moderate predictors of the hybrid GHA for MOE, while the MOE GCA for *P. tecunumanii* low and high varieties are non-predictors of the MOE GHA value.

For the PEExPCH hybrid, Kain (2003) reported moderate to high GHA-GCA correlations for growth traits and very high correlations for wood density. Specifically, for 11-year DBH, he found r_{HP} = 0.51 for PEE, and r_{HP} = 0.83 for PCH, while for density the correlations were r_{HP} = 0.94 for PEE and r_{HP} = 0.85 for PCH. In contrast, in a study of 56 PEE parents tested with a PCH pollen mix, no correlation of GHA and GCA was found for the PEE parental effect for 11-year growth traits height, DBH, or volume (average r_{HP} = 0.02) (Brawner et al., 2004)

4.4.6 Breeding Recommendations

A sizeable genetic gain could be achieved in volume gain, considering the high variability found between the hybrid families of PATxTECH and PATxTECL in SAFCOL and PATxTECL clones in Sappi. It should be possible to find individuals ranging over ±25% in volume gain

in SAFCOL and near to ±40% in PATxTECL variety in Sappi clonal population in both varieties. Particularity in PATxTECL, a considerable gain can be accomplished by increasing the selection intensity of TECL parents since these parents demonstrate a larger GHA variance than PAT in both companies. On the PATxTECH variety, no GHA variance was found for TECH parents, and most of the genetic effect was attributed to SHA variance, followed by PAT GHA variance. Under this scenario, a conservative strategy is to concentrate the effort on selecting more PAT parents to be crossed with some TECH.

For volume, *P. patula* contributes substantial GHA variance to PATxTECH and PATxTECL varieties, and in both hybrids, there is a moderately strong GCA-GHA correlation. On the *P. tecunumanii* side, TECH does not contribute GHA variance for volume, but TECL contributes important GHA variance and displays a moderately strong GCA-GHA correlation. Thus, to improve volume growth in the hybrids, *P. patula* parents and TECL parents could be selected based on their performance in pure species breeding programs. There is some evidence of an unfavorable scale effect, where a high GCA might translate to a lower GHA (e.g., GCA_{PAT} = 10% might be GHA_{PAT} = 5% in the PATxTECL hybrid; see Figure 4.6), but still, the relationship is positive, and preselection of parents based on pure species breeding programs should be possible.

For MOE, it might be possible to increase MOE as much as ±2.78 and ±3.2 GPa in PATxTECH and PATxTECL, respectively, by selecting the best full-sib families. However, most of the MOE gain possible would come from *P. patula* GHA variance, which is greater in both hybrid varieties than GHA variance from TECH and TECL. In addition, there were moderately strong GHA-GCA correlations for MOE for *P. patula* parents in both hybrid varieties. Thus, it should be possible to select good *P. patula* hybrid parents for MOE based on progeny performance in pure species tests. The *P. tecunumanii* parents contributed less GHA variance and also had a very weak to zero correlation between GHA and GCA. Therefore, parental selection to improve MOE in hybrid progeny of PATxTECL or PATxTECH should be made based primarily on *P. patula*.

4.5 Conclusions

A high Genotype × Environment interaction was found in SAFCOL field tests for PATxTECH variety, and this was due primarily to one site with a different growth behavior ascribable to the high altitude. Therefore, it is recommended that some type of regionalization or zoning of the SAFCOL landbase should be done, perhaps based on altitude, for the testing and deployment of this variety.

For the PATxTECH variety, the PAT parents demonstrated a considerable impact in the volume growth of this variety, which makes it essential to test more PAT parents in future interspecific crosses in SAFCOL. However, the GHA variance of TECH parents was zero, and under this result, it would be preferable to use fewer TECH fathers in future interspecific crosses or to select these parents for their merits in other traits such as MOE or *Fusarium* tolerance.

For the PATxTECL variety, the TECL parents consistently demonstrate higher impact than PAT for growth, with higher GHA variance than PAT in Sappi and SAFCOL field tests. The recommendation is to test more TECL parents in future interspecific crosses to increase the volume gain. The PAT parents demonstrated a higher GHA on MOE; thus, future selection should be made for PAT in MOE and for TECL in volume performance.

The very low GxE interaction is a good indicator that the selections made within the company breeding programs will perform similarly across sites, without ranking changes.

There is some evidence of a low amount of epistasis variation for tree volume in the PATxTECL population. If commercial clonal deployment of the hybrid becomes a possibility, more research should be done to confirm this type of genetic variance.

The positive correlations found between GHA and GCA for volume allow parent selection for future interspecific crosses based on the GCA performance in a pure species breeding program. However, due to a scale effect where a high GCA seems to translate to lower GHA values, more genetic gain could be made by selecting parents for new hybrid crosses based on their performance observed in hybrid progeny tests.

4.6 References

Araújo, J. A., Borralho, N. M. G., & Dehon, G. (2012). The importance and type of non-additive genetic effects for growth in Eucalyptus globulus. Tree Genetics and Genomes, 8(2), 327–337. https://doi.org/10.1007/s11295-011-0443-x

Baltunis, B. S., Huber, D. A., White, T. L., Goldfarb, B., & Stelzer, H. E. (2005). Genetic effects of rooting loblolly pine stem cuttings from a partial diallel mating design. Canadian Journal of Forest Research, 35(5), 1098–1108. https://doi.org/10.1139/x05-038

Baltunis, B. S., Wu, H. X., Dungey, H. S., Mullin, T. J. "Tim," & Brawner, J. T. (2009). Comparisons of genetic parameters and clonal value predictions from clonal trials and seedling base population trials of radiata pine. Tree Genetics and Genomes, 5(1), 269–278. https://doi.org/10.1007/s11295-008-0172-y

Belaber, E. C., Gauchat, M. E., Reis, H. D., Borralho, N. M., & Cappa, E. P. (2018). Genetic Parameters for Growth, Stem Straightness, and Branch Quality for Pinus elliottii var. elliottii × Pinus caribaea var. hondurensis F1 Hybrid in Argentina. Forest Science, 64(6), 595–608. https://doi.org/10.1093/forsci/fxy021

Bison, O., Ramalho, M. A. P., Peçanha Rezende, G. D. S., Aguiar, A. M., & de Resende, M. D. V. (2007). Combining ability of elite clones of Eucalyptus grandis and Eucalyptus urophylla with Eucalyptus globulus. Genetics and Molecular Biology, 30(2), 417–422. https://doi.org/10.1590/S1415-47572007000300019

Brawner, J. T., Dieters, M. J., & Nikles, D. G. (2004). Correlations between pure and hybrid combining abilities of slash pine parents. Forest Genetics, 10(3), 241–248.

Burdon, R. D. (1977). Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. Silvae Genet., 26(5–6), 168–175.

Butler, D. G. (2019). asreml: Fits the Linear Mixed Model. In R package version 4.1.0.110. www.vsni.co.uk

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2017). ASReml-R Reference Manual Version 4. ASReml-R Reference Manual, 176.

http://www.homepages.ed.ac.uk/iwhite/asreml/uop.

Cerda, D. (2012). Geographical Variation of Cold Hardiness in Pinus patula Provenances and Genetic Inheritance of Cold Hardiness in Pinus patula × Pinus tecunumanii Hybrids.

Chen, Z.-Q., Hai, H. N. T., Helmersson, A., Liziniewicz, M., Hallingbäck, H. R., Fries, A., Berlin, M., & Wu, H. X. (2020). Advantage of clonal deployment in Norway spruce (Picea abies (L.) H. Karst). Annals of Forest Science, 77(1), 14. https://doi.org/10.1007/s13595-020-0920-1

Costa E Silva, J., Borralho, N. M. G., Araújo, J. A., Vaillancourt, R. E., & Potts, B. M. (2009). Genetic parameters for growth, wood density and pulp yield in Eucalyptus globulus. Tree Genetics and Genomes, 5(2), 291–305. https://doi.org/10.1007/s11295-008-0174-9

Costa E Silva, J., Borralho, N. M. G., & Potts, B. M. (2004). Additive and non-additive genetic parameters from clonally replicated and seedling progenies of Eucalyptus globulus. Theoretical and Applied Genetics, 108(6), 1113–1119. https://doi.org/10.1007/s00122-003-1524-5

Dieters, M. J., Nikles, D. G., Toon, P. G., & Pomroy, P. (1997). Genetic parameters for F1 hybrids of Pinus caribaea var. hondurensis with both Pinus oocarpa and Pinus tecunumanii. Canadian Journal of Forest Research, 27(7), 1024–1031. https://doi.org/10.1139/x97-053

Dieters, M. J., White, T. L., Littell, R. C., & Hodge, G. R. (1995). Application of approximate variances of variance components and their ratios in genetic tests. Theoretical and Applied Genetics, 91(1), 15–24. https://doi.org/10.1007/BF00220853

Dvorak, W. S., Hodge, G. R., Gutierrez, E. A., Malan, F. S., Osorio, L. F. and, & Stanger, T. K. (2000). Pinus tecunumanii. In Consevation & Testing of Tropical and Subtropical Forest Tree Species by the CAMCORE Cooperative (pp. 188–209). College of Natural Resources, NCSU.

Dvorak, W. S., Hodge, G. R., Kietzka, J. E., Malan, F., Osorio, L. F., & Stanger, T. K. (2000). Pinus patula. In Consevation & Testing of Tropical and Subtropical Forest Tree Species by the CAMCORE Cooperative (pp. 148–173). College of Natural Resources, NCSU.

Falconer, D. . S., & Mackay, T. F. . (1996). Introduction to quantitative genetics (4th edn). Pearson. https://doi.org/10.1002/bimj.19620040211

Foster, G. S., & Shaw, D. V. (1988). Using clonal replicates to explore genetic variation in a perennial plant species. Theoretical and Applied Genetics, 76(5), 788–794. https://doi.org/10.1007/BF00303527

Fukatsu, E., Hiraoka, Y., Matsunaga, K., Tsubomura, M., & Nakada, R. (2014). Genetic relationship between wood properties and growth traits in Larix kaempferi obtained from a diallel mating test. Journal of Wood Science, 61(1), 10–18. https://doi.org/10.1007/s10086-014-1436-9

Gapare, W. J., Baltunis, B. S., Ivković, M., & Wu, H. X. (2009). Genetic correlations among juvenile wood quality and growth traits and implications for selection strategy in Pinus radiata D. Don. Annals of Forest Science, 66(6), 606–606. https://doi.org/10.1051/forest/2009044

Gapare, W. J., & Musokonyi, C. (2002). Provenance performance and genetic parameter estimates for Pinus caribaea var. hondurensis planted at three sites in Zimbabwe. Forest Genetics, 9(3), 183–189.

Gezan, S. A., de Carvalho, M. P., & Sherrill, J. (2017). Statistical methods to explore genotype-by-environment interaction for loblolly pine clonal trials. Tree Genetics & Genomes, 13(1), 1. https://doi.org/10.1007/s11295-016-1081-0

Gwaze, D. P., Dungey, H. S., Dieters, M. J., Toon, P. G., & Nikles, D. G. (2000). Interspecific pine hybrids. I. Genetic parameter estimates in Australia. Forest Genetics, 7(1), 11–20.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72(358), 320–338. https://doi.org/10.1080/01621459.1977.10480998

Hodge, G. R., & Dvorak, W. S. (2012). Growth potential and genetic parameters of four Mesoamerican pines planted in the Southern Hemisphere. Southern Forests, 74(1), 27–49. https://doi.org/10.2989/20702620.2012.686192

Hodge, G. R., & Dvorak, W. S. (2015). Provenance variation and within-provenance genetic parameters in Eucalyptus urophylla across 125 test sites in Brazil, Colombia, Mexico, South Africa and Venezuela. Tree Genetics and Genomes, 11(3). https://doi.org/10.1007/s11295-015-0889-3

Hongwane, P., Mitchell, G., Kanzler, A., Verryn, S., Lopez, J., & Chirwa, P. (2018). Alternative pine hybrids and species to Pinus patula and P. radiata in South Africa and Swaziland. Southern Forests: A Journal of Forest Science, 80(4), 301–310. https://doi.org/10.2989/20702620.2017.1393744

Hongwane, P., Mitchell, R., Kanzler, A., Verryn, S., & Chirwa, P. (2017). Growth and dynamic modulus of elasticity of Pinus patula × Pinus tecunumanii hybrids in Mpumalanga, South Africa. Southern Forests: A Journal of Forest Science, 79(4), 277–285. https://doi.org/10.2989/20702620.2016.1254899

Isik, F., Li, B., & Frampton, J. (2003). Estimates of additive, dominance and epistatic genetic variances from a clonally replicated test of loblolly pine. Forest Science, 49(1), 77–88. https://doi.org/10.1093/forestscience/49.1.77

Kabir, M., Akpa, G. N., Nwagu, B. I., Adeyinka, I. A., Shehu, D. M., Galadima, M. A., & Yahaya, H. K. (2014). General combining ability (GCA), specific combining ability (SCA) and reciprocal effects on average daily gain in body weights at various ages of rabbit in northern guinea savannah zone of Nigeria. IOSR Journal of Agriculture and Veterinary Science, 7(4), 48–51. https://doi.org/10.9790/2380-07414851

Kain, D. (2003). Genetic parameters and improvement strategies for the Pinus elliottii var. elliottii x Pinus caribaea var. hondurensis hybrid in Queensland, Australia (Doctoral dissertation, The Australian National University, Camberra, Australia). Retrieved from https://openresearch-repository.anu.edu.au/handle/1885/9236

Kanzler, A., Nel, A., & Ford, C. (2014). Development and commercialisation of the Pinus patula × P. tecunumanii hybrid in response to the threat of Fusarium circinatum. New Forests, 45(3), 417–437. https://doi.org/10.1007/s11056-014-9412-1

Kanzler, A., Payn, K., & Nel, A. (2012). Performance of two Pinus patula hybrids in southern Africa. Southern Forests: A Journal of Forest Science, 74(1), 19–25. https://doi.org/10.2989/20702620.2012.683639

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. In R package version 0.7.0. https://cran.r-project.org/package=rstatix

Ladrach, W. E. (1986). Comparaciones entre procedencias de siete coníferas en la Zona Andina al finalizar los ochos años.

Ladrach, W. E., & Mazuera, H. (1978). Growth and development of the Chupillauta arboretum at 7 years and the Mexican pines in San Jose at 6.3 years. In Cartón de Colombia Research Report (Vol. 34, p. 4).

Lopez, J. L., Abt, R. C., Dvorak, W. S., Hodge, G. R., & Phillips, R. (2018). Tree breeding model to assess financial performance of pine hybrids and pure species: deterministic and stochastic approaches for South Africa. New Forests, 49(1), 123–142. https://doi.org/10.1007/s11056-017-9609-1

Lynch, M., & Walsh, B. (1998). Genetic and Analysis of Quantitative Traits. Sinauer. https://books.google.com/books?id=UhCCQgAACAAJ

Madhibha, T., Murepa, R., Musokonyi, C., & Gapare, W. (2013). Genetic parameter estimates for interspecific Eucalyptus hybrids and implications for hybrid breeding strategy. New Forests, 44(1), 63–84. https://doi.org/10.1007/s11056-011-9302-8

Mitchell, R. G., Wingfield, M. J., Hodge, G. R., Steenkamp, E. T., & Coutinho, T. A. (2013). The tolerance of Pinus patula × Pinus tecunumanii, and other pine hybrids, to Fusarium circinatum in greenhouse trials. New Forests, 44(3), 443–456. https://doi.org/10.1007/s11056-012-9355-3

Mutete, P., Murepa, R., & Gapare, W. J. (2015). Genetic parameters in subtropical pine F1 hybrids: heritabilities, between-trait correlations and genotype-by-environment interactions. Tree Genetics & Genomes, 11(5), 93. https://doi.org/10.1007/s11295-015-0926-2

Nöjd, P., & Isango, J. (2003). Growth Periodicity of Pinus patula in Southern Tanzania. Journal of Tropical Forest Science, 15(2), 352–358. https://www.jstor.org/stable/43594507

Patterson, H. D. and, & Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. Biometrika, 58(3), 545–554. https://www.jstor.org/stable/2334389

Team R. Core. (2019). R: A language and environment for statistical computing. In R Foundation for Statistical Computing. https://doi.org/10.1108/eb003648

van den Berg, G. J., Verryn, S. D., Chirwa, P. W., & Deventer, F. Van. (2015). Genetic Parameters of Interspecific Hybrids of Eucalyptus grandis and E. urophylla Seedlings and Cuttings. Silvae Genetica, 64(1–6), 291–308. https://doi.org/10.1515/sg-2015-0027

White, T. L., Adams, W. T., & Neale, D. B. (2007). Forest Genetics. In CABI Publishing. CABI Publishing. https://doi.org/10.1038/167764a0

Wielinga, B., Raymond, C. A., James, R., & Matheson, A. C. (2009). Effect of green density values on pinus radiata stiffness estimation using a stress-wave technique. New Zealand Journal of Forestry Science, 39(1), 71–79.

Zhu, Y., Wu, S. J., Xu, J., Lu, Z., Li, G., Hu, Y., Yang, X., & Bush, D. (2017). Genetic parameters for growth traits and stem-straightness in Eucalyptus urophylla × E. camaldulensis hybrids from a reciprocal mating design. Euphytica, 213(7), 1–16. https://doi.org/10.1007/s10681-017-1923-3

Chapter 5:

Conclusions

5.1 Summary of Research Chapters

There were three main research topics in this manuscript. Chapter 2 covered a simulation study on the precision of variance component estimation for hybrid populations. Chapter 3 examined volume and three wood property traits for a population of hybrid clones of *E. nitens x E. globulus* (GloNi) in Chile. Data were available from 27 tests of more than 1,200 clones from crosses among 12 *E. nitens* and 8 *E. globulus* parents (28 full-sibs families in total). Chapter 4 examined volume and MOE for the hybrid of *P. patula x P. tecunumanii* (PATxTEC) in South Africa. Data were available from hybrid breeding programs from two different companies, seedling for SAFCOL, with a total of 40 and 97 full-sib families from *P. patula x P. tecunumanii* variety high-elevation and P. *patula x P. tecunumanii* variety low-elevation, respectively, and 26 full-sib families with clonal propagation (271 clones) of *P. patula x P. tecunumanii* variety low-elevation for Sappi company.

5.2 Simulation Results

In Chapter 2, results were reported for a simulation study examining the precision of variance component estimates deriving from mixed model analysis and REML estimation, using linear models and underlying genetic structures representative of possible genetic architecture of hybrid forest trees for the trait of 8-year volume. Hybrid breeding with trees is difficult and slow, and hybrid mating designs often involve a relatively small number of parents. One focus of Chapter 2 was the inspection of the frequency of serious errors, either over-estimation or under-estimation of variance components. A set of 12 scenarios, with different genetic architecture and different numbers of parents (12, 24, or 48) and

crosses per parent (an average of 2, 4, or 8), was examined with simulated hybrid progeny in a set of four field trials, with each scenario repeated over 1000 iterations.

The results from Chapter 2 indicated that the use of a full linear model specifying GHA effects for each parent species, SHA effects, and genetic x test interaction effects for simulated hybrid progeny populations produced, in general, very accurate mean variance component estimates over 1000 iterations for all scenarios tested. As expected for variance component analysis, there was a substantial variation of variance component estimates for the random effects GHA_A, GHA_B, and SHA over the 1000 iterations. In other words, the variance component estimates for a given scenario can differ widely from iteration to iteration (where one iteration represents one set of experiments in the "real world").

In general, variance component estimates for all random effects were close to the target variance values set up for each scenario, and hybrid breeders using the estimates to plan breeding strategy would not be seriously mistaken. SHA variance estimates were consistently stable across all scenarios since this parameter was estimated with a large number of degrees of freedom (i.e., a large number of crosses relative to the number of parents). For GHA variance estimates, serious under- or over-estimates were observed in some scenarios. Factors associated with more frequent serious errors were low "true" underlying GHA variance, high true SHA variance, smaller numbers of parents tested, and smaller numbers of crosses per parent. If the mating design included 24 parents tested in at least 4-crosses, it is unlikely to find a high proportion of severe errors in the estimates of GHA variance. In the "worst-case" scenario (with low GHA variance, high SHA variance, 48 parents with 2 crosses per parent), a zero GHA variance estimate was observed in 11% of the iterations. In this case, a breeder would observe zero GHA variance for one of the parental species, even though the true GHA variance in the population is non-zero (small, but possibly still economically important).

5.3 Genetic Parameters for GloNi

The results from Chapter 3 for the GloNi (*E. nitens x E. globulus*) hybrid in Chile showed that large genetic gains in volume could be achieved with this variety in the two breeding zones of Arauco Company, and similarly, substantial gains are possible for all the wood traits analyzed (basic density, pulp yield, and specific consumption). There is substantial GHA variance associated with the NIT parents in both zones, making it important to test a large number of NIT parents in future interspecific crosses in Arauco and Valdivia zones. Another important result was that *E. nitens* pure species GCA for volume is a good predictor *of E. nitens* hybrid GHA volume in both breeding zones. This means that it is possible to select parents for future interspecific crosses based on the performance of their pure species progeny.

Somewhat surprisingly, the estimated GHA variance for tree volume for GLO parents was zero in both breeding zones and for the wood trait pulp yield in Valdivia. The simulation results from Chapter 2 indicated that a low number of parents might give a zero GHA variance estimate even if true GHA variance is not zero. In the *Eucalyptus nitens x E. globulus* (GloNi) population, there were 12 *E. nitens* and 8 *E. globulus* parents, with each parent in an average of 2 and 3 crosses, respectively. The simulation results show that for a low number of crosses tested, there is around a 5% to 13% chance of getting a false zero or near zero estimates of GHA variance, and with a low number of parents, there is around a 3% to 9% chance of a false zero or near zero estimates of GHA variance. Perhaps the combination of both a low number of parents and a low number of crosses (which was not tested in the simulation) would increase the probability of obtaining zero GHA variance estimates for one of the species.

Assuming that the zero GHA_{GLO} variance estimates for volume and pulp yield in Valdivia are true, it should be preferable to use fewer GLO fathers in future interspecific crosses, or to select those parents primarily for their merits in other traits, such as wood properties or rooting abilities, a very important trait for commercial propagation of hybrid clones.

The pure species GCA is also a useful predictor of the hybrid GHA for wood properties, but this appears to be influenced by the environment. Specifically, for the Arauco zone, the *E. globulus* pure species GCA is a remarkable predictor of hybrid GHA; thus, globulus parents could be selected based on their GCA or GHA values for hybrid crosses. However, this relationship did not hold for Valdivia, where the GCA-GHA correlation was zero. On the other hand, in Valdivia, the *E. nitens* GCA is a good predictor of the GHA for wood properties but a poor predictor in Arauco. In all likelihood, any GloNi clones produced in the breeding program would be tested in both zones, so some weight should be placed on wood properties of both pure species *E. globulus* and *E. nitens*.

The epistasis effect was very high for volume in both breeding zones, accounting for 42% to 56%, which is near to values reported for other hybrid *Eucalyptus* populations. For wood properties, zero epistasis variance was found. The epistasis variance for volume is an important part of the additional genetic gain for volume possible with clonal propagation of GloNi.

In summary, a good strategy for future GloNi interspecific crosses in the current hybrid breeding program is : in Arauco, the selection of the best *E. nitens* mothers for volume gain, and the best *E. globulus* parents for wood properties, by evaluation of the parents either in pure species or hybrid testing; in Valdivia, the selection of the best *E. nitens* for volume gain either by pure or hybrid progeny testing, and the best *E. globulus* parents from hybrid tests (GHA values).

5.4 Genetic Parameters for PATxTEC

Data from two varieties of the *Pinus patula x P. tecunumanii* hybrid were available (from low-elevation and high-elevation varieties of *P. tecunumanii*) and were analyzed separately. For the variety of *Pinus patula x P. tecunumanii* high (PATxTECH), there was substantial *P. patula* GHA variation for tree volume and a zero estimate for TECH GHA variance. For the wood trait MOE, GHA variance was found for both PAT and TECH. There was also a good correlation of pure species GCA and hybrid GHA for *P. patula* parents. Thus, more PAT

parents should be tested in future interspecific crosses with TECH to increase volume gain, and the PAT parents could be selected based on pure species progeny performance. Simulation results from Chapter 2 indicated that low numbers of parents and high SHA variance could both increase the frequency of false zero GHA variance estimates. Regarding the zero GHA variance estimate for TECH parents, there were only 11 TECH parents tested, and there was a significant estimated SHA variance for this dataset. However, assuming that the zero GHA variance for volume from TECH is true, the selection of TECH parents should be based mainly on MOE performance, or perhaps could be based on other traits, such as *Fusarium* resistance or frost tolerance.

For the *P. patula x P. tecunumanii* low (PATxTECL) hybrid, independent populations were tested by different companies, and the genetic parameters from the two datasets were very similar. For tree volume, GHA variance from TECL parents was more important in both populations (clonal population of Sappi and seedling population of SAFCOL) than GHA variance from PAT parents. This suggests that future crosses should emphasize selecting more TECL parents to maximize genetic gain for volume. There were moderate positive correlations between GCA and GHA of both *P. patula* and *P. tecunumanii* that should facilitate parent selection for new interspecific crosses, and these selections could be made based on the performance in pure or hybrid field tests results of each parental species.

In the clonal population of PATxTECL (Sappi), 29% of the total genetic variation was attributed to epistasis, which is concordant with other conifer studies reported in the literature. However, these values were not statistically significant (high SE) but should be considered in future data analysis since they are not small values.

APPENDICES

Appendix A. Summary of true, sample, and REML variance estimates.

Table A 1. Summary table of true (σ^2), sample ($\hat{\sigma}^2$) and REML ($\hat{\hat{\sigma}}^2$) variance estimates for all simulated parameters per scenario, with their respective SD across scenarios. Npar and NC were abbreviations of Number of Parents and Number of Crosses for each species, respectively.

Scenario 1	$Npar_A$= 24; $Npar_B$= 12; NC_A= 4; NC_B= 8 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	62.5	125	31.25	15.625	31.25	2109.375
$\hat{\sigma}^2$	126.4 ± 36.3	62.2 ± 25.5	125.1 ± 18.4	31.1 ± 26.2	15.6 ± 13	31.3 ± 0.6	2109.9 ± 22.3
$\hat{\hat{\sigma}}^2$	126.7 ± 50.9	61.4 ± 36	125.4 ± 26.2	31.5 ± 8.7	15.7 ± 6.6	31.6 ± 7.6	2109.7 ± 22.3

Scenario 2	$Npar_A$= 24; $Npar_B$= 24; NC_A= 4; NC_B= 4 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	62.5	125	31.25	15.625	31.25	2109.375
$\hat{\sigma}^2$	122.3 ± 36	62.9 ± 18.5	124.1 ± 18	31.2 ± 25.4	15.4 ± 13.5	31.2 ± 0.5	2109.3 ± 21.9
$\hat{\hat{\sigma}}^2$	123.1 ± 51.8	63.4 ± 31.8	125.1 ± 29	30.5 ± 9	15.6 ± 6.6	31.1 ± 8.1	2109.3 ± 22.1

Scenario 3	$Npar_A$= 24; $Npar_B$= 48; NC_A= 4; NC_B= 2 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	62.5	125	31.25	15.625	31.25	2109.375
$\hat{\sigma}^2$	122.7 ± 35.8	62.1 ± 12.8	123.5 ± 18.3	31.4 ± 24.7	15.4 ± 12.8	31.3 ± 0.3	2109.8 ± 22.3
$\hat{\hat{\sigma}}^2$	124.6 ± 55	62.2 ± 35.8	125.5 ± 35	31.1 ± 8.7	15.7 ± 8.1	30.7 ± 9.7	2109.8 ± 22.5

Scenario 4	$Npar_A$= 24; $Npar_B$= 12; NC_A= 4; NC_B= 8 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	62.5	250	31.25	15.625	62.5	1953.125
$\hat{\sigma}^2$	124.8 ± 36.6	62.3 ± 26.3	247 ± 35.8	30.6 ± 25.7	15.5 ± 13.4	62.6 ± 1.2	1954.5 ± 20.1
$\hat{\hat{\sigma}}^2$	124.5 ± 60.3	61.6 ± 46.2	249.2 ± 48.6	31.2 ± 10.6	15.7 ± 7.5	62.6 ± 10.2	1954.5 ± 20.2

Scenario 5	$Npar_A$= 24; $Npar_B$= 24; NC_A= 4; NC_B= 4 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	62.5	250	31.25	15.625	62.5	1953.125
$\hat{\sigma}^2$	123.1 ± 35.8	62.6 ± 18.9	246.6 ± 35.8	31.4 ± 26.1	15.2 ± 11.8	62.5 ± 0.9	1953 ± 20.5
$\hat{\hat{\sigma}}^2$	121.9 ± 62.9	63.3 ± 43.9	249.8 ± 52.5	30.8 ± 10.2	16 ± 7.8	61.9 ± 11.3	1953.1 ± 20.6

Scenario 6	$Npar_A$= 24; $Npar_B$= 48; NC_A= 4; NC_B= 2 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	62.5	250	31.25	15.625	62.5	1953.125
$\hat{\sigma}^2$	124.2 ± 37.3	62.7 ± 13.1	247.6 ± 35.8	30.5 ± 26.4	15.4 ± 13.2	62.5 ± 0.6	1952.6 ± 20.8
$\hat{\hat{\sigma}}^2$	127.3 ± 67.8	67.9 ± 52.3	245.8 ± 58.4	31 ± 10.4	15.5 ± 10.4	62.3 ± 13.1	1952.5 ± 21.1

Scenario 7	$Npar_A$= 24; $Npar_B$= 12; NC_A= 4; NC_B= 8 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	187.5	125	31.25	46.875	31.25	1953.125
$\hat{\sigma}^2$	124.7 ± 34.1	184.4 ± 77.7	123.3 ± 18.5	32.2 ± 27.4	48.9 ± 40.4	31.2 ± 0.6	1953.5 ± 19.9
$\hat{\hat{\sigma}}^2$	126.6 ± 51.1	185.2 ± 93.8	123.7 ± 25.9	30.9 ± 8.9	47.6 ± 14.2	31.5 ± 7.5	1953.5 ± 20

Scenario 8	$Npar_A$= 24; $Npar_B$= 24; NC_A= 4; NC_B= 4 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	187.5	125	31.25	46.875	31.25	1953.125
$\hat{\sigma}^2$	125.5 ± 38.5	184 ± 54.1	123.3 ± 18	31 ± 23.9	47.3 ± 38.8	31.3 ± 0.5	1953.8 ± 20.5
$\hat{\hat{\sigma}}^2$	125.2 ± 54.4	184.3 ± 73.2	124.2 ± 28.5	31 ± 9.2	46.9 ± 11.6	31.2 ± 8.2	1953.9 ± 20.8

Scenario 9	$Npar_A$= 24; $Npar_B$= 48; NC_A= 4; NC_B= 2 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	187.5	125	31.25	46.875	31.25	1953.125
$\hat{\sigma}^2$	123.8 ± 36.4	185.9 ± 38.4	124 ± 18.2	31.4 ± 25.4	46.1 ± 38.8	31.3 ± 0.3	1953.2 ± 19.8
$\hat{\hat{\sigma}}^2$	124.9 ± 58.5	186.1 ± 66.7	125 ± 33.3	30.4 ± 9.5	46.7 ± 11	31.2 ± 9.4	1953.2 ± 20

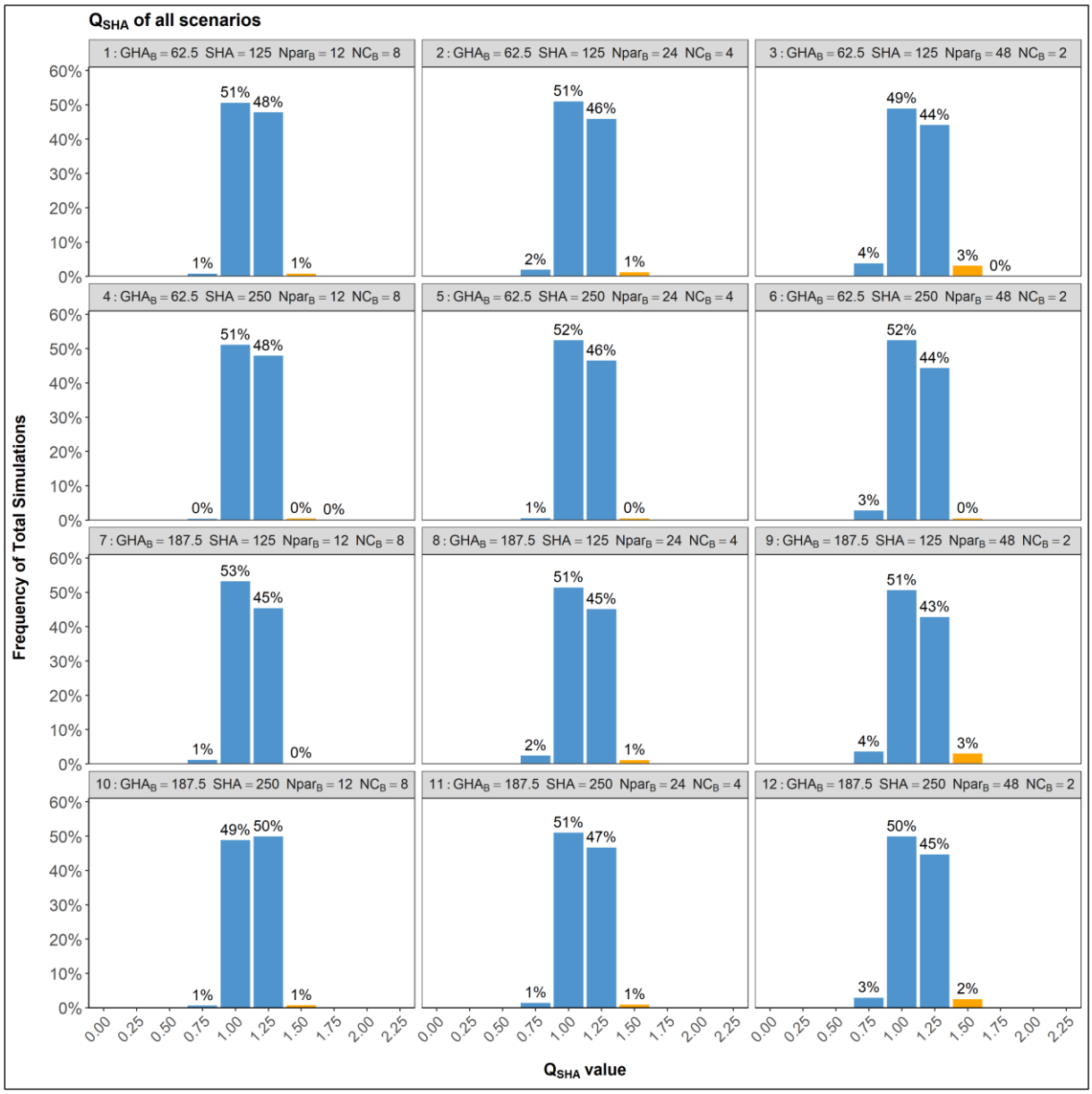
Scenario 10	$Npar_A$= 24; $Npar_B$= 12; NC_A= 4; NC_B= 8 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	187.5	250	31.25	46.875	62.5	1796.875
$\hat{\sigma}^2$	128.3 ± 37.2	184.1 ± 78	248 ± 35.5	30.1 ± 24.4	48.1 ± 39.8	62.5 ± 1.2	1797.3 ± 18.3
$\hat{\hat{\sigma}}^2$	127.3 ± 63.1	180.7 ± 97.4	251.8 ± 49.7	31.4 ± 9.8	47.7 ± 15.1	62.3 ± 10.1	1797.4 ± 18.5

Scenario 11	$Npar_A$= 24; $Npar_B$= 24; NC_A= 4; NC_B= 4 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	187.5	250	31.25	48.875	62.5	1796.875
$\hat{\sigma}^2$	125.8 ± 37.2	187.7 ± 54.2	248.2 ± 35.7	30.4 ± 25.3	48.7 ± 38.6	62.5 ± 1	1795.9 ± 18.5
$\hat{\hat{\sigma}}^2$	124.2 ± 65.4	187.3 ± 82.4	251.5 ± 54.2	31 ± 10.3	46.3 ± 13.2	62.8 ± 11.2	1795.8 ± 18.7

Scenario 12	$Npar_A$= 24; $Npar_B$= 48; NC_A= 4; NC_B= 2 || Variance components ± SD						
Parameter	GHA_A	GHA_B	SHA	GHA_AxE	GHA_BxE	SHAxE	Err
σ^2	125	187.5	250	31.25	48.875	62.5	1796.875
$\hat{\sigma}^2$	124 ± 36.8	188.7 ± 39.1	249.4 ± 37.5	31.5 ± 25.7	45.6 ± 36.4	62.5 ± 0.6	1796 ± 18.9
$\hat{\hat{\sigma}}^2$	123.3 ± 74.2	194 ± 84	249.3 ± 64.2	31.1 ± 11.3	46.3 ± 14.3	62.4 ± 12.8	1796 ± 19

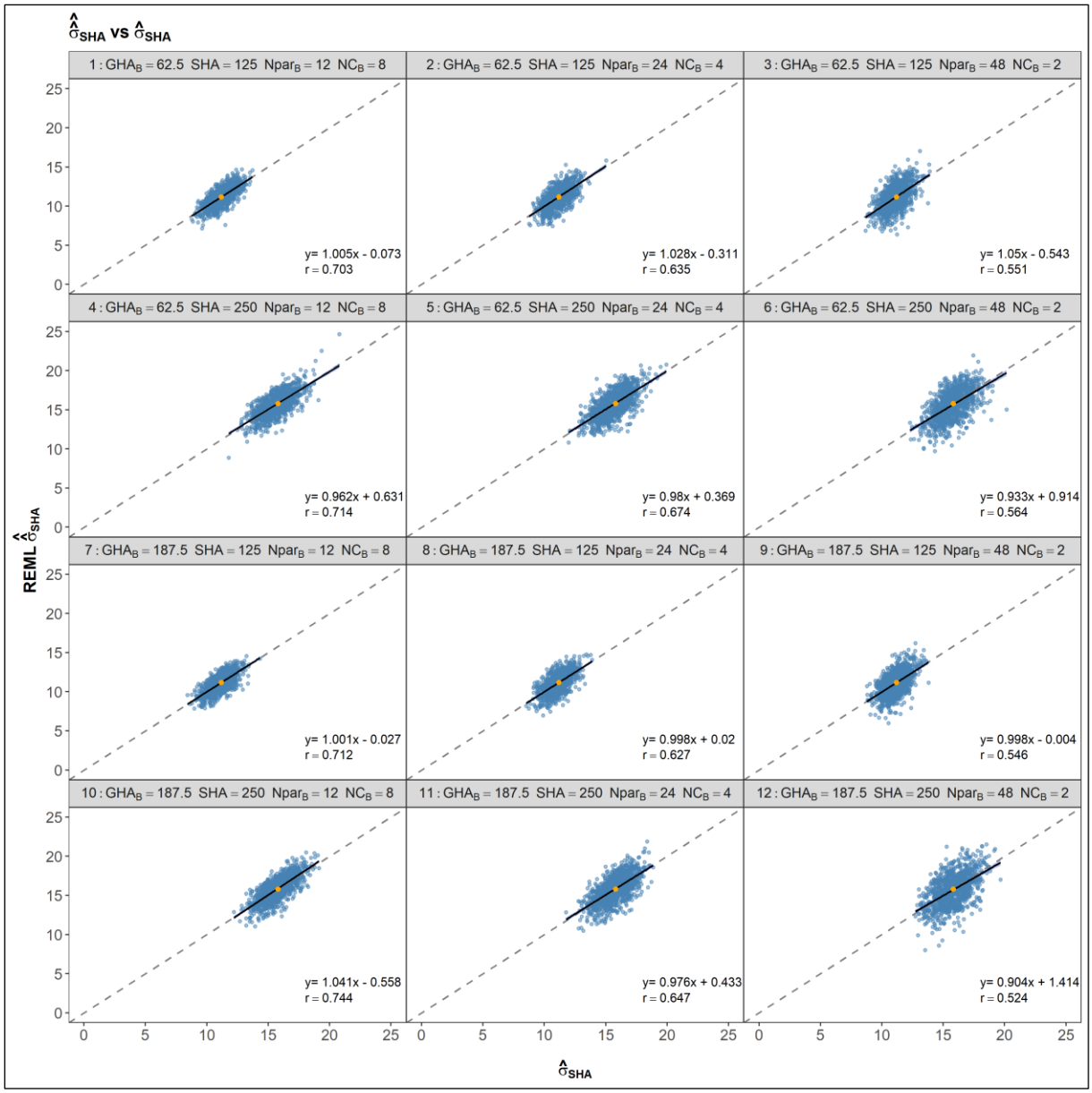
Appendix B. Histogram of Q-values for SHA.

Figure B 1. Histogram of Q-values for SHA. $Q = \hat{\hat{\sigma}}/\sigma$, where $\hat{\hat{\sigma}}$ = REML estimate of σ_{SHA}, and σ = the true parameter value for σ_{SHA}. Histograms show results from 1000 simulations of 12 scenarios with different genetic parameters and mating designs. Bars in red indicate high deviation of true values, yellow bars middle deviations of true value and blue bars precise estimates. $Npar_B$ and NC_B were abbreviations of Number of Parents and Number of Crosses for species B.



Appendix C. Scatterplot of predicted REML SHA variance vs sample SHA variance

Figure C 1. Scatterplots of predicted REML $\hat{\hat{\sigma}}_{SHA}$ vs $\hat{\sigma}_{SHA}$, where the orange dot represents the true σ value of each scenario. Blue dots represent the result from 1000 simulations of 12 scenarios. The segmented line is the reference line of equivalence in each scenario. $Npar_B$ and NC_B were abbreviations of Number of Parents and Number of Crosses for species B.



Appendix D. Diallel of crosses

I. Diallel with 24 parents of species B and 12 parents of species A, with an average of 4 crosses for species A and 8 crosses for species B.

Species B \ Species A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		x			x					x	x	x				x	x					x	x	
2	x				x	x				x							x	x					x	x
3		x				x	x					x	x					x	x					
4			x				x	x					x	x					x	x				
5			x	x				x	x					x	x	x				x	x			
6			x						x	x				x							x	x	x	
7	x				x					x	x					x	x					x	x	
8	x	x			x	x					x	x					x	x	x					x
9		x				x	x	x				x	x					x	x					
10			x			x							x	x				x		x		x	x
11			x	x				x	x					x	x					x	x			
12				x					x	x					x	x					x	x		x

II. Diallel with 24 parents of species A and B, with an average of 4 crosses each.

Species B \ Species A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	x	x											x		x									x
2	x														x	x								
3		x	x													x	x							
4			x	x													x	x						
5				x	x	x											x		x					
6					x														x	x				
7						x	x													x	x			
8							x	x													x	x		
9								x	x	x											x		x	
10									x														x	x
11	x									x	x													x
12	x	x									x	x												
13	x		x									x	x	x										
14			x	x									x											
15				x	x									x	x									
16					x	x									x	x								
17					x		x									x	x	x						
18							x	x									x							
19								x	x									x	x					
20									x	x									x	x				
21									x		x									x	x	x		
22											x	x									x			
23												x	x									x	x	
24													x	x									x	x

III. Diallel with 24 parents of species B and 48 parents of species A, with an average of 4 crosses for species A and 2 crosses for species B

Species B \ Species A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	x												x											
2	x												x											
3			x												x									
4				x												x								
5					x												x							
6						x												x						
7							x												x					
8								x												x				
9									x												x			
10										x												x		
11											x												x	
12												x												x
13	x												x											
14		x												x										
15			x												x									
16				x												x								
17					x												x							
18					x												x							
19							x												x					
20								x												x				
21									x												x			
22										x												x		
23											x												x	
24												x												x
25	x												x											
26		x												x										
27			x												x									
28				x												x								
29					x												x							
30						x												x						
31							x												x					
32								x												x				
33									x												x			
34									x												x			
35											x												x	
36												x												x
37	x												x											
38		x												x										
39			x												x									
40				x												x								
41					x												x							
42						x												x						
43							x												x					
44								x												x				
45									x												x			
46										x												x		
47											x												x	
48												x												x

Appendix E. Arauco NIR Model Parameters

Table E 1. NIR models for wood properties in Arauco Company, Calibration, and Validation.

	Calibration				Validation		
Model	**n samples**	**Factors**	**RMSCV**	**R^2_C**	**n samples**	**RMSEP**	**R^2_P**
Basic density (km/m^3)	365	8	24.4	0.8	64	23.1	0.79
Pulp Yield (%)	365	8	1.18	0.8	64	1.29	0.77
Specific Wood Consumption (m^3/Adt)	365	7	0.25	0.81	64	0.22	0.84

R^2_C is the coefficient of determination for calibration and RMSECV root mean square error of cross-validation. R^2_P is the coefficient of determination for prediction performance, and RMSEP is the root mean square error of prediction.

Appendix F. *Pinus tecunumanii* provenances used in SAFCOL and Sappi

Table F 1. *Pinus tecunumanii* var Low and High elevation provenances sources, used for interspecific crosses in SAFCOL and Sappi

		Company	
Variety	Provenance	SAFCOL	Sappi
TECH	Chempil	x	
TECH	El Carrizal	x	
TECH	Las Trancas	x	
TECH	Montebello	x	
TECH	San Jerónimo	x	
TECL	Culmi		x
TECL	Mountain Pine Ridge		x
TECL	San Esteban	x	
TECL	San Francisco	x	
TECL	Villa Santa	x	x
TECL	Yucul		x

Appendix G. Single-Site genetic parameters on SAFCOL and Sappi field tests

Table G 1. Single-site genetic parameters of standardized volume (stVol) and MOE analysis in SAFCOL field test.

Species	Site	Trait	σ^2_{plot}	σ^2_{pxt}	σ^2_{pat}	σ^2_{tec}	σ^2_{err}	σ^2_{phen}	$H^2_b \pm SE$
PATxTECH	SAF 1	MOE	0.74	0.16	0.40	0.17	6.36	7.82	0.23 ± 0.09
PATxTECH	SAF 2	MOE	0.69	0.00	0.42	0.04	10.55	11.7	0.08 ± 0.05
PATxTECH	SAF 3	MOE	1.18	1.23	0.46	0.00	6.55	9.43	0.62 ± 0.21
PATxTECL	SAF 4	MOE	-	0.44	0.46	0.74	4.24	5.87	0.7 ± 0.21
PATxTECL	SAF 1	MOE	0.66	0.56	0.34	0.21	6.28	8.05	0.41 ± 0.08
PATxTECL	SAF 2	MOE	3.61	0.38	0.24	0.03	14.72	18.98	0.11 ± 0.08
PATxTECL	SAF 3	MOE	2.19	0.92	0.67	0.11	7.42	11.29	0.46 ± 0.11
PATxTECH	SAF 1	stvol	346.56	163.25	90.78	0.00	1533.42	2134	0.39 ± 0.13
PATxTECH	SAF 5	stvol	0.00	90.53	211.65	0.00	1856.34	2159	0.36 ± 0.15
PATxTECH	SAF 2	stvol	0.45	254.09	165.93	0.00	1738.19	2159	0.62 ± 0.18
PATxTECH	SAF 3	stvol	73.75	138.50	111.76	0.00	1799.66	2124	0.37 ± 0.12
PATxTECL	SAF 4	stvol	-	107.69	0.00	81.78	1941.36	2131	0.28 ± 0.14
PATxTECL	SAF 1	stvol	249.26	146.72	46.68	45.22	1638.39	2126	0.36 ± 0.08
PATxTECL	SAF 2	stvol	9.39	36.56	51.47	138.40	1934.30	2170	0.24 ± 0.07
PATxTECL	SAF 3	stvol	48.02	125.67	35.93	51.88	1873.40	2135	0.32 ± 0.07

Table G 2. Single-site genetics parameters of stVol on Sappi field tests

Site	Trait	σ^2_{plot}	σ^2_{Clw}	σ^2_{pxt}	σ^2_{pat}	σ^2_{tecl}	σ^2_{err}	σ^2_{phen}	$H^2_b \pm SE$
SAP 1	stvol	659.96	199.43	209.244	0	174.531	1197.74	2440.9	0.24 ± 0.07
SAP 2	stvol	283.8	471.249	1.698	54.19	124.686	1437.887	2373.51	0.27 ± 0.04