ABSTRACT

MARTINS, GUSTAVO SALGADO. Reproductive Systems in *Acacia crassicarpa* Leveraging Breeding Opportunities for Accelerated Delivery of Genetic Gains in Important Quantitative Forestry Traits (Under the direction of Dr. Gary R. Hodge and Dr. Juan J. Acosta).

Acacia crassicarpa is an important tree species in Southeast Asia. Its vigorous growth, resistance to pests and diseases, and good bole form are valuable features for tropical forestry. There are still constraints to maximizing the delivery of genetic gains. Although feasible, controlled pollination is impractical for advancing breeding populations requiring a huge effort to produce more than a few crosses per year. Vegetative propagation is possible only with juvenile ortets. Finally, there is limited knowledge about the wood properties of the species.

This study applied genomics to unveil mating dynamics, reconstruct full-sib families for trial testing, and evaluate the potential of genomic selection to accelerate genetic gains for the species. Additionally, important wood and pulping properties were characterized, and within-tree patterns of variation were investigated for efficient phenotyping strategies.

One season of reproduction in a seed orchard was characterized by genotyping 84,315 seedlings with forty-two SNP markers. The analysis indicated that 67.8% of the seed collected was derived from male parents within the orchard. The average number of male parents per open-pollination family was 50, with the average dominant male proportion equal to 23%. The reproductive success of genotypes was highly variable. Cumulative combined male-female reproductive success indicated that 50% of parents produced 80% of the seed. Spatial analysis showed a moderate-high spatial correlation between the mother tree's distance to the pollen source and its proportion within the open-pollination family, with a rapid decay with distance increase. On the wood properties, forty trees were selected for destructive sampling at age 50 months and assessed for wood density, kraft pulp yield, α-cellulose, carbohydrate composition, lignin content,

and syringil/guayacil ratio. The mean whole-tree disc basic density was 481 kg m^{-3} and screened kraft pulp yield was 53.8%. Ground-level sampling could reliably predict the whole-tree property for basic density, pulp yield, and glucose content. Using NIR predictions to indirectly measure basic density, correlations with whole-tree density values were sufficient to allow accurate ranking and efficient selection of genotypes in a breeding program context.

Genetic control of quantitative traits was studied with full-sib multi-environmental progeny trials measured at 36 months. The traits were predominantly controlled by additive effects, with heritability ranging between 0.09 for survival to 0.45 for basic density. The genetic correlation across sites was high for all traits showing the low impact of genotype-by-environment interaction. The trait-trait correlation showed that straightness was independent of other traits, survival was only correlated with mean annual increment, and growth traits were highly correlated among themselves. Surprisingly, wood basic density was highly correlated with growth traits.

Integrating genomic methods into the breeding program of *A. crassicarpa* made possible the construction of genomic models with excellent breeding value prediction ability. Genomic models outperformed pedigree-based models for all traits, and accurate individual tree selection resulted in valuable gains for all units of selection: individual trees for generation advancement or within-family genomic selection for deployment with family forestry. The average gain from the within-family genomic selection practiced with a selection intensity of 10% on the top five ranked families was 18%, demonstrating the opportunity to effectively double the gains achieved in a generation compared to deployment based on family means.

The results of this study provide fundamental answers to the conservation and breeding efforts of *A. crassicarpa*, allowing the maximization of the delivery of genetic gains.

Reproductive Systems in *Acacia crassicarpa* Leveraging Breeding Opportunities for Accelerated Delivery of Genetic Gains in Important Quantitative Forestry Traits

by
Gustavo Salgado Martins

A thesis 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

2023

APPROVED BY:

Gary R. Hodge Co-Chair of Advisory Committee	Juan J. Acosta Co-Chair of Advisory Committee
Ross W. Whetten	James Holland

DEDICATION

This dissertation is dedicated to Mariana, my lovely wife, mother of my sons, and my companionship in this adventure we call our lives. You give me strength every day at your side. Thank you! I love you.

To Henrique, for being this strong boy traveling the world with dad and doing better than myself against all our challenges together. I love you! To Daniel, for being the joy of our experience in the USA. I love you too! I love you all.

To my family in Brazil. My dad, my mother, my sisters, and my in-laws.

BIOGRAPHY

Gustavo Salgado Martins was born in Ituiutaba, Minas Gerais, Brazil on the 11th March 1984. His father, Luiz Antônio de Souza Martins, is an agronomist, and his mother, Lícia Maria Salgado, is a historian. DNA analysis (myheritage.com) revealed ancestral lineages connecting him strongly to Portugal and Italy via his birthplace in Minas Gerais, Brazil, a state with a long history of European migration into South America.

Flag of Minas Gerais state, Brazil

51% of its DNA traces back to the Iberian peninsula. The region of Iberia, which encompasses Spain and Portugal, has been historically shaped by multiple civilizations and distinct populations, from the ancient Iberian tribes to modern-day Spanish and Portuguese people. In 1492, Christopher Columbus set sail for the Americas, kicking off the age of exploration and conquest in the New World. Iberian explorers spread across the Americas, parts of Africa, the Indian subcontinent, the Portuguese islands of Azores and Madeira, and the Spanish Canary Islands, leaving their genetic mark on these areas.

15% of Gustavo's DNA traces to the Sardinian people from his maternal grandmother's Italian side. The people, language, and culture of Sardinia are distinct from those of neighboring regions; Sardinians preserve an ancient Neolithic European genetic legacy thanks to the relative isolation of the population. Communities of Sardinian descent are present in other parts of Europe, including mainland Italy, Germany, and the United Kingdom. People of Sardinian descent can also

be found in Brazil and Argentina. Another 10% leads to the Baltics from his maternal grandfather's side. The Baltic ethnic group is Indo-European from an ethnolinguistic point of view. Many small ethnic groups in the area, including those of Baltic origin, eventually merged into the larger groups, including the Prussians and those today considered to hail from the Baltic states: Estonia, Latvia, and Lithuania.

African blood is in the pot, with 11.5% matching the Maghreb (which encompasses modern-day Morocco, Algeria, Tunisia, and Libya), tracing their roots to the Berbers of antiquity. North Africans share some genetic traits with their Southern European neighbors, though the cultures and religions of the region were affected significantly by Arabization and the rise of Islam after the 6th century CE. Colonization and waves of immigration over the last two centuries have created a large North African diaspora.

Another 6.5% of his DNA is Nigerian. Nigeria has the largest population in Africa, and its people belong to over 250 sub-ethnic groups, the Esan and the Yoruba. Nigeria was ruled by various kingdoms and tribal rulers and became a colony of England in the 19th century. Many people of African descent in the Americas have some Nigerian ethnicity.

Lastly, but most proudly, indigenous American blood is there for 3.5% of Gustavo's DNA. The population of South America includes several indigenous groups with pre-Columbian roots. Most were hunter-gatherers living in socially and culturally complex communities. European colonial expansion in the 1500s brought the decimation and displacement of many indigenous groups due to unfamiliar diseases and land disputes. Still, the legacy of indigenous cultures continues to reverberate across the continent, with many place names in Brazil originating in various indigenous languages. "Ituiutaba," Gustavo's birthplace, is an indigenous name from the "Caiapó" people.

He attended elementary and middle school at Anglo Ituiutaba, moving to his mother's hometown Lavras, MG, Brazil, in 1999, where he attended high school at Gammon institute. He would soon move to the “Escola Superior de Agricultura de Lavras,” founded in 1908 by North-American Presbyterian missionaries under the motto “Dedicated to the Glory of God and Human Progress.” He commenced his superior education in 2004, studying Biology and obtaining a B.Sc. in Forest Engineering in 2010.

Gustavo was granted an assistantship and introduced to forest research during his first term in undergrad forestry school, working with silvicultural and management of *Eucalyptus* spp., teak (*Tectona grandis*), and Australian red cedar (*Toona ciliata*). He worked with professors and researchers from Votorantim Metais and Bela Vista Florestal enterprises in Minas Gerais state, getting interested in tree breeding. In 2010 he continued his studies and was accepted into the Genetics and Plant Breeding Program of the Federal University of Lavras, where he obtained an M.Sc. in 2012, working with competition effects in eucalypts.

In May 2012, Gustavo joined as a tree breeding specialist in the Research Department of PLANTAR, a forestry enterprise based in Belo Horizonte, MG, Brazil, with operations for the production of green pig iron and structural wood for construction. In 2014 he entered the pulp industry, accepting a tree breeding specialist job at VERACEL, a large pulp mill in Bahia state, Brazil.

In 2016, Gustavo moved abroad to work as a tree breeder for APRIL Group, a large vertically integrated pulp industry in Pangkalan Kerinci, Riau, Indonesia. APRIL Group joined Camcore in 2018. Working for the company, he had the opportunity to start a Doctoral Program in Forestry. In 2019 the family moved to Raleigh, North Carolina, USA. Upon completing his Ph.D., he will return to Indonesia for the APRIL Asia tree improvement program leader position.

ACKNOWLEDGMENTS

Thank North Carolina State University. I couldn't be better welcomed in this great country of yours. I want to thank my co-advisor, Dr. Gary Hodge, for allowing me to pursue my Ph.D. at NC State. You have made everything possible since our first conversation, welcomed me at Camcore, and taught me a good amount of quantitative genetics and forestry. It was a privilege to have your advice. I thank my co-advisor Dr. Juan Jose Acosta, for teaching me so much during the course. You have been a friend and improved our time in the United States.

Thank you to all Camcore friends and colleagues, Willi Woodbridge, Romeo Jump, Dr. Robert Jetton, Dr. Juan Luis Lopez, Robert McGee, Dr. Luis Ibarra and family, Dr. Colin Jackson, Dr. Evandro Tambarussi, Leonida Cherotich, Gina Zabala, and Mindoro Razoki. I enjoyed having you as my friend.

I thank Dr. Ross Whetten and Dr. James Holland for agreeing to be part of my thesis committee and for all your contributions and time dedicated to my Ph.D. project. I appreciate it. To the professors Dr. Fikret Isik and Dr. Steve McKeand for all the lessons. To the friends Dr. Marcelo Mollinari and Dr. Gabriel Gesteira for all your time and help with the bioinformatics. To the professors Dr. Magno Antônio Patto Ramalho, Dr. Flávia Maria Avelar Gonçalves, Dr. Marcos Deon Vilela de Resende, Dr. César Brasil, Dr. Dário Grattapaglia, and Dr. Teotônio Assis. Without learning from you, I would never be here. I'm standing on the shoulders of giants.

I am grateful to the APRIL Asia group, Anderson Tanoto, Dr. Mukesh Sharma, Wong Ching Yong, Muhammad Yuliarto, Dr. Alvaro Durán, François Van Deventer, and Adriano Almeida for supporting me in achieving this degree. Without you, all of this would have never happened. I am proud to be part of this project. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Acacia crassicarpa is one of Southeast Asia's most important tree species for forestry. This research explored a breeding population employing state-of-the-art technology to improve the species for timber production. Breeding efforts with the species are relatively recent, and biological and technological constraints remain to maximize its genetic potential for forestry.

The manuscript was organized with an introduction followed by chapters prepared for submission to research journals as original articles. The first chapter presents a general introduction to the species biology and tree improvement strategies. It contains important concepts on which the following chapters develop. Some redundancy between the text in the introduction and chapters will then be present.

Chapter Two deals with mating dynamics in open-pollination systems revealed by large-scale pedigree reconstruction. It is currently with the referees of a reputable research journal.

Chapter Three is a dive into the species' wood properties. The article was published in the journal *Forests* and named "Wood and pulping properties variation of *A. crassicarpa* and sampling strategies for accurate phenotyping."

Chapter Four provides information on the genetic control of quantitative traits in full-sib families tested in multi-environmental trials for a large collection of important traits for the species.

Chapter Five demonstrates the improvement in genetic modeling and increased genetic gains delivered by genomic selection.

CHAPTER 1

Introduction to *Acacia crassicarpa*

1.1 Introduction

Acacia crassicarpa A.Cunn. ex Benth. (thick-podded salwood, red wattle, Papua New Guinea red wattle, northern wattle, brown salwood) is an important tree species in Southeast Asia (Nambiar and Harwood 2014; Harwood and Nambiar 2014), where hundreds of thousands of hectares of planted forests are supported by advancements in silviculture (Mendham and White 2019) and genetic improvement (Harwood *et al*. 2015; Nirsatmanto and Sunarti 2019). The name "*crassicarpa"* comes from Greek in reference to its thick pod. In Sumatra, Indonesia, flowering starts 30-36 months after planting, while the seed is produced in abundance after four years. Seeds mature 5-6 months after flowering. The main flowering season is January-February, but light flowering may occur as late as September. The peak fruiting season is June-July. However, there is variation between locations and from year to year.

The species is naturally distributed in the humid tropical North-eastern Australia and New Guinea island. Provenance testing of wild collections started in the 1990s to explore its potential as an alternative species to *A. mangium*. With good bole form, vigorous growth on poorly drained acidic soils, and resistance to pests and diseases, *A. crassicarpa* is a valuable option for tropical plantation forestry (Turnbull *et al*. 1998; Midgley and Turnbull 2003). Small foresters and large integrated pulp and paper industries have *A. crassicarpa* as an important component of their wood supply (Martins *et al*. 2020; Nambiar *et al*. 2018).

Acacia Mill. is a large woody genus of *Fabaceae* Lindl. family. The Australian acacias, formerly placed in *Acacia* subgenus *Phyllodineae* (Gibson *et al*. 2011; Miller *et al*. 2011), consists of 1,012 described species, of which at least a third have been introduced in different parts of the

world (Richardson *et al*. 2011). About nine of these Australian species extend northwards into New Guinea Island (Skelton 1987), including the five most important species and provenances for Southeast Asia's tropical forestry: *A. aulacocarpa*, *A. auriculiformis*, *A. crassicarpa*, *A. leptocarpa*, and *A. mangium*.

Floral morphology is a conserved trait, with small tubular flowers grouped into spherical or elongated flower heads, with pollen presented on the inflorescence surface (Stone *et al*. 2003). Pollen grains are clustered into a composite unit, named a "polyad," which provides an efficient means of dispersal via pollinators and is a key component of the pollination efficiency of all acacias. Their flowers have a small cup-shaped stigma into which only one polyad can fit. There are always fewer ovules per ovary than pollen grains per polyad, so one polyad from a single pollination event can potentially fertilize all the ovules (Kenrick and Knox 1982). The stigmas are also distributed over the surface of the flower heads, opening gradually and asynchronously. A generalist entomophilous pollination syndrome is identified and provides accessible floral rewards to almost any insect visitor (Bernhardt 1989). The release of floral scent often enhances the recruitment of insects just before pollen release, and visual advertisement is maximized by opening flowers (Stone 2003; Kenrick 2003). Acacia inflorescences show no apparent adaptations for capturing wind-borne pollen, in contrast to typical wind-pollinated species, which have feathery stigmas and aerodynamic features that aid in capturing the pollen grains. Acacia's pollen-to-ovule ratio is also very low, compatible with dependence on animal pollen vectors (Gibson *et al*. 2011).

To a certain extent, outcrossing rates are highly promoted by dichogamous flowering and self-incompatibility (George *et al*. 2008). Partial self-compatibility and intraspecific variation in self-compatibility seem relatively common, with some ability to reproduce by selfing (Millar *et al*. 2012). Among Australian *Acacia* species, selfing varies from self-incompatible to completely self-

compatible and autogamous (Kenrick and Knox 1989; Morgan *et al*. 2002). Regarding inbreeding depression, a significantly reduced early growth was verified at 18 months of age in *A. mangium*, with lower height and diameter at breast height of self-fertilized individuals compared to outcrossed, demonstrating the importance of minimizing selfing in operational seed production (Harwood *et al*. 2004). For *A. crassicarpa*, a study with progenies of single trees collected from two populations (15 trees near Coen, Queensland, Australia, and 21 trees near Wemenever, Papua New Guinea) showed high outcrossing rates with little variation between them. Differences between single and multi-locus estimates and fixation indices between populations were not significant, indicating that these populations followed Hardy-Weinberg equilibrium frequencies with a low level of inbreeding (Moran *et al*. 1989).

A. crassicarpa is a hermaphrodite with strictly protogynous flowers where the stigma is receptive before the anthers produce pollen. Although technically feasible, controlled pollination of acacias is difficult as the individual flowers are very small and difficult to emasculate. Also, the percentage of flowers that develop into pods is typically less than 5%, even if pollen is not limiting. In the breeding program of April Asia, even recommended methods such as "inflorescence pollination" (Griffin *et al*. 2010) require a huge effort to produce more than a few crosses yearly. Therefore, controlled pollination is not practical for advancing breeding populations which require hundreds of crosses for each breeding cycle. Breeding populations have been bred by open pollination. Typically, progeny trials are established, and family and within-family selections are made to convert them into "seedling seed orchards" (SSO). After thinning and a subsequent general flowering, the open-pollinated seed is collected from the best trees of the better-performing families to establish second-generation progeny trials. Clonal breeding orchard (CBO) establishment combines selections captured by marcotting or grafting and is also practiced to

increase selection intensity and genetic gain over that achievable from SSOs (Harwood *et al*. 2015).

In principle, clonal forestry should deliver the greatest genetic gain at any stage of a breeding program by deploying both additive and non-additive gain (Zobel 1993). It requires mass-production of juvenile propagules of selected trees as implemented for *Eucalyptus* (Rezende *et al*. 2014). For *A. crassicarpa*, vegetative propagation is cost-effective, but is possible only with juvenile ortets. With current vegetative propagation techniques, by the time phenotypes can be measured for assessment of genetic value, the selected tree can only be propagated by grafting. *Family forestry* can be defined as the commercial deployment of half-sib (open-pollinated) or full-sib (control-pollinated) family blocks. Family forestry can be done using either seedlings or vegetative propagules, such as rooted cuttings. In Indonesia, family forestry using vegetative propagules (sometimes called clonal family forestry) has been successfully employed with acacia (Wong and Yuliarto 2014). Selected female parents tested using progeny in open-pollination families will produce a limited amount of improved seeds in seed orchards. This seed can be sown and vegetatively propagated to bulk-up nursery hedges supporting the deployment of cuttings. The main limitation of this strategy with open-pollination families is that selection is based only on female additive effects. Still, it has been widely used due to the inefficiency of controlled crossing. The advantage of full-sib family deployment over open-pollinated family deployment is significant. The full parental control and potential capture of gain from specific combining ability can capitalize on the full-sib family genetic value, delivering larger genetic gains and higher uniformity (White *et al*. 2007).

The past decades have seen considerable progress in forest tree genomics research, and cost-effective genotyping platforms of breeder-friendly single nucleotide polymorphism markers

(SNP) are now available for all mainstream plantation forest trees (Grattapaglia 2022). Various tools and methods have been successfully employed in applied forest tree improvement programs (Whetten *et al*. 2023). Parentage analysis using different genotyping platforms, statistical methods, and software has been reported for several animal and plant species (Jones and Ardren 2003; Jones *et al*. 2010). The value of parentage reconstruction is to allow full pedigree determination of open-pollinated seed lots. It may be the only way full parental control can be applied to estimate genetic parameters for species where controlled crossing is inefficient or impossible. For *A. crassicarpa*, using molecular markers to reconstruct pedigrees could be of great value in determining full-sib families and managing kinship in breeding populations. With full-sib family models, breeders can better model the genetics of complex quantitative traits and their association with phenotypic variation. Quantitative genetics researchers are often interested in partitioning observed phenotypic variance into causal genetic and environmental components. Typically, this involves the use of linear mixed models that can handle unbalanced data and complex experimental designs to deal with a high level of field heterogeneity due to differences in soil type, fertility, water holding capacity, etc., and a high number of genetic treatments (Isik *et al*. 2017), common to tree breeding trials with large tree plot sizes.

Furthermore, several published forest tree studies show that utilizing genomic information for modeling the genetic contribution on the phenotypic expression matches or surpasses the performance of phenotypic selection for growth and wood properties traits (Grattapaglia 2022). Also, realized relationship matrices constructed from genome-wide SNP markers can accurately measure relatedness (Hayes *et al*. 2009), capturing the deviation from the expected value due to the Mendelian sampling of alleles during sexual recombination. Within-family genomic selection is feasible, allowing the breeder to explore the genetic variance present within a full-sib family

equivalent to ½ of the additive genetic variance plus ¾ of the dominance genetic variance (Falconer and MacKay 1996) and deliver additional genetic gains beyond family mean selection (Lynch and Walsh 1998). This strategy would be useful with tree species with difficult vegetative propagation, as demonstrated for loblolly pine, with accurate within-family genomic selection (Walker *et al*. 2021). In addition, upon the breakthrough development of techniques for mass-scale production of juvenile propagules, superior individual genotypes can be sourced with genomic selection, maximizing the genetic gain obtained per breeding generation with clonal forestry.

1.2 References

Bernhardt, P. (1989) The floral biology of Australian *Acacia*. In: Advances in legume biology. Proceedings of the Second International Legume Conference, St. Louis, Missouri, 23-27 June 1986; Stirton, CH, Zarucchi JL, Eds.; Missouri Botanical Garden, St Louis, U.S.A, 263–281.

Falconer DS, Mackay TF (1996) Introduction to Quantitative Genetics (4th edition). 480 p. Longman Group Ltd.

George N, Byrne M, Yan G (2008) Mixed mating with preferential outcrossing in *Acacia saligna* (Labill.) H. Wendl. (*Leguminosae: Mimosoideae*). Silvae Genetica, 57, 139–145. https://doi.org/10.1515/sg-2008-0021

Gibson MR, Richardson DM, Marchante E, Marchante H, Rodger JG, Stone GN *et al*. (2011) Reproductive biology of Australian Acacias: Important mediator of invasiveness? Diversity and Distributions, 17:5, 911-933. https://doi.org/10.1111/j.1472-4642.2011.00808.x

Grattapaglia D (2022) Twelve Years into Genomic Selection in Forest Trees: Climbing the Slope of Enlightenment of Marker Assisted Tree Breeding. Forests, 13, 1554. https://doi.org/10.3390/f13101554

Griffin AR, Vuong TD, Harbard JL, Wong CY, Brooker C, Vaillancourt RE (2010) Improving controlled pollination methodology for breeding *Acacia mangium* Willd. New Forests, 2010, 40:2, 131–142. https://doi.org/10.1007/s11056-010-9188-x

Harwood CE, Ha HT, Tran HQ, Butcher PA, Williams ER (2004) The effect of inbreeding on early growth of *Acacia mangium* in Vietnam. Silvae Genetica, 53, 65–69. https://doi.org/10.1515/sg-2004-0012

Harwood, CE, Nambiar EKS (2014) Sustainable plantation forestry in South East Asia. ACIAR Technical reports No. 84. Australian Centre for International Agricultural Research: Canberra. 100 pp.

Harwood, CE, Hardiyanto EB, Wong CY (2015) Genetic improvement of tropical Acacias: achievements and challenges. Southern Forests, 77, 11-18. https://doi.org/10.2989/20702620.2014.999302

Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research, 91:1, 47–60. https://doi.org/10.1017/S0016672308009981

Isik F, Holland J, Maltecca C (2017) Genetic data analysis for plant and animal breeding. Cham, Switzerland: Springer International Publishing.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. Molecular Ecology, 12, 2511–2523. https://doi.org/10.1046/j.1365-294X.2003.01928.x

Jones, AG, Clayton, MS, Paczolt KA, Ratterman, NL (2010) A practical guide to methods of parentage analysis. Molecular Ecology Resources, 10, 6–30. https://doi.org/10.1111/j.1755-0998.2009.02778.x

Kenrick J, Knox RB (1982) Function of the polyad in reproduction of *Acacia*. Annals of Botany, 50, 721–727. https://www.jstor.org/stable/42758606

Kenrick J, Knox RB (1989) Quantitative analysis of self-incompatibility in trees of seven species of *Acacia*. Journal of Heredity, 80, 240–245. https://doi.org/10.1093/oxfordjournals.jhered.a110842

Kenrick J (2003) Review of pollen-pistil interactions and their relevance to the reproductive biology of *Acacia*. Australian Systematic Botany, 16, 119–130. https://doi.org/10.1071/SB02005

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits (1st ed.). Sunderland: Sinauer Associates.

Martins GS, Yuliarto M, Antes R, Sabki, Prasetyo A, Unda F, Mansfield SD *et al*. (2020) Wood and Pulping Properties Variation of *Acacia crassicarpa* A.Cunn. ex Benth. and Sampling Strategies for Accurate Phenotyping. Forests, 11, 1043. https://doi.org/10.3390/f11101043

Mendham DS, White DA (2019) A review of nutrient, water and organic matter dynamics of tropical Acacias on mineral soils for improved management in Southeast Asia. Australian Forestry, 82:sup1, 45-56. https://doi.org/10.1080/00049158.2019.1611991

Midgley SJ, Turnbull JW (2003) Domestication and use of Australian Acacias: case study of five important species. Australian Systematic Botany, 16, 89-102. https://doi.org/10.1071/SB01038

Millar MA, Byrne M, Nuberg I, Sedgley M (2012) High levels of genetic contamination in remnant populations of *Acacia saligna* from a genetically divergent planted stand. Restoration Ecology, 20:2, 260-267. https://doi.org/10.1111/j.1526-100X.2010.00758.x

Miller JT, Murphy DJ, Brown GK, Richardson DM, González-Orozco CE (2011) The evolution and phylogenetic placement of invasive Australian *Acacia* species. Diversity and Distributions, 17:5, 848–860. https://doi.org/10.1111/j.1472-4642.2011.00780.x

Moran GF, Muona O, Bell JC (1989) Breeding Systems and Genetic Diversity in *Acacia auriculiformis* and *A. crassicarpa*. Biotropica, 21:3, 250-256. https://doi.org/10.2307/2388652

Morgan A, Carthew SM, Sedgley M (2002) Breeding system, reproductive efficiency and weed potential of *Acacia baileyana*. Australian Journal of Botany, 50, 357–364. https://doi.org/10.1071/BT01088

Nambiar EKS, Harwood CE (2014) Productivity of Acacia and eucalypt plantations in Southeast Asia. 1. Bio-physical determinants of production: opportunities and challenges. International Forestry Review, 16:2, 225–248. https://doi.org/10.1505/146554814811724757

Nambiar EKS, Harwood CE, Mendham DS (2018) Paths to sustainable wood supply to the pulp and paper industry in Indonesia after diseases have forced a change of species from Acacia to eucalypts. Australian Forestry, 81:3, 148-161. https://doi.org/10.1080/00049158.2018.1482798

Nirsatmanto A, Sunarti S (2019) Genetics and Breeding of Tropical Acacias for Forest Products: *Acacia mangium*, *A. auriculiformis* and *A. crassicarpa*. In: Advances in Plant Breeding Strategies: Industrial and Food Crops; Al-Khayri J, Jain S, Eds.; Springer, Cham: Springer Nature, Switzerland, 2019; Volume 6, pp. 3-28.

Rezende GDSP, Resende MDV, Assis TF (2014) Eucalyptus breeding for clonal forestry. In: Fenning T (ed). Challenges and Opportunities for the World's Forests in the 21st Century. Springer Netherlands: Dordrecht, pp 393–424.

Richardson DM, Carruthers J, Hui C, Impson FAC, Robertson MP, Rouget M *et al*. (2011) Human-mediated introductions of Australian Acacias – a global experiment in biogeography. Diversity and Distributions, 17:5, 771–787. https://doi.org/10.1111/j.1472-4642.2011.00824.x

Skelton DJ (1987) Distribution and ecology of Papua New Guinea Acacias. In: Australian Acacias in developing countries. Proceedings of an international workshop held at Gympie,

Australia, 4-7 August 1986; Turnbull, J.W., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 16., Canberra, Australia, pp. 38-44.

Stone GN, Raine NE, Prescott M, Willmer PG (2003) Pollination ecology of Acacias (*Fabaceae*, *Mimosoideae*). Australian Systematic Botany, 16, 103–118. https://doi.org/10.1071/SB02024

Turnbull JW, Midgley SJ, Cossalter C (1998) Tropical Acacias planted in Asia: an overview. In: Recent developments in Acacia planting. Proceedings of an international workshop held in Hanoi, Vietnam, 27-30 October 1997; Turnbull, J.W., Crompton, H.R., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 82., Canberra, Australia, pp. 155-160.

Walker TD, Cumbie WP, Isik F (2021) Single-step genomic analysis increases the accuracy of within-family selection in a clonally replicated population of *Pinus taeda* L. Forest Science, 68:37–52. https://doi.org/10.1093/forsci/fxab054

Whetten RW, Jayawickrama KJS, Cumbie WP, Martins GS (2023) Genomic Tools in Applied Tree Breeding Programs: Factors to Consider. Forests, 14, 169. https://doi.org/10.3390/f14020169

White TL, Adams WT, Neale DB (2007) Forest genetics. Wallingford: CAB International.

Wong CY, Yuliarto M (2014) Deployment of acacias in short rotation pulpwood plantation. In: Acacia 2014 “Sustaining the Future of Acacia Plantation Forestry” International Conference, IUFRO Working Party 2.08.07: Genetics and Silviculture of Acacias, Hue, Vietnam, 18–21 March 2014, Compendium of Abstracts.

Zobel BJ (1993) Clonal Forestry in the Eucalypts. In: Ahuja, MR., Libby, W.J. (eds) Clonal Forestry II. Springer, Berlin, Heidelber.

CHAPTER 2

Pollination and Mating Dynamics Unveiled by Orchard-Wide Pedigree Reconstruction in *Acacia crassicarpa*

2.1 Introduction

Acacia crassicarpa A.Cunn. ex Benth. is an important tree species in Southeast Asia (Nambiar and Harwood 2014; Harwood and Nambiar 2014), where hundreds of thousands of hectares of planted forests are supported by advancements in silviculture (Mendham and White 2019) and genetic improvement (Harwood *et al*. 2015; Nirsatmanto and Sunarti 2019). The species is naturally distributed in the humid tropical North-eastern Australia and New Guinea island. Provenance testing of wild collections started in the 1990s to explore its potential as an alternative species to *A. mangium*. With good bole form, vigorous growth on poorly drained acidic soils, and resistance to pests and diseases, *A. crassicarpa* is a valuable option for tropical plantation forestry (Turnbull *et al*. 1998; Midgley and Turnbull 2003). Small forestry companies and large integrated pulp and paper industries have *A. crassicarpa* as an important component of their wood supply (Martins *et al*. 2020; Nambiar *et al*. 2018).

Acacia Mill. is a large woody genus of *Fabaceae* Lindl. family. The Australian acacias, formerly placed in *Acacia* subgenus *Phyllodineae* (Gibson *et al*. 2011; Miller *et al*. 2011), consisting of 1,012 described species, of which at least a third have been introduced in different parts of the world (Richardson *et al*. 2011). About nine of these Australian species extend northwards into New Guinea Island (Skelton 1987), including the five most important species and provenances for Southeast Asia's tropical forestry: *A. aulacocarpa*, *A. auriculiformis*, *A. crassicarpa*, *A. leptocarpa*, and *A. mangium*.

Floral morphology is a conserved trait, with small tubular flowers grouped into spherical or elongated flower heads, with pollen presented on the inflorescence surface (Stone *et al*. 2003). Pollen grains are clustered into a composite unit, named a "polyad," which provides an efficient means of dispersal via pollinators and is a key component of the pollination efficiency of all acacias. Their flowers have a small cup-shaped stigma into which only one polyad can fit. There are always fewer ovules per ovary than pollen grains per polyad, so one polyad from a single pollination event can potentially fertilize all the ovules (Kenrick and Knox 1982). The stigmas are also distributed over the surface of the flower heads, opening gradually and asynchronously. A generalist entomophilous pollination syndrome is identified and provides accessible floral rewards to almost any insect visitor (Bernhardt 1989). The release of floral scent often enhances the recruitment of insects just before pollen release, and visual advertisement is maximized by opening flowers (Stone 2003; Kenrick 2003). Acacia inflorescences show no apparent adaptations for capturing wind-borne pollen, in contrast to typical wind-pollinated species, which have feathery stigmas and aerodynamic features that aid in capturing the pollen grains. Acacia's pollen-to-ovule ratio is also very low, compatible with dependence on animal pollen vectors (Gibson *et al*. 2011).

To a certain extent, normalized outcrossing rates are highly promoted by dichogamous flowering and self-incompatibility (George *et al*. 2008). Partial self-compatibility and intraspecific variation in self-compatibility seem relatively common, with some ability to reproduce by selfing (Millar *et al*. 2012). Among Australian *Acacia* species, selfing varies from self-incompatible to completely self-compatible and autogamous (Kenrick and Knox 1989; Morgan *et al*. 2002). Butcher *et al*. (2004), studying the breeding system of *A. mangium*, observed high self-pollination rates in the offspring of trees derived from natural populations with inbreeding, while a seed orchard based on highly outcrossing Papua New Guinea (PNG) populations produced outcrossed

seed. However, high outcrossing rates can be maintained in populations with less than half the allelic richness of the more genetically diverse PNG populations as long as a high proportion of heavily flowering trees and sufficient pollinator activity are present. In regards to inbreeding depression, a significantly reduced early growth was verified at 18 months of age in *A. mangium*, with lower height and diameter at breast height of self-fertilized individuals compared to outcrossed, demonstrating the importance of minimizing selfing in operational seed production (Harwood *et al*. 2004). For *A. crassicarpa*, a study with progenies of single trees collected from two populations (15 trees near Coen, Queensland, Australia, and 21 trees near Wemenever, PNG) showed high outcrossing rates with little variation between them. Differences between single and multi-locus estimates and fixation indices between populations were not significant, indicating that these populations followed Hardy-Weinberg equilibrium frequencies with a low level of inbreeding (Moran *et al*. 1989).

A. crassicarpa flowers are hermaphrodites with strictly protogynous flowers where the stigma is receptive before the anthers produce pollen. Although technically feasible, controlled pollination of acacias is difficult as the individual flowers are very small and difficult to emasculate. Also, the percentage of flowers that develop into pods is typically less than 5%, even if pollen is not limiting. In the breeding program of April Asia, even recommended methods such as "inflorescence pollination" (Griffin *et al*. 2010) require a huge effort to produce more than a few crosses per year. Therefore, controlled pollination is not practical for advancing breeding populations which require hundreds of crosses for each breeding cycle. Breeding populations have been bred by open pollination.

Typically, progeny trials are established, and then family and within-family selection is done to convert them into "seedling seed orchards" (SSO). After thinning and a subsequent general

flowering, the open-pollinated seed is collected from the best trees of the better-performing families to establish second-generation progeny trials. Clonal breeding orchard (CBO) establishment, which brings together elite selections captured by marcotting or grafting, is also practiced to increase selection intensity and genetic gain over that achievable from SSOs (Harwood *et al*. 2015).

The successful use of molecular markers to perform parentage analysis using different genotyping platforms, statistical methods, and software has been reported for several animal and plant species (Jones and Ardren 2003; Jones *et al*. 2010). In our context, using molecular markers to reconstruct pedigrees could be of great value in determining full-sib families and managing kinship in breeding populations. It also allows for better modeling of the genetics of complex traits and their association with phenotypic variation. Furthermore, pedigree reconstruction can help breeders understand the mating dynamics of orchards and give insights into the open-pollination family structure. Therefore, this study aimed to examine a pollination season of an *A. crassicarpa* orchard with parentage analysis provided by large-scale use of single nucleotide polymorphism (SNP) markers to unveil recombination patterns and reconstruct full-sib families for breeding trials establishment.

2.2 Material and Methods

2.2.1 Germplasm and DNA-Sampling

The orchard studied is an elite seed orchard (ESO) established by the breeding program of April Asia in August 2012 on a mineral soil compartment in Riau province, Sumatra Island, Indonesia (Figure 2.1). Silvicultural practices followed the company's standard operating procedure for soil cultivation, fertilization, and weeding. The orchard design consisted of single-tree plots of 15 genetic treatments replicated in 15 randomized blocks, totaling 225 plots of 100

square meters each (10m x 10m) with a total area of 2.25 hectares. The ESO compartment is located on the outskirts of a commercial nursery, adjacent to several eucalyptus and acacia experiments and seed orchards. It comprises the ESO and a neighboring CBO in a eucalyptus plantation buffer surrounded by access roads and a natural riverine forest.

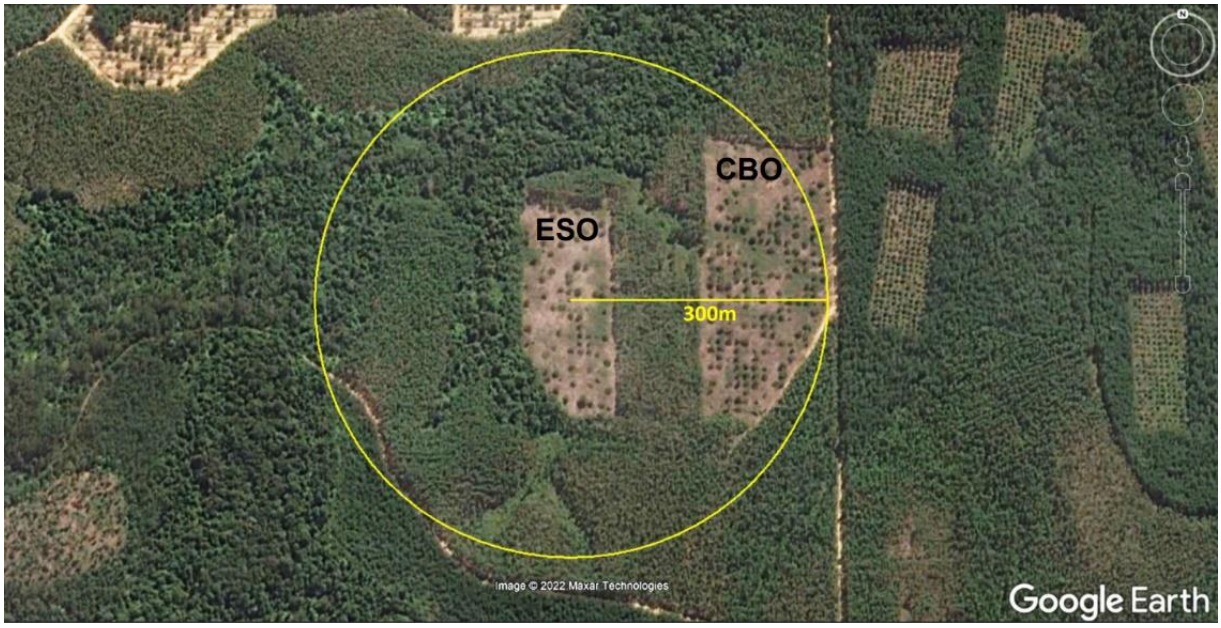


Figure 2.1. An aerial overview of the orchard's compartment highlights the Elite Seed Orchard (ESO) and the Clonal Breeding Orchard (CBO). Imagery date: July 8, 2015.

In early 2017, a flowering census with leaf sampling was performed on all seventy-seven standing trees in the ESO, and all live trees in the neighboring CBO, to confirm the genetic identity of mother trees. Pod harvesting occurred from March to September, peaking in June and July. All pods produced from individual open-pollinated trees over the reproductive season were collected and processed, generating 50 open-pollination families. A selection of 24 of these families, balanced by maternal representation and seed weight, was sown in January 2018. After germination and plantlet development in the nursery, 84,315 seedlings were produced and sampled for DNA analysis and parentage reconstruction.

2.2.2 DNA Extraction and SNP Genotyping

A proprietary panel of 42 SNP markers was developed from an internal genome sequencing effort to provide a high-throughput and cost-effective genotyping platform capable of accurately assigning *A. crassicarpa* genotypes. SNPs were selected based on minor allele frequency (MAF), broad distribution across the genome, including representation of each of the 13 chromosomes, and uniqueness of sequence surrounding the SNP. Orion Biosains (Puchong, Malaysia) designed, optimized, and performed assays utilizing the LGC Array Tape genotyping platform. A set of 173 trees was genotyped to validate the panel, comprising the 77 ESO mother trees and 96 progenies. Orchard tree samples were genotyped in triplicate, and nursery samples were genotyped in quadruplicate, with 25,830 total SNP genotyping events. DNA was extracted from dried leaf samples using Orion Biosains' proprietary automated DNA extraction protocols. For the parentage reconstruction campaign, 84,315 orchard offspring seedlings were sampled and genotyped following the abovementioned procedures.

2.2.3 Parentage Reconstruction

Seedlings were sampled from known mothers and assigned to candidate fathers using the paternity analysis of the software CERVUS version 3.0.7 (Marshall *et al*. 1998; Kalinowski *et al*. 2007). Estimates of expected heterozygosity (H_E), observed heterozygosity (H_O), polymorphism information content (PIC), and average non-exclusion probability (NE) were calculated for each locus based on allele frequencies of the offspring genotyped. The simulation of paternity analysis, utilized by the software to estimate the resolving power of the markers given their allele frequencies and to estimate critical values of the log-likelihood statistics, was performed with the following parameters: 10,000 individuals; 58 candidate parents; 0.95 proportion of sampled parents; 0.01 proportion of mistyped loci; and the option to test for self-fertilization. All other

parameters of the paternity analysis followed the software default. Offspring assigned to a candidate father, given the known mother, with a trio confidence level equal to or above 95% were considered for downstream analysis.

2.3 Results

2.3.1 Pedigree Reconstruction and Family Structure

Seventy-seven standing trees were counted in the census performed in early 2017, with flowers recorded in all of them. The genetic identity analysis pointed out 18 genotypes in the ESO and 40 genotypes in the neighbor CBO. The ESO layout is presented in Figure 2.2, with the 24 open-pollinated trees with seedlots selected for this study highlighted in green.

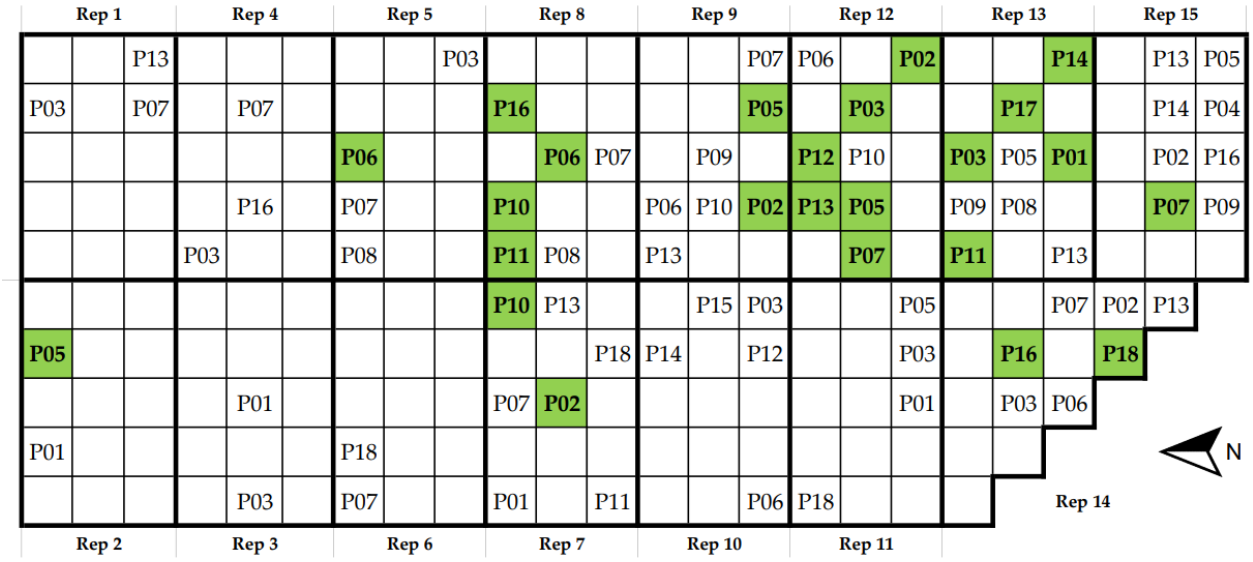


Figure 2.2. Elite Seed Orchard design with 77 trees comprising 18 unique parents (P) confirmed by genetic identity analysis. The 24 cells highlighted in green represent the open-pollinated trees with a seedlot sown in this study.

The genotype call rate for the validation set was 98.4%. As measured by concordance across replicated genotyping assays of the same samples, genotype call accuracy was 99.2%

(Appendix A1). The allele frequency analysis summary is presented in Appendix A2. There were 347 samples with less than half the number of markers called, and following the software preset, these samples were disregarded for downstream analysis. The remaining 83,968 seedlings, successfully genotyped with more than 21 SNP, showed a 0.973 average proportion of SNP called. The mean minor allele frequency (MAF) was 0.425, the mean observed heterozygosity (H_O) was 0.508, the mean expected heterozygosity (H_E) was 0.482, and the mean polymorphic information content (PIC) was 0.365. The summary of the paternity analysis is presented in Table 2.1.

Table 2.1. Paternity analysis output by male groups representing progeny assigned as full-sibs of ESO parents (Group 1), ESO females pollinated by a CBO male parent (Group 2), and progeny with an ESO female and unknown male parent (Group 3). Group 3 was subdivided into progeny with a positive female LOD score (Group 3.1) and a negative female LOD score (Group 3.2).

Group	Male Parent Assignment	Number of Progeny	% of Total
1	ESO male	56,938	67.8%
2	CBO male	10,764	12.8%
3	Not Assigned	16,266	19.4%
3.1	Female LOD ≥ 0	15,878	18.9%
3.2	Female LOD < 0	388	0.5%
Total		83,968	100%

Of all progeny tested, 67.8% were assigned to an ESO male parent, i.e., a full-sib of ESO parents (Group 1), and 12.8% were assigned to a CBO male parent, being a full-sib of an ESO female and a known male parent present in the neighbor CBO (Group 2). Overall, 80.6% of the progeny analyzed were assigned to an ESO female and one of the 58 (18 ESO + 40 CBO) male parents with a known genotype. The analysis did not assign any known male parent with a 95% confidence level for the remaining 19.4% of the progeny (Group 3). In parentage analysis, a positive LOD score means that the candidate parent is more likely to be the true parent than not

the true parent. A true parent almost always has a positive LOD score. The Group 3 breakdown showed that 18.9% of the total progeny was not assigned a male parent but had a positive LOD score for the female parent. Only 0.5% showed a negative female LOD score, indicating that 99.5% of the progeny analyzed matched the known female.

The genetic composition of the open-pollination families (OP) derived from the 24 selected mother trees is presented in Figure 2.3.

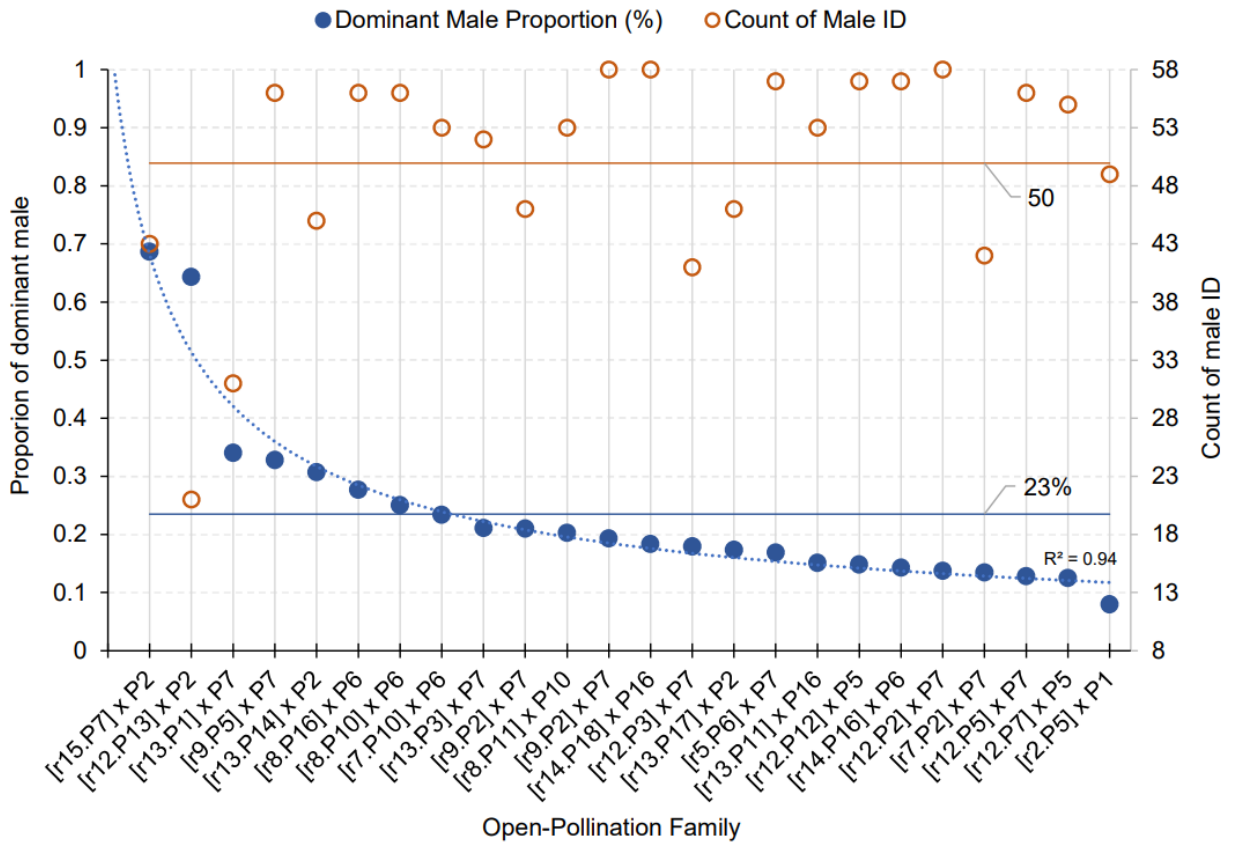


Figure 2.3. Open pollination family structure as revealed by parentage analysis. Families were named as a key "[r . P_f] x P_m" with the female tree within brackets with "r" for the replicate number, "P_f" for the female parent, "P_m" for the dominant male parent. The decay of the dominant male proportion fits a power function ($y = 0.77x^{-0.58}$) with $R^2 = 0.94$ with an average dominant male proportion of 23%. The average number of males per family was 50.

The full diallel with parents and reciprocal crosses matrix is presented in Appendix A3. A total of 236 full-sib families were determined, accounting for 73% of the 324 families with 18 parents ($N^2 = 18^2 = 324$). Ignoring the crosses' directionality and selfings, i.e., merging reciprocal crosses discounting selfings, 136 full-sib families were determined, accounting for 89% of the 153 total possible (N(N-1)/2 = (18*17/2) = 153). If we consider only families with at least 30 progeny, enough to establish a progeny trial with replication, 95 families were determined, representing 62% of the total possible families. A 3-D representation of the crossing structure of the population derived from the pedigree reconstruction is presented in Figure 2.4.

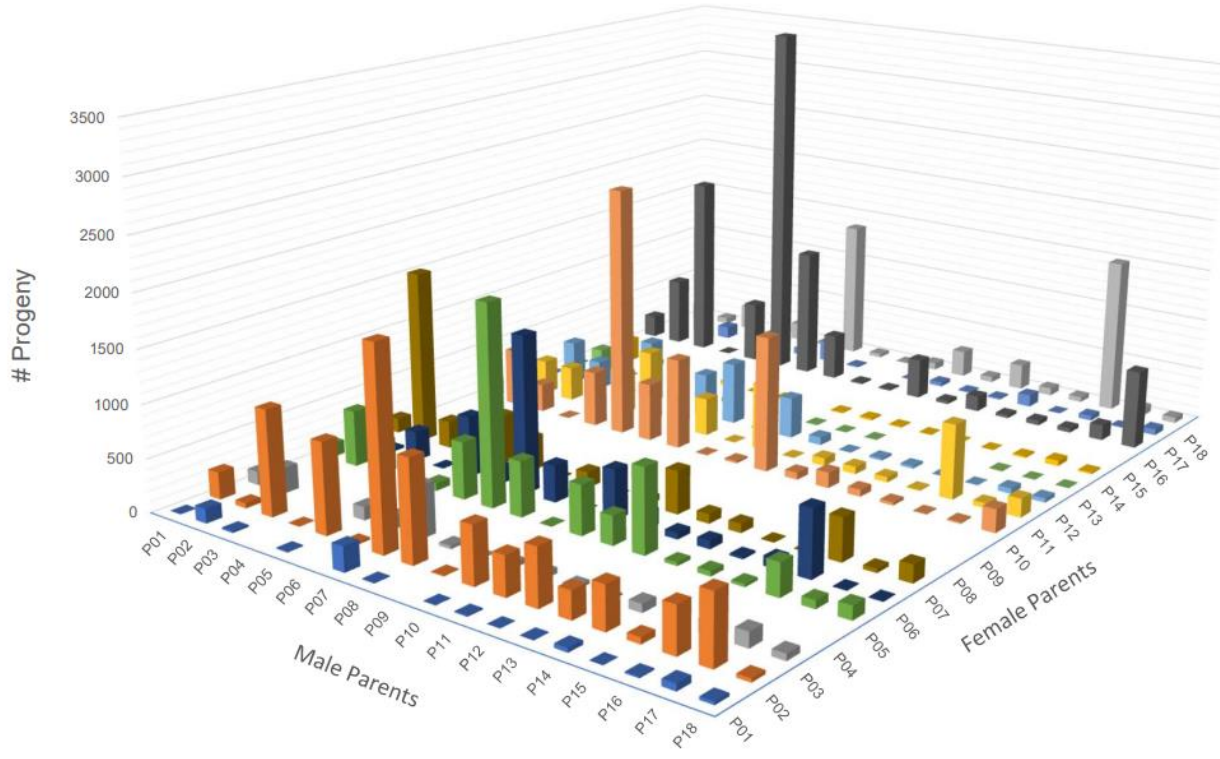


Figure 2.4. 3-D representation of the full diallel revealed by the pedigree reconstruction. Each column represents a full-sib family, with its height the number of progeny. No seedlots of female parents P04, P08, P09, and P15 were sowed. Thus, they are blank in the chart above on the female side.

2.3.2 Reproductive Success

The elite seed orchard progeny output is summarized in Table 2.2. The total seed weight harvested in the 2017 seed production season was approximately 15 kg. Exploring the entire seed collection was impossible, and a 4.6 kg sample of the orchard's total production was sowed to match the genotyping budget. The Pearson correlation coefficient between the female progeny count and the harvest seed weight was 0.88. The correlation between the number of ramets of each genotype with the male progeny count was 0.23.

Table 2.2. ESO progeny output of the 2017 reproductive season. For each parent, the number of ramets, total seed weight harvested and sowed, and the number of progeny counted as female and male given by the pedigree reconstruction are presented, along with their respective selfing rates and progeny count rank.

Parent	Ramets	Weight (g)		Progeny Count				
		Harvest	Sow	Female	Rank	Self (%)	Male	Rank
P01	5	105	38	687	13	0.4	1,562	12
P02	5	3,200	635	11,546	2	0.4	5,424	3
P03	9	716	174	2,407	10	0	4,954	4
P04	1	0	0	0	15	0	25	18
P05	6	715	449	8,577	4	0.8	4,533	5
P06	6	2,891	446	8,221	5	2.0	8,236	2
P07	10	302	279	5,394	8	0.2	10,231	1
P08	3	54	0	0	16	0	4,332	6
P09	3	0	0	0	17	0	49	17
P10	4	1,723	589	9,823	3	0.2	2,743	9
P11	3	586	451	6,512	7	0	3,339	8
P12	2	382	210	4,138	9	0.1	1,774	10
P13	7	43	35	359	14	0	1,132	14
P14	3	77	36	746	12	0.3	957	15
P15	1	0	0	0	18	0	343	16
P16	4	3,590	793	15,257	1	0.3	4,212	7
P17	1	81	81	1,765	11	0.1	1,350	13
P18	4	815	407	8,148	6	0.7	1,742	11
Total	77	15,280	4,623	83,580	-	-	56,938	-
Average	4.3	849	257	4,643	-	0.3	3,163	-

The cumulative reproductive success of the ESO parents normalized by the number of ramets to account for their unbalanced representation in the orchard is presented in Figure 2.5.

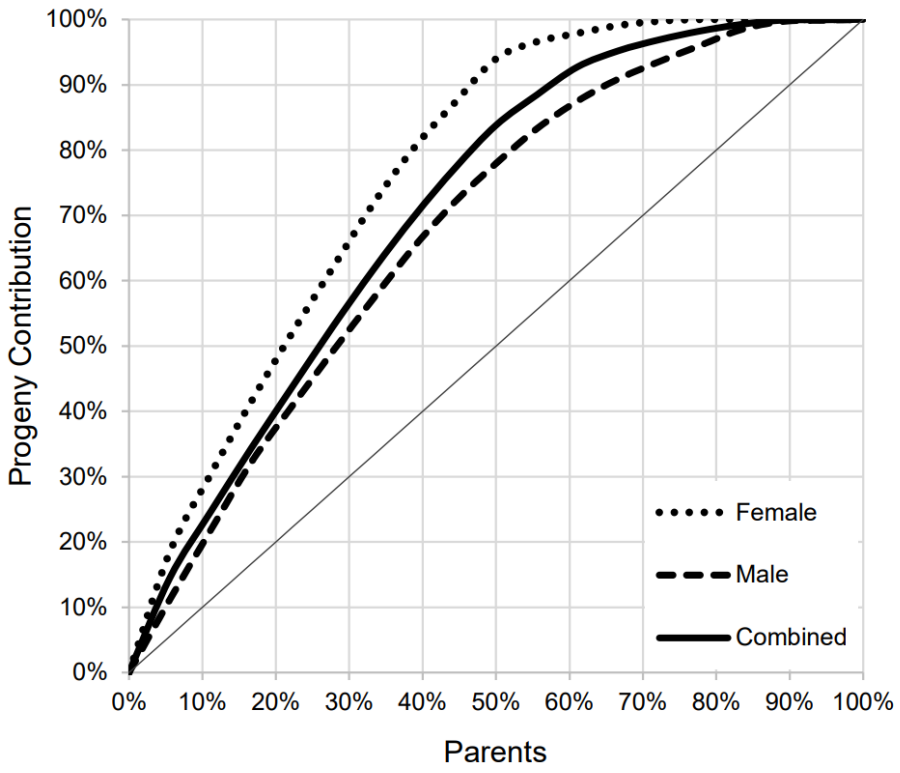


Figure 2.5. Cumulative female, male, and combined reproductive success. The diagonal line represents an equal contribution of the breeding parents.

2.3.3 Pollination Spatial Pattern

Three genotypes were present with a single tree/ramet: P04, P15, and P17. Their male reproductive outputs, i.e., the number of progeny with the genotype as a male parent, were 24, 343, and 1,350, respectively. We have used the parent with the largest male contribution, P17, to model the pollen spatial dispersion pattern. Of the 24 OP families sowed, 21 had P17 as one of its male parents, not accounting for P17 itself. We calculated the proportional contribution to the entire open-pollination family for these 21 mother trees that received P17's pollen. The largest proportional contributions

were detected in the proximity of the pollen source, decaying with distance following a power function with $R^2 = 0.77$, as presented in Figure 2.6.

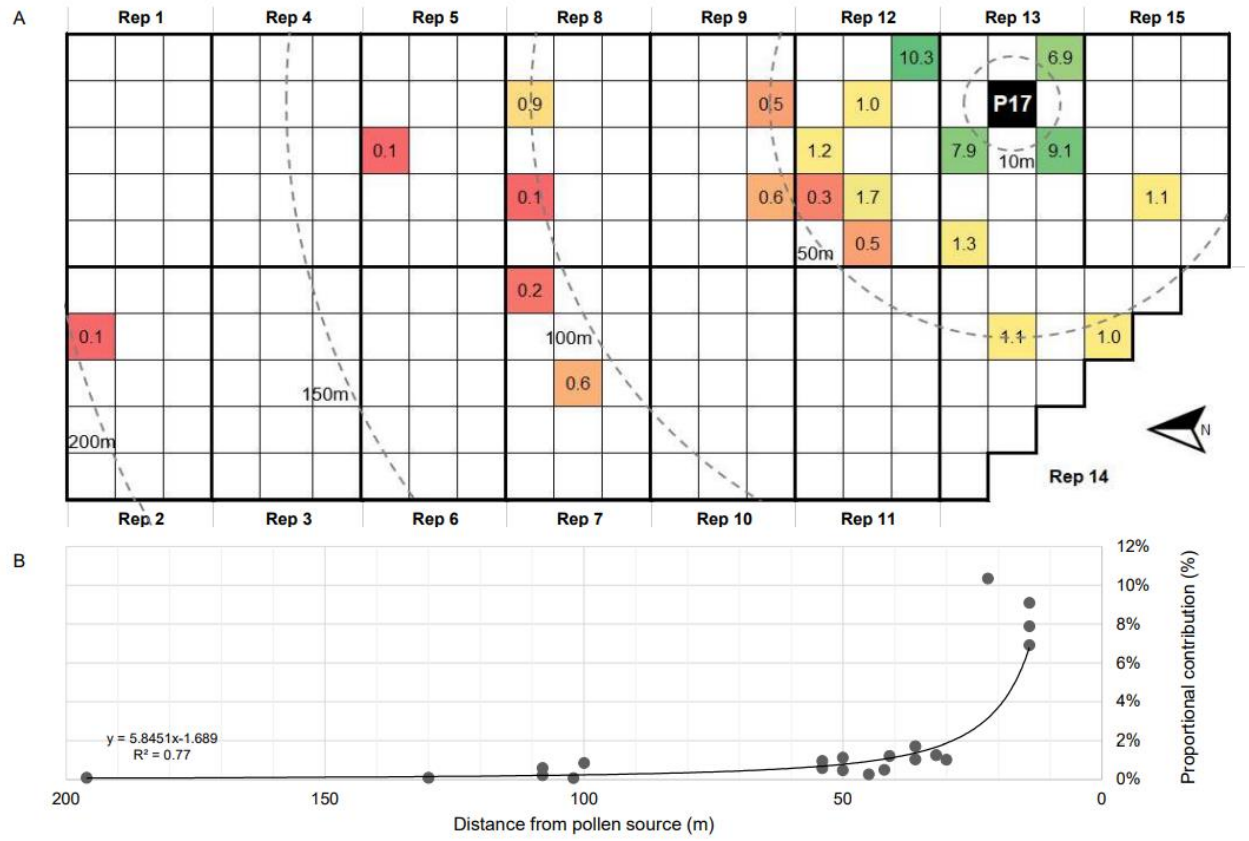


Figure 2.6. P17's spatial pollen distribution across the orchard is revealed by pedigree reconstruction. The values in each cell of the orchard design (A) are the proportional contribution of P17 as a pollen donor relative to the entire open-pollination family size, expressed as a percentage of the total progeny count. The contribution decay follows a power function with $R^2 = 0.77$ (B).

2.4. Discussion

Seed orchards are a key component of forest breeding cycles and are assumed error-free regarding their genetic integrity until proven differently. On the other hand, identification errors

are rather common (Slavov *et al*. 2005; Kaya and Isik 2010; Przybylski *et al*. 2019). Quality control of genetic identity is critical throughout the breeding cycle. Genetic gains can be compromised if genetic identities are lost or confused. "While proper labeling and care during all aspects of the tree improvement program are essential, genetic fingerprinting is also an important tool that should be used to enhance quality control" (White *et al*. 2007). Molecular markers are useful to validate the orchard's genetic integrity, correcting errors such as misidentification of ramets or the presence of "external" genotypes. In our study, supported by SNP markers, we detected three identification errors, i.e., genotypes not intended to be included in the original design, and therefore precisely confirmed the actual genotypic composition of all trees in the orchard. Furthermore, large-scale pedigree reconstruction was made possible by accurately determining parent-offspring pairs and mating system dynamics.

Despite the apparent complexity and availability of a suite of software and methods for parentage analysis, simple Mendelian inheritance is the fundamental mechanism behind it. Thus, it is logical that genotyping errors will impact the accuracy of parentage assignments because they will likely result in apparent incompatibilities between parents and offspring. Compared with genotype-by-sequencing (GbS), SNP-PCR marker systems usually produce more reliable genotypes (Whalen *et al*. 2019) and are the best method considering both reliability and cost per sample. Furthermore, choosing informative SNPs will provide higher power for parentage analysis and maximize cost efficiency. To maximize power, SNP minor allele frequency and genotype call accuracy should be high (Flanagan and Jones 2019). We used a validated SNP panel to successfully perform 3.38 million SNP genotyping events with a call rate of 98.4%, call accuracy of 99.2%, and an average MAF of 0.425. These were excellent parameters that added confidence to assigning parentage accurately.

Paternity analysis was performed assuming a known mother, as we carefully harvested pods from female trees. Also, the analysis is more powerful if the female parent is known (Marshall *et al*. 1998). Indeed, 99.5% of the progeny analyzed matched with confidence the female parent informed *a priori*, indicating the good traceability of the harvesting campaign and the accuracy of the parentage assignment on the maternal side. On the male side, 67.8% of the assignments were from an ESO parent, 12.8% were from the neighbor CBO located 150 meters away, and 19.4% were unknown, i.e., pollen contamination from an unknown male parent.

This level of external pollen contamination was lower than what has been found for other tree species. In *Eucalyptus grandis*, Jones *et al*. (2008) reported 46% assignment of pollen parents out of the population beyond 192 meters from the respective mother trees outcrossed. Grattapaglia *et al*. (2004) found 29% external pollen contamination for hybrid *E. grandis x E. urophylla*. In *Pinus sylvestris* L., Hall *et al*. (2020) found a 33.6% external pollen contamination with the closest Scots pine stands to the seed orchard more than 500 m away.

With parentage resolved, we found a surprisingly rich male contribution to the open-pollination family structure. The average number of male parents per family was 50, representing 86% of the 58 known male parents with at least one progeny in the family. The contribution of each male parent was uneven, and the most successful male parent within an open-pollinated family accounted, on average, for 23% of the total number of individuals. In contrast, the smallest male within OP accounted for a single progeny in all families tested.

We could also demonstrate the spatial pollen dispersion pattern for a single genotype, P17. This parent was a low-tier genotype for reproductive success, with only 81g of seeds harvested, with a progeny count of 1,765 as female (rank 11 out of 18) and 1,350 as male (rank 13 out of 18). Its pollen dispersion showed a moderate-high spatial correlation between the mother tree's distance

to the pollen source and its proportion within the open-pollination family, with a rapid decay with distance increase. This result is consistent with the study of Le *et al*. (2019) dealing with natural hybridization between *A. auriculiformis* and *A. mangium*. They concluded that the frequency of pollen flow decreased exponentially with increasing distance from the pollen source.

The correlation between the number of ramets and male reproductive success was significant (r = 0.72). More ramets scattered around by an appropriate orchard design, such as the Optimum Neighborhood Algorithm design (Chaloupková *et al.* 2016), will likely increase the overall male success. Many ramets of a given genotype pollinating their closest female trees in larger proportions, similar to what was observed for P17, will increase their overall male success when accounting for the entire orchard as a bulk. However, even for female parents receiving P17's pollen not farthest than 20 meters, its average contribution to the open-pollination family was only about 10%, showing that distance alone is not the only factor determining pollination success. We can expect synchronicity driven by dichogamous flowering, pollinator availability, and reproductive success to play a major role, especially in the absence of adaptations for capturing wind-borne pollen, a pattern consistent with most species in the genus *Acacia* (Gibson *et al*. 2011).

For *A. mangium*, Yuskianati and Isoda (2013) indicated that approximately 80% of all crosses were between trees separated by 40 m or less. Nurtjahjaningsih (2016) found that pollinations occurred over distances ranging between 15 and 150 m while studying a seed orchard in Wonogiri, Central Java, Indonesia. In contrast, some *Acacia* species potentially pollinated by birds have recorded longer distances of pollen movement. Maximum pollinator dispersal distances exceeded 1,870 m in *A. woodmaniorum* (Millar *et al*. 2014), and occasional hybridization between native populations of *A. saligna* subsp. *saligna* and subsp. *lindleyi* was detected over distances of

around 1,600 m (Millar *et al*. 2008). We did not find any previous reports of pollination dynamics for *A. crassicarpa* for a direct comparison.

The female and male reproductive success of genotypes was variable. The cumulative maternal reproductive success indicated that 40% of the parents contributed 80% of the seed production, whereas, on the paternal side, 50% contributed 80% of the pollination. The reproductive success distortion was slightly more pronounced on the maternal side. We could verify that for the combined cumulative reproductive success, 20% of the parents contributed 40% of the seed production. We attribute *A. crassicarpa* to a 20/40 distortion pattern. These results differ somewhat from wind-pollinated species, such as pines, generally showing larger distortions with a 20/80 pattern (Park *et al*. 2020; Lai *et al*. 2010).

Our study describes the open-pollination recombination pattern in *A. crassicarpa* as rich and broad, without a single predominant male parent, and with many pollen donors present within the family structure, even at varying proportions. From a breeding perspective, this feature is advantageous when the objective is to generate many different crossing combinations, with recombination events approximating panmixis, allowing for the reconstruction of full-sib families in sufficient quantity to fill a diallel structure. For *A. crassicarpa*, the lack of efficiently controlled pollination methods is an important limitation. Using genomic methods, we could reconstruct 95 full-sib families from an OP seed collection, representing 62% of the total number of families not accounting for reciprocals, with at least 30 seedlings per family, sufficient for full-sib progeny testing. Full-sib mating designs allow the estimation of both additive and non-additive components of genetic variance and specific combining effects, which may prove valuable in future breeding efforts with *A. crassicarpa*.

Finally, the selfing rates estimated in our study were very low, with an average of 0.3%, indicative of an efficient mechanism to prevent large-scale selfing in *A. crassicarpa*. Among selfed individuals pointed out by the parentage analysis, it was impossible to discriminate true-to-type self-pollinations from cross-pollinations with pollen from the same male parent genotype. Nevertheless, selfing was of minor scale in the open-pollination family structure analyzed.

2.5 Conclusion

Parentage analysis using an informative SNP panel successfully characterized a reproductive season of a breeding orchard of *A. crassicarpa*. The species has a generalist entomophilous pollination syndrome with a mating dynamic favoring a rich male composition of open-pollination families. Even though we could show a case of spatial pollen pattern distribution with the rapid decay of pollination with distance to pollen source, in general, most families showed many male parents without a single dominant one. On average, the male that produced the most progeny in a given family had a 23% contribution. If seed orchard managers and breeders target orchard designs that provide optimal mating, open pollination should result in good admixture minimizing preferential mating. Our study showed a low selfing rate of 0.3%, corroborating an outcrossed breeding system for the species. From an applied tree breeding perspective, the parentage reconstruction efficiently generated many full-sib families, allowing progeny testing and overcoming the controlled pollination limitation to produce the crosses.

2.6 References

Bernhardt, P. The floral biology of Australian *Acacia*. In: Advances in legume biology. Proceedings of the Second International Legume Conference, St. Louis, Missouri, 23-27 June 1986; Stirton, CH, Zarucchi JL, Eds.; Missouri Botanical Garden, St Louis, U.S.A, 1989, pp. 263–281.

Butcher P, Harwood C, Quang TH (2004) Studies of mating systems in seed stands suggest possible causes of variable outcrossing rates in natural populations of *Acacia mangium*. Forest Genetics, 11(3-4), 303-309.

Chaloupková K, Stejskal J, El-Kassaby, YA *et al*. (2016) Optimum neighborhood seed orchard design. Tree Genetics & Genomes, 12, 105. https://doi.org/10.1007/s11295-016-1067-y

Flanagan S, Jones AG (2019) The future of parentage analysis: From microsatellites to SNPs and beyond. Molecular Ecology, 28, 544-567. https://doi.org/10.1111/mec.14988

George N, Byrne M, Yan G (2008) Mixed mating with preferential outcrossing in *Acacia saligna* (Labill.) H. Wendl. (*Leguminosae: Mimosoideae*). Silvae Genetica, 57, 139–145. https://doi.org/10.1515/sg-2008-0021

Gibson MR, Richardson DM, Marchante E, Marchante H, Rodger JG, Stone GN *et al*. (2011) Reproductive biology of Australian Acacias: Important mediator of invasiveness? Diversity and Distributions, 17:5, 911-933. https://doi.org/10.1111/j.1472-4642.2011.00808.x

Grattapaglia D, Ribeiro VJ, Rezende GDSP (2004) Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus*. Theoretical and Applied Genetics, 109, 192-199. https://doi.org/10.1007/s00122-004-1617-9

Griffin AR, Vuong TD, Harbard JL, Wong CY, Brooker C, Vaillancourt RE (2010) Improving controlled pollination methodology for breeding *Acacia mangium* Willd. New Forests, 2010, 40:2, 131–142. https://doi.org/10.1007/s11056-010-9188-x

Hall D, Zhao W, Wennström U, Gull BA, Wang X (2020) Parentage and relatedness reconstruction in *Pinus sylvestris* using genotyping-by-sequencing. Heredity, 124, 633-646. https://doi.org/10.1038/s41437-020-0302-3

Harwood CE, Ha HT, Tran HQ, Butcher PA, Williams ER (2004) The effect of inbreeding on early growth of *Acacia mangium* in Vietnam. Silvae Genetica, 53, 65–69. https://doi.org/10.1515/sg-2004-0012

Harwood, CE, Nambiar EKS (2014) Sustainable plantation forestry in South East Asia. ACIAR Technical reports No. 84. Australian Centre for International Agricultural Research: Canberra. 100 pp.

Harwood, CE, Hardiyanto EB, Wong CY (2015) Genetic improvement of tropical Acacias: achievements and challenges. Southern Forests, 77, 11-18. https://doi.org/10.2989/20702620.2014.999302

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. Molecular Ecology, 12, 2511–2523. https://doi.org/10.1046/j.1365-294X.2003.01928.x

Jones ME, Shepherd M, Henry R (2008) Pollen flow in *Eucalyptus grandis* determined by paternity analysis using microsatellite markers. Tree Genetics & Genomics, 4, 37-47. https://doi.org/10.1007/s11295-007-0086-0

Jones, AG, Clayton, MS, Paczolt KA, Ratterman, NL (2010) A practical guide to methods of parentage analysis. Molecular Ecology Resources, 10, 6–30. https://doi.org/10.1111/j.1755-0998.2009.02778.x

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology, 16, 1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Kaya N, Isik K (2010) Genetic identification of clones and the genetic structure of seed crops in a *Pinus brutia* seed orchard. Turkish Journal of Agriculture and Forestry, 34, 127–134. https://doi.org/10.3906/tar-0904-11

Kenrick J, Knox RB (1982) Function of the polyad in reproduction of *Acacia*. Annals of Botany, 50, 721–727. https://www.jstor.org/stable/42758606

Kenrick J, Knox RB (1989) Quantitative analysis of self-incompatibility in trees of seven species of *Acacia*. Journal of Heredity, 80, 240–245. https://doi.org/10.1093/oxfordjournals.jhered.a110842

Kenrick J (2003) Review of pollen-pistil interactions and their relevance to the reproductive biology of *Acacia*. Australian Systematic Botany, 16, 119–130. https://doi.org/10.1071/SB02005

Lai SK, Funda T, Liewlaksaneeyanawin C, Klàpste J, Van Niejenhuis A, Cook C *et al*. (2010) Pollination dynamics in a Douglas-fir seed orchard as revealed by pedigree reconstruction. Annals of Forest Science, 67, 808. https://doi.org/10.1051/forest/2010044

Le S, Harwood CE, Nghiem CQ, Griffin AR, Vaillancourt RE (2019) Patterns of hybrid seed production in adjacent seed orchards of *Acacia auriculiformis* and *A. mangium* in Vietnam. Annals of Forest Science, 76, 46. https://doi.org/10.1007/s13595-019-0823-1

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology, 7, 639-655. https://doi.org/10.1046/j.1365-294x.1998.00374.x

Martins GS, Yuliarto M, Antes R, Sabki, Prasetyo A, Unda F, Mansfield SD *et al*. (2020) Wood and Pulping Properties Variation of *Acacia crassicarpa* A.Cunn. ex Benth. and Sampling Strategies for Accurate Phenotyping. Forests, 11, 1043. https://doi.org/10.3390/f11101043

Mendham DS, White DA (2019) A review of nutrient, water and organic matter dynamics of tropical Acacias on mineral soils for improved management in Southeast Asia. Australian Forestry, 82:sup1, 45-56. https://doi.org/10.1080/00049158.2019.1611991

Midgley SJ, Turnbull JW (2003) Domestication and use of Australian Acacias: case study of five important species. Australian Systematic Botany, 16, 89-102. https://doi.org/10.1071/SB01038

Millar MA, Byrne M, Nuberg I, Sedgley M (2008) High outcrossing and random pollen dispersal in a planted stand of *Acacia saligna* subsp. *saligna* revealed by paternity analysis using microsatellites. Tree Genetics & Genomes, 4, 367–377. https://doi.org/10.1007/s11295-007-0115-z

Millar MA, Byrne M, Nuberg I, Sedgley M (2012) High levels of genetic contamination in remnant populations of *Acacia saligna* from a genetically divergent planted stand. Restoration Ecology, 20:2, 260-267. https://doi.org/10.1111/j.1526-100X.2010.00758.x

Millar MA, Coates DJ, Byrne M (2014) Extensive long-distance pollen dispersal and highly outcrossed mating in historically small and disjunct populations of *Acacia woodmaniorum*

(*Fabaceae*), a rare banded iron formation endemic. Annals of Botany, 114, 961–971. https://doi.org/10.1093/aob/mcu167

Miller JT, Murphy DJ, Brown GK, Richardson DM, González-Orozco CE (2011) The evolution and phylogenetic placement of invasive Australian *Acacia* species. Diversity and Distributions, 17:5, 848–860. https://doi.org/10.1111/j.1472-4642.2011.00780.x

Moran GF, Muona O, Bell JC (1989) Breeding Systems and Genetic Diversity in *Acacia auriculiformis* and *A. crassicarpa*. Biotropica, 21:3, 250-256. https://doi.org/10.2307/2388652

Morgan A, Carthew SM, Sedgley M (2002) Breeding system, reproductive efficiency and weed potential of *Acacia baileyana*. Australian Journal of Botany, 50, 357–364. https://doi.org/10.1071/BT01088

Nambiar EKS, Harwood CE (2014) Productivity of Acacia and eucalypt plantations in Southeast Asia. 1. Bio-physical determinants of production: opportunities and challenges. International Forestry Review, 16:2, 225–248. https://doi.org/10.1505/146554814811724757

Nambiar EKS, Harwood CE, Mendham DS (2018) Paths to sustainable wood supply to the pulp and paper industry in Indonesia after diseases have forced a change of species from Acacia to eucalypts. Australian Forestry, 81:3, 148-161. https://doi.org/10.1080/00049158.2018.1482798

Nirsatmanto A, Sunarti S (2019) Genetics and Breeding of Tropical Acacias for Forest Products: *Acacia mangium*, *A. auriculiformis* and *A. crassicarpa*. In: Advances in Plant Breeding Strategies: Industrial and Food Crops; Al-Khayri J, Jain S, Eds.; Springer, Cham: Springer Nature, Switzerland, 2019; Volume 6, pp. 3-28.

Nurtjahjaningsih I (2016) Parent identification in a multi location trial seed orchard of *Acacia mangium* using microsatellite markers. Indonesian Journal of Forestry Research, 3, 19–26. https://doi.org/10.20886/ijfr.2016.3.1.19-26

Park J, Kang H, Yeom D, Kang K, EL-Kassaby YA, Lee K (2020) Gender, reproductive output covariation and their role on gene diversity of *Pinus koraiensis* seed orchard crops. BMC Plant Biology, 20, 418. https://doi.org/10.1186/s12870-020-02632-9

Przybylski P, Kowalczyk J, Odrzykoski I, Matras J (2019) Identyfing alien genotypes and their consequences for genetic variation in clonal seed orchards of *Pinus sylvestris* L. Dendrobiology, 81, 40-46. http://dx.doi.org/10.12657/denbio.081.005

Richardson DM, Carruthers J, Hui C, Impson FAC, Robertson MP, Rouget M *et al*. (2011) Human-mediated introductions of Australian Acacias – a global experiment in biogeography. Diversity and Distributions, 17:5, 771–787. https://doi.org/10.1111/j.1472-4642.2011.00824.x

Skelton DJ (1987) Distribution and ecology of Papua New Guinea Acacias. In: Australian Acacias in developing countries. Proceedings of an international workshop held at Gympie,

Australia, 4-7 August 1986; Turnbull, J.W., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 16., Canberra, Australia, pp. 38-44.

Slavov GT, Howe GT, Adams WT (2005) Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. Canadian Journal of Forest Research, 35, 1592–1603. https://doi.org/doi:10.1139/x05-082

Stone GN, Raine NE, Prescott M, Willmer PG (2003) Pollination ecology of Acacias (*Fabaceae*, *Mimosoideae*). Australian Systematic Botany, 16, 103–118. https://doi.org/10.1071/SB02024

Turnbull JW, Midgley SJ, Cossalter C (1998) Tropical Acacias planted in Asia: an overview. In: Recent developments in Acacia planting. Proceedings of an international workshop held in Hanoi, Vietnam, 27-30 October 1997; Turnbull, J.W., Crompton, H.R., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 82., Canberra, Australia, pp. 155-160.

Whalen A, Gorjanc G, Hickey JM (2019) Parentage assignment with genotyping-by-sequencing data. Journal of Animal Breeding and Genetics, 136, 102–112. https://doi.org/10.1111/jbg.12370

White TL, Adams WT, Neale DB (2007) Forest Genetics. CABI, Cambridge, MA.

Yuskianti V, Isoda K (2013) Detection of pollen flow in the seedling seed orchard of *Acacia mangium* using DNA marker. Indonesian Journal of Forestry Research, 10, 31–41. https://doi.org/10.20886/ijfr.2013.10.1.31-41

CHAPTER 3

Wood and Pulping Properties Variation of *Acacia crassicarpa* and Sampling Strategies for Accurate Phenotyping

Gustavo Salgado Martins[1,3], Muhammad Yuliarto[1], Rudine Antes[1], Sabki[1], Agung Prasetyo[1], Faride Unda[2], Shawn D. Mansfield[2], Gary Hodge[3], and Juan Jose Acosta[3]

[1]*Asia Pacific Resources International Ltd., Jl. Lintas Timur, Pangkalan Kerinci, Kabupaten Pelalawan, Riau 28300, Indonesia.*

[2]*Department of Wood Science, Faculty of Forestry, 4030 - 2424 Main Mall, University of British Columbia, Vancouver, B.C., Canada.*

[3]*Camcore, Department of Forestry & Environmental Resources, College of Natural Resources, North Carolina State University, Raleigh, NC, USA.*

This chapter consists of a manuscript prepared for publication in the scientific journal *Forests* 2020, 11(10), 1043; doi:10.3390/f11101043, reproduced with permission from MDPI, Basel, Switzerland.

ABSTRACT

Research Highlights: This study provides a comprehensive set of wood and pulping properties of *Acacia crassicarpa* to assess variation and efficient sampling strategies for whole-tree level phenotyping. *Background and Objectives: A. crassicarpa* is an important tree species in Southeast Asia, with limited knowledge about its wood properties. The objective of this study was to characterize important wood properties and pulping performance of improved germplasm of the species. Furthermore, we investigated within-tree patterns of variation and evaluated the efficiency of phenotyping strategies. *Materials and Methods:* Second-generation progeny trials were studied, where forty 50-month-old trees were selected for destructive sampling and assessed for wood density, kraft pulp yield, α-cellulose, carbohydrate composition, and lignin content and composition (S/G ratio). We estimated the phenotypic correlations among traits, determined within-tree longitudinal variation and its importance for whole-tree level phenotyping. *Results:* The mean whole-tree disc basic density was 481 kg m^{-3}, and the screened kraft pulp yield was 53.8%. The reliabilities of each sampling position to predict whole-tree properties varied with different traits. For basic density, pulp yield, and glucose content, the ground-level sampling could reliably predict the whole-tree property. With NIR predictions as an indirect measurement method for disc basic density, we verified reduced reliability values for breast height sampling but sufficiently correlated to allow accurate ranking and efficient selection of genotypes in a breeding program context. *Conclusions:* We demonstrated the quality of *A. crassicarpa* as a wood source for the pulping industry. The wood and pulping traits have high levels of phenotypic variation, and standing tree sampling strategies can be performed for both ranking and high accuracy phenotyping purposes.

3.1 Introduction

Acacia crassicarpa A.Cunn. ex Benth. is a fast-growing tree species largely used as a wood source in Southeast Asia (Nambiar and Harwood 2014), where planted forests are supported by advancements in silviculture (Mendham and White 2019), genetic improvement by recurrent selection strategies, vegetative propagation, and recent, use of molecular tools (Nirsatmanto and Sunarti 2019, Harwood *et al*. 2015, McKinnon *et al*. 2018). The primary use of the species' wood is for pulp and paper production, where large vertically integrated companies in the region have hundreds of thousands of hectares of *A. crassicarpa* forests comprising a major component of their wood supply chain (Nambiar *et al*. 2018, Griffin *et al*. 2018).

Targeting higher pulp production efficiency, one of the main breeding objectives is to improve the wood quality to optimize cellulose yield and chemical consumption during the pulping process. Furthermore, the pulp industry is evolving to expand the portfolio of products obtained from forest biomass beyond bleached kraft pulp, potentially including dissolving pulp, chemicals, fuels, and polymers, in alignment with sustainable wood supply and the biorefinery concept for waste-free processing of wood into value-added products (Kumar and Christopher 2017, Lundberg *et al*. 2014, Marinova *et al*. 2010).

Wood chemistry must be considered when evaluating pulping processes and product quality (Schmidt 2005), and is therefore essential in ensuring that the breeding objectives are aligned with the long-term business strategy. To realize genetic gains in wood quality traits, understanding the extent of genetic control is fundamental to the choice of the breeding strategy (ies) to be implemented (White 1987, Borralho *et al*. 1993). Furthermore, knowing how different wood properties are correlated is important, as this information can then establish selection traits that may be used to explain the majority of the variability of interest for a given product attribute.

Finally, the larger the number of characters involved, the smaller the attainable gain from selection for each trait (Chen *et al*. 2016, Rezende *et al*. 2014, Greaves *et al*. 1997).

There is scarce literature dealing with phenotypic and genetic parameters of wood quality traits of *A. crassicarpa*. The first reports of its pulping and paper-making qualities were published in the early 1990s (Clark *et al.* 1991, Clark *et al*. 1992), where three native trees of unknown age showed an average basic density of 638 kg m^{-3} and screened kraft pulp yield of 47.2%. Following on, Laurila 1995 ranked *A. crassicarpa* along with *A. mangium* and *G. arborea* as the most suitable species for pulp and paper in a study comparing eight species in a reforestation project in South Kalimantan (Indonesian portion of Borneo Island) for wood density, strength, fiber properties, lignin content, and extractives content. The basic density, compared at the same age, was higher than *A. mangium* and similar to *A. auriculiformis*. Provenance variation was evaluated by Shukor *et al*. 1998 on four-year-old trees from six provenances that originated from Australia, Papua New Guinea and Irian Jaya (Papua, the Indonesian portion of New Guinea Island). The authors showed significant differences among provenances in shrinkage, compression, and shear parallel to the grain, but none for specific gravity, or the flexural properties MOR (modulus of resistance) and MOE (modulus of elasticity). Yao *et al*. 2012 evaluated four *A. crassicarpa* trees at varying ages, obtaining an average klason lignin content of 21.5%, surprisingly, the highest value among the five *Acacia* species tested, suggesting a low lignin content for the genus.

For other relevant species of the *Acacia* genus, a collection of studies dealing with different wood properties supports an ample understanding of the phenotypic variation found in the genus, as shown for *A. mangium* (Griffin *et al*. 2014, Chowdhury *et al*. 2005, Nugroho *et al.* 2012, Karlinasari *et al.* 2014, Moya *et al*. 2010), *A. auriculiformis* (Chowdhury *et al*. 2009, Tonouéwa *at al*. 2020) and their interspecific hybrid (Dinh Kha *et al*. 2012, Rokeya *et al*. 2010, Rafeadah

and Rahim 2007, Jusoh *et al.* 2014, Barry *et al*. 2005, Yahya *et al*. 2010, Bakri *et al*. 2018). Moreover, there is abundant literature dealing with wood quality traits for other important tree species commercially planted worldwide, contextualizing variation in wood quality traits in a wider range of tree species, as shown for eucalypts (Bouvet *et al*. 2020, Nickolas *et al*. 2020, Stackpole *et al*. 2011, Del-Río *et al*. 2005, Ohra-aho *et al*. 2013, Lima *et al*. 2019, Raymond 2002), pines (Atwood *et al.* 2002, Sykes *et al*. 2003, Neale *et al.* 2002, Li *et al*. 2020, Fundová 2020) and poplars (Schimleck *et al*. 2005, Jin *et al*. 2019, Niemczyk *et al*. 2019). Generally, wood quality traits show intermediate to strong genetic control, with greater genetic stability across environments than growth traits, with the tree's phenotype providing a reliable indicator of its genetic merit, making wood quality traits amenable to genetic advancement through selection (Cornelius 1994, Zobel and Talbert 1984).

To accurately assess the wood properties of a tree, it is important to consider the inherent dimensional changes in wood properties that are unequal along the main bole of the tree, requiring destructive multi-spatial sampling for the accurate determination of whole-tree level phenotype (Zobel and van Buijtenen 1986, Barnett and Jeronimidis 2003, Plomion *et al*. 2001, Von Arx *et al*. 2016, Katz *et al*. 2008, Jones *et al*. 2008, Schimleck *et al*. 2006, Ohshima *et al*. 2005, Igartúa *et al*. 2003, Ona *et al*. 2001, Raymond and Muneri 2001, Raymond *et al*. 2001, Kube and Raymond 2001). On the other hand, the phenotyping procedure for important pulping traits is resource demanding. Thus, wood quality traits are expensive and time-consuming to measure, requiring laboratory facilities and technical expertise to process wood and assess the properties correctly. To overcome the challenge of characterizing a tree with a non-destructive, cost-effective, and faster procedure, wood technologists have developed indirect measurement methods (Gao *et al*. 2017, Schimleck *et al*. 2019) that inherently introduce a trade-off between accuracy and ease of

measurement that depends on the trait and the objective of the characterization. Ultimately, the goal is to capture enough information to adequately quantify the mean and variability in a particular property, accounting for sampling and subsequent analysis costs.

In the context of a tree breeding program, breeders have been able to efficiently characterize families and clones for selection balancing the loss in phenotyping accuracy, due to indirect measurements and single position sampling, with replication. Usually, families and clones are ranked for a particular trait with an average value across multiple trees/ramets measured by a given indirect measurement method. Notably, near-infrared spectroscopy (NIRS) models have been developed and successfully used to predict wood properties (Hodge *et al*. 2018, Meder *et al*. 2010, Schimleck *et al*. 2005, Schimleck *et al*. 2004). Despite the significant investments required for calibration of NIR models, this technique is well-suited for ranking purposes and allows for the inclusion of wood quality traits in the selection criterion and provides the ability to estimate genetic parameters at the population level (Schimleck 2008, Downes *et al*. 1997).

In this study we used 240 50-month-old wood samples, representing six stem positions taken from 40 *A. crassicarpa* trees, to evaluate wood density, screened kraft pulp yield, α-cellulose, carbohydrate composition, and lignin content and composition (S/G ratio) to (1) assess variation in important wood and pulping properties of *A. crassicarpa*; (2) estimate phenotypic correlations among different wood quality and pulping traits, and (3) understand within-tree longitudinal patterns of variation and its importance for whole-tree level phenotyping. With this comprehensive set of wood and pulping property estimates, we aim to increase the amount of information available for this important tree species, and better address efficient sampling procedures for different phenotyping objectives.

3.2 Materials and Methods

3.2.1 Field Trials

A. crassicarpa progeny trials established by the breeding program of Asia Pacific Resources International Limited - APRIL (www.aprilasia.com) were employed in this study. Two replicates of the same group of families in a common breeding zone were established in November 2012 in hemic peat soil with an average bulk density of 0.22 gram cm^{-3} at Pelalawan Regency, Riau Province, Indonesia. Each test consisted of 25 treatments in an eight replication randomized complete block design with ten trees per plot (two rows of five trees). The specimens consisted of open-pollinated families derived from a clonal breeding orchard with second-generation selections from Papua New Guinea provenance.

3.2.2 Wood Properties Measurements

3.2.2.1 Wood Sampling and Basic Density Determination

Utilizing the company's NIR model routinely used in the breeding program, all families were ranked at age three years for basic density, and four families covering the observed range in the trials were selected. Trees were field inspected for tree stem defects, diseases, and straightness, and ten trees of each of the four families were then chosen for destructive wood sampling. The trees were felled, and commercial heights (HTcom) were determined from the base to the point representing a 4.5 cm diameter at the top. The bole positions corresponding to breast height (1.3 m), and 0%, 25%, 50%, 75%, and 100% of the commercial height were then marked, the over-bark diameter was recorded, and cores, discs, and logs were extracted (Figure 3.1) and transported to APRIL's Wood Tech Laboratory.

Two discs (2.5 cm of thickness) were collected at all positions. From the first disc, wedges were ground to produce woodmeal following TAPPI T 257 as source material for NIR spectral

collection and associated wet chemical analysis. The second disc was used to determine the disc basic density (DBD), measured in kg m^{-3}, following TAPPI T 258. At positions representing 0%, 25%, 50%, and 75%, logs of 1.2 m length were produced, and from their central region, the two discs were extracted. The remaining billets from each log were chipped (FARMI CH 260 OEM), and screened following SCAN CM:40-01 standard using a wood chip classifier (TMI 71-01 Chip Class) to provide chips for basic density (CBD) and screened kraft pulp yield (KPY) measurements. The CBD was measured following TAPPI 258 procedure. In addition, samples representing whole-tree composition were prepared by combining volumes of chips from each position proportionally to its area, and composite chips basic density (CBDc) and composite screened kraft pulp yield (KPYc) were determined.

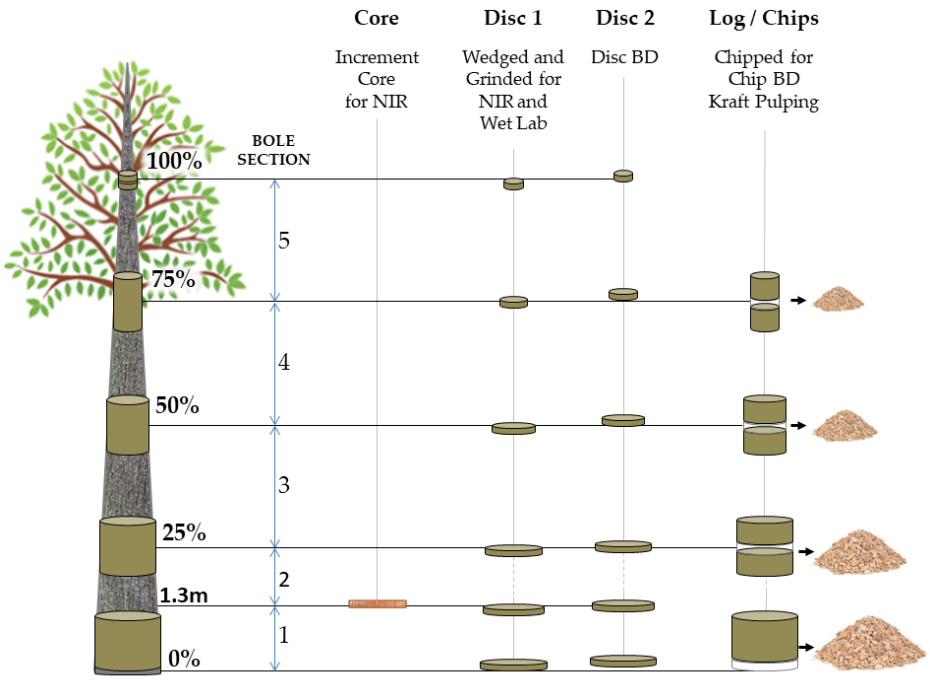


Figure 3.1. Destructive sampling for direct measurements of wood properties and correspondent laboratory analyses performed at each position along the tree bole. The figure is a schematic and not to scale and is intended for illustration only.

3.2.2.2 Wood Carbohydrates and Lignin

Five grams of wood meal were supplied to the University of British Columbia for wood chemical analysis, according to Hart *et al*. 2013. The sample representing the 100% position was not included in the sample set for wood chemical determination. The wood samples were then ground using a Wiley mill to pass through a 40 mesh screen and Soxhlet extracted with hot acetone for 24 hours, and then oven-dried at 105 ºC for 24 hours. Three mL of 72% (w/w) H_2SO_4 was pipetted into a test tube containing approximately 200mg of dried material and was mixed for 30 seconds every 10 minutes. After two hours, the contents of the test tube were transferred to a serum bottle using 112 mL nanopure water. The serum bottles were then sealed and autoclaved at 121 ºC for 60 minutes. After autoclaving, the contents of the bottles were allowed to cool, then vacuum filtered through a pre-weighed medium coarseness crucible (Pyrex, USA) and 15 mL filtrate collected for further analysis. The retentate was rinsed with 60 mL of deionized water to remove any residual sugars and acid. The crucibles containing the retentate were oven-dried at 105 ºC for 24 hours, and then re-weighed to obtain the insoluble lignin content (INS) of the wood gravimetrically. A sample of the filtrate was analyzed for acid-soluble lignin (SOL) at 205 nm. The total lignin (LIG) (soluble and insoluble) is expressed as a proportion of the initial extractive free wood.

Approximately 0.9 mg of the solubilized filtrate and 0.1 mg of fucose (5mg/mL) internal standard was mixed and filtered through a 0.45 µm nylon filter into a glass vial. The total carbohydrate content (arabinose, rhamnose, galactose, glucose, mannose, and xylose) was determined using an anion exchange high-performance liquid chromatograph (Dx-600; Dionex, Sunnyvale, CA, USA) equipped with an ion exchange PA1 (Dionex) column, a pulsed amperometric detector with a gold electrode, and a SpectraAS3500 auto injector (Spectra-Physics, USA). The concentrations of

arabinose (ARA), galactose (GAL), glucose (GLU), mannose (MAN), rhamnose (RHA), and xylose (XYL) were calculated in proportion to the initial extractive free wood and combined as total carbohydrate concentration.

3.2.2.3 Syringyl-Guaiacyl Ratio (S/G)

A 10 mg sample of oven-dried extract-free wood was used to determine the lignin monomer composition. One mL of a reaction mixture (8.75 mL dioxane, 250 µL BF_3, and 1 mL ethanol) was added to a 6 mL reaction vial containing the dried material and purged with N_2 gas before the lid was tightly sealed. Vials were placed in a heating block at 100 ºC for 4 hours with periodic (hourly) agitation. The vials were transferred to a -20 ºC fridge for 5 minutes to halt the reaction. Then, 200 µL of internal standard (5 mg tetracosane /1 mL methylene chloride) and 300 µL 0.4 M $NaHCO_3$ were added to the vial to bring the pH between 3 and 4. Next, 2 mL of nanopure water and 1 mL methylene chloride were added to the vial, which was then recapped, vortexed, and allowed to separate into two phases. One mL of the lower phase was drawn by pipette, filtered through anhydrous Na_2SO_4, and finally transferred directly into a 2 mL polypropylene safe-lock microfuge tube. The sample was evaporated to dryness in a Speedvac set to 45 ºC and then resuspended in 700 µL of methylene chloride. Twenty uL of resuspended sample was derivatized by combining it with 20 µL of pyridine and 100 µL of N,O-bis(trimethylsilyl) acetamide in a glass insert within an amber-glass vial. The vial was sealed and inverted to mix, and allowed to incubate for at least 2 hours at 25 ºC prior to analysis. Finally, 1 µL of solution was analyzed by a gas chromatograph (HP 5890 Series II, Agilent Tech., Ontario, Canada) on an HP 6890 series II column equipped with an auto injector and flame ionization detector (FID) (Agilent Tech., Ontario, Canada) as per Robinson and Mansfield 2009.

3.2.2.4 Alpha Cellulose

Alpha cellulose (αCEL) was determined according to Porth *et al*. 2013, with minor modifications. In brief, the acetone extracted wood was allowed to dry overnight at 50 °C. 3.5 mL of solution A (60 mL glacial acetic acid + 1.3 g NaOH L^{-1}) and 1.5 mL of 20% sodium chlorite solution (20 g $NaClO_2$ in 80 mL distilled water) were added to exactly recorded amounts (~150 mg) of extract-free wood meal to initiate the chlorite delignification. The reaction tube was tightly sealed and then gently shaken at 60 ºC for 14 h. The reaction was quenched by placing the tubes in an ice bath, and the reaction solution was then thoroughly removed by pipetting while not disturbing the settled reacted wood meal. This procedure was repeated using fresh aliquots of each reactant. Finally, the reacted wood meal was transferred to a pre-weighed coarse sintered crucible, and washed twice with 50 mL of 1 % glacial acetic acid (under vacuum), followed by two washes with 10 mL acetone under vacuum, then dried at 50 ºC overnight to obtained holocellulose yield. To obtain the alpha cellulose content of the woody material, alkaline extractions using two different sodium hydroxide extractions were performed sequentially to remove the hemicelluloses. Exact weights of ~100 mg of holocellulose were transferred to a small beaker and left at room temperature for 30 min to allow moisture equilibration. To this 8 mL of 17.5% NaOH (from sodium hydroxide 50 % w/w) was added and the material was left to react for 30 min at 40 ºC. Then, 8 mL of distilled water were added and the material was stirred for one min, and left to react for 29 min. The reaction solution was carefully removed, and the process was repeated with fresh reactants. After the second reaction, all retentate was filtered through a pre-weighed coarse sintered crucible by washing with distilled water (3 x 50 mL). Subsequently, the reaction was neutralized by soaking in 1.0 M acetic acid for 5 min. After additional washing with distilled water (3 x 50 mL), the material was dried at 50 ºC overnight to obtain the alpha cellulose content gravimetrically.

3.2.2.5 Kraft Pulping

The screened wood chip samples of positions 0%, 25%, 50%, 75%, plus the whole-tree composite sample were measured for moisture content following TAPPI T 550, and screened kraft pulp yield (KPY) calculated using gravimetry. The cooking equipment used was a CRS autoclave with six 3-L rotating vessels, with every vessel carrying 350 bone-dry grams of chips and the liquor-to-wood ratio at 1:4. The batch digestors were controlled electronically by electric heating, and the cooks were carried out at 165 °C with a constant variable alkali charge of 15.5% to 19.5% EA as NaOH and fix sulfidity of 35%. Both parameters were chosen based on the kappa target 18 ± 1. The kappa number was determined by TAPPI T 236. Every result presented is an average of at least five cooks per sample at the same kappa number interval.

3.2.2.6 NIR Modelling

A NIR model was trained for basic density utilizing the 240 disc basic density (DBD) direct measurements, and the corresponding NIR spectra generated by 32 scans averaged to produce a single reflectance spectrum for each sample. The equipment used was a "FOSS NIR XDS Rapid Content Analyser" measuring reflectance at 660 NIR wavelengths covering the range of 1100 to 2500 nm.

A data analysis pipeline, written in R, was used for model development in two separate phases: 1) transformations and outlier detection, and 2) model training, cross-validation, and model selection (Hodge *et al*. 2018, Acosta *et al*. 2020). The first phase of the program applies mathematical transformations to NIR spectra to remove the scattering of diffuse reflections associated with sample particle size and to improve subsequent regression analyses. Scatter-correction methods and spectral derivatives were applied to the spectral data. Scatter-correction methods included: multiplicative scatter correction (MSC), standard normal variate (SNV) and

detrend (DT); spectral derivative methods included second-order polynomial, the second derivative of Savitzky-Golay smoothing with two different window sizes of 5 and 7 points (SG5 and SG7), and combination of transformation by pairs (SNV + SG, MSC + SG and DT + SG). For all observations on each spectral database, local outlier factors (LOF) were calculated and used to identify outliers based on density and distance (Breunig *et al*. 2000). Individuals with LOF values greater than two were excluded from the analysis.

The second phase uses the outlier free and transformed databases to develop NIR prediction models between spectral data and DBD direct measurements. Partial least squares regression (PLS) was implemented in R using the R-package "pls" (Mevik *et al*. 2018), and model performance was evaluated using leave-one-out (LOO) cross-validation. Desirable models are those that: maximize the cross-validation coefficient of determination (R^2_{cv}); minimize the standard error of cross-validation ($RMSEP_{cv}$); have a small number of latent variables (projection factors). Based on the criterion mentioned above, the best model was selected to predict the DBD at each position sampled.

3.2.3 Within-tree Level Analysis

At the bole positions level, exploratory analysis was performed with the following descriptive statistics: the number of observations (N), mean, standard deviation (SD), coefficient of variation (CV), and box plots. Within-tree longitudinal variation patterns were estimated with simple linear regression models, fitted with the positions as regressors of the wood property along the bole. The predictor is the position, numerically expressed as the percentage of commercial height, and the response is the trait considered. Thus, the slope shows how the trait is varying for every 1% of the commercial height, from base to top. To transform it into an objective statistic, we multiplied the slope coefficient by 100 (Slope100) to express the total variation in the

regression line, corresponding to the difference of fitted values at 0% and 100% of the commercial height. For all traits, the "Slope100" descriptive statistics were calculated with their statistical significance obtained via t-tests for the mean, minimum, and maximum. Also, Shapiro-Wilk normality tests were performed to determine if the slopes are normally distributed.

3.2.4 Whole-tree Level Analysis

For each bole section, determined for the log in between the positions measured, we calculated the frustum volume and its mean wood property averaging the lower and upper measures weighted by their areas. The whole-tree phenotypes were then estimated as the average value of all bole sections weighted by their volumes. Pearson's correlation (r) of all pairwise combinations of variables at the whole-tree level were calculated, and t-tests with n-2 degrees of freedom for their significance.

The precision of reduced sets of sampling positions to predict the whole-tree phenotype – referred to as reliability in the remainder of this manuscript – was investigated by comparing simple linear regression models with positions taken singly, in pairwise and three-way combinations. For pairwise and three-way combinations, the predictor was the mean value of the corresponding positions. The reliability was assessed using both the model R^2 and the non-parametric Spearman's rank correlation coefficient (ρ). The effect of indirect measurements was explored considering the NIR DBD predictions at each position. Similarly, as described above for the direct measurements of DBD, the reliability and ρ between combinations of positions were calculated and compared with the whole-tree value.

All statistical analyses were done using the R software environment version 3.6.2 (R core Team, 2020).

3.3 Results

3.3.1 Within-tree Level Wood Properties

The average commercial height was 17.88 m, and the volumetric proportions of each bole section are presented in Table 3.1 and Figure 3.2. The descriptive statistics at the position level are shown in Appendix A4. The box plots express the longitudinal variation at the population level. For each trait, the variation was fairly constant across positions, with similar coefficients of variation.

Table 3.1. Average tree bole sections volumes and their proportion of the total volume.

Position	Section	Diameter (cm)		Length (m)	Volume (m^3)	% Total Volume	Cumulative Volume
		Lower	Upper				
0% - 1.3 m	1	20.9	18.2	1.30	0.0396	15.5%	15.5%
1.3 m - 25%	2	18.2	15.9	3.17	0.0738	28.9%	44.4%
25% - 50%	3	15.9	13.1	4.47	0.0749	29.4%	73.8%
50% - 75%	4	13.1	9.9	4.47	0.0474	18.6%	92.4%
75% - 100%	5	9.9	4.5	4.47	0.0194	7.6%	100%

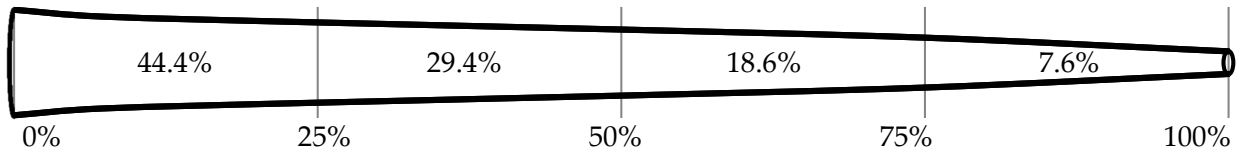


Figure 3.2. Graphical representation of volumetric proportion of the average tree bole sections.

To assess within-tree longitudinal patterns of variation, the positions must be considered as a group of measurements taken on the same specimen. The slope of a simple linear regression model, fitted with the positions as regressors of the wood property along the bole, reflects the tree's linear longitudinal pattern of variation. The Slope100 descriptive statistics for all traits are presented in Table 3.2. For DBD, the average tree showed a decreasing density with -68.2 kg m^{-3} difference from bottom to top. All trees showed a decreasing trend for DBD, although at varying

rates. The trees with the minimum and maximum slope100 values of DBD are presented in Figure 3.3(a). Similar patterns were found for CBD, LIG, and INS. The screened kraft pulp yield (KPY) had slope100 with the opposite pattern, with a mean of 3.25%, varying from -0.22% to 8.06%, showing that all trees have an increasing trend with varying positive slope100 values. The traits GLU, XYL, MAN, and SOL have positive means, but with slope100 ranging from negative to positive values, indicating signal reversions when comparing different trees. The trees with the minimum and maximum slope100 values of GLU are presented in Figure 3.3(b). For αCEL, ARA, GAL, RHA, and S/G, the signal reversion was also present, but with the average tree showing no longitudinal variation with slope100 mean equals zero. For all traits, the Slope100 was normally distributed with Shapiro-Wilk's normality test p-values > 0.05.

Table 3.2. Slope100 descriptive statistics for diameter (DIA), disc basic density (DBD), chips basic density (CBD), screened kraft pulp yield (KPY), alpha cellulose (αCEL), glucose (GLU), arabinose (ARA), galactose (GAL), rhamnose (RHA), xylose (XYL), mannose (MAN), total lignin (LIG), insoluble lignin (INS), acid-soluble lignin (SOL) and syringyl-guaiacyl ratio (S/G).

Variable	N	Mean	Min	Max	Range	SD	Skew	Kurtosis	SW_p [1]
DIA	40	-14.9*	-20.1*	-9.7*	10.5	2.5	-0.18	-0.85	0.69
DBD	40	-68.2*	-133.7*	-10.0 ^{ns}	123.8	31.5	-0.03	-0.98	0.68
CBD	40	-59.1*	-149.4*	15.7*	165.1	37.1	-0.55	-0.04	0.30
KPY	40	3.25*	-0.22 ^{ns}	8.06*	8.28	2.00	0.49	-0.70	0.13
αCEL	40	0.03 ^{ns}	-10.83*	9.47*	20.30	3.89	-0.32	0.48	0.47
GLU	40	1.34*	-8.29*	9.33*	17.62	3.47	-0.37	0.20	0.77
ARA	40	-0.02 ^{ns}	-0.20*	0.17*	0.37	0.07	-0.23	0.29	0.53
GAL	40	0.01 ^{ns}	-0.47*	0.30*	0.77	0.17	-0.73	0.26	0.09
RHA	40	-0.05*	-0.21*	0.19*	0.40	0.08	0.40	0.31	0.75
XYL	40	1.08*	-1.40*	4.19*	5.58	1.40	0.31	-0.37	0.42
MAN	40	0.55*	-0.64*	1.63*	2.27	0.53	0.19	-0.45	0.39
LIG	40	-3.82*	-7.26*	0.70*	7.96	1.83	0.08	-0.52	0.91
INS	40	-4.20*	-8.00*	-0.52 ^{ns}	7.48	1.84	-0.13	-0.88	0.44
SOL	40	0.36*	-0.29*	1.25*	1.55	0.37	0.44	-0.15	0.41
S/G	40	0.02 ^{ns}	-0.27*	0.18*	0.45	0.10	-0.63	0.26	0.19

[1] Shapiro-Wilk normality test p-value, with null hypothesis H_0 = data are normally distributed.

* significant at 0.05; ^{ns} not-significant.

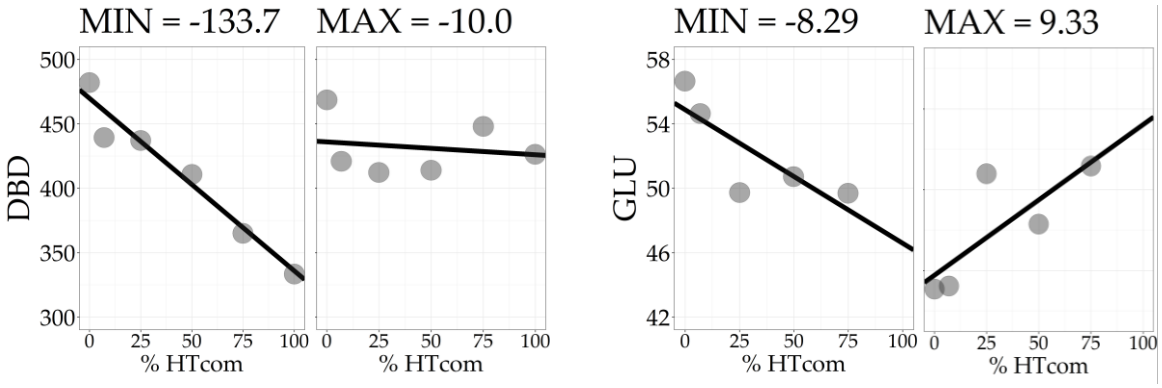


Figure 3.3. Longitudinal pattern of variation of the trees with the minimum and maximum Slope100 for disc basic density (a) and glucose content (b).

3.3.2 Whole-tree Level Wood Properties

The whole-tree level statistics for all traits are presented in Table 3.3. For the growth traits, the mean total height was 21.3 m, and the mean diameter at breast height was 18.2 cm. The mean total tree volume was 0.257 m^3. The basic density mean values were 481.7 kg m^{-3} for disc basic density, 467.3 kg m^{-3} for chips basic density, and 474.8 kg m^{-3} for composite chips basic density. As expected, the range for basic density was wide, of the order of 157.5 kg m^{-3}. The mean screened kraft pulp yield was 53.8%, the same value was observed for the composite KPY, both with a 7% range. The mean alpha cellulose content was 44.4%, with an 8.4% range. The two major carbohydrates were glucose, with mean 50.4% and 5.6% range, and xylose, with mean 13.8% and 2.4% range. The contents of the minor carbohydrates mannose, galactose, arabinose, and rhamnose were 1.31%, 0.62%, 0.26%, and 0.23%, respectively. Mean total lignin was 29.35%, with a range of 4.4%, whereas insoluble lignin was 27%, and acid-soluble lignin was 2.35%. The mean lignin monomers syringil-guaiacyl ratio was 1.67, with a range of 0.3.

Table 3.3. Descriptive statistics of commercial height (HTcom), diameter at breast height (DBH), tree volume (VOL), and whole-tree level disc basic density (DBD), chips basic density (CBD), composite chips basic density (CBDc), screened kraft pulp yield (KPY), composite screened kraft pulp yield (KPYc), alpha cellulose (αCEL), glucose (GLU), arabinose (ARA), galactose (GAL), rhamnose (RHA), xylose (XYL), mannose (MAN), total lignin (LIG), insoluble lignin (INS), acid-soluble lignin (SOL) and syringyl-guaiacyl ratio (S/G).

Variable	Units	N	Mean	SD	CV	Median	Min	Max	Range	SE
HTcom	m	40	21.3	0.9	0.04	21.2	17.4	22.9	5.5	0.14
DBH	cm	40	18.2	2.2	0.12	18.0	13.5	22.5	9.0	0.34
VOL	m^3	40	0.257	0.055	0.21	0.250	0.141	0.390	0.249	0.01
DBD	kg m^{-3}	40	481.7	34.4	0.07	485.6	400.9	558.4	157.5	5.44
CBD	kg m^{-3}	40	467.3	31.3	0.07	470.5	392.7	534.3	141.6	4.95
CBDc	kg m^{-3}	37	474.8	31.6	0.07	479.4	398.6	528.6	130.0	5.19
KPY	%	40	53.8	1.6	0.03	54.0	49.7	56.8	7.1	0.25
KPYc	%	39	53.8	1.5	0.03	53.8	49.4	56.4	7.0	0.25
αCEL	%	40	44.4	1.7	0.04	44.3	40.1	48.5	8.4	0.27
GLU	%	40	50.4	1.6	0.03	50.5	47.5	53.1	5.6	0.25
ARA	%	40	0.26	0.06	0.23	0.30	0.10	0.30	0.20	0.01
GAL	%	40	0.62	0.11	0.18	0.60	0.40	0.90	0.50	0.02
RHA	%	40	0.23	0.05	0.24	0.20	0.10	0.30	0.20	0.01
XYL	%	40	13.8	0.5	0.03	13.8	12.6	15.0	2.4	0.08
MAN	%	40	1.31	0.29	0.22	1.30	0.80	2.20	1.40	0.05
LIG	%	40	29.4	1.0	0.03	29.4	27.4	31.8	4.4	0.16
INS	%	40	27.0	1.0	0.04	27.1	25.3	29.4	4.1	0.16
SOL	%	40	2.35	0.18	0.08	2.40	1.90	2.80	0.90	0.03
S/G	ratio	40	1.67	0.08	0.05	1.70	1.50	1.80	0.30	0.01

The Pearson's correlation matrix is presented in Appendix A5, with elements below diagonal corresponding to the correlations at the whole-tree level. For wood density, very high positive relationships (> 0.9) were apparent between disc basic density, chips basic density, and the composite CBD, indicating that regardless of the measurement method, similar estimates were obtained at the whole-tree level. Similarly, the KPY and the KPYc have a 0.93 correlation, showing strong evidence that the chips composite may be a resource-wise sampling strategy,

especially for expensive phenotyping traits, such as pulp yield. Wood basic density was positively correlated with pulp yield, alpha cellulose, arabinose, galactose, and rhamnose; and negatively correlated with mannose and soluble lignin. Alpha cellulose content was strongly associated with glucose (0.70), and both traits negatively correlated with lignin content (-0.48 and -0.60, respectively). Regarding correlations among carbohydrates, arabinose, galactose, and rhamnose were positively correlated among themselves. Galactose was related to density, to all other sugars, and lignin. Xylose was not associated with any other trait than galactose.

The correlation between total lignin and insoluble lignin was 97%, demonstrating that the insoluble lignin is the major component driving the complete lignin response. Total lignin was negatively associated with glucose, alpha cellulose, and pulp yield, showing consistency among the data, with the antagonistic relationship expected between lignin and pulp yield-related traits. The lignin monomer ratio (syringil-guaiacyl) was significantly correlated only with the carbohydrate mannose.

3.3.2.1 Whole-tree Properties Prediction

The results of the bole positions reductionist analyses for whole-tree property prediction are presented in Table 3.4 for the selected traits KPY, DBD, GLU, XYL, and INS.

In general, when comparing single positions, samples at breast height and 25% were the most reliable to predict the whole-tree value. However, pairwise and three-way combinations of positions increased the correlations. For KPY, the position 25% alone showed reliability of 0.94, and rank correlations of 0.97 to predict whole-tree values, with marginal gains when adding more positions. For DBD, the 1.3 m and 25% alone showed $R^2 \approx 0.84$, and when their mean value was used, it increased to 0.95. For the carbohydrates and insoluble lignin content, single sampling positions had less power to predict whole-tree properties. For glucose, 1.3 m was the best single

position with R^2 = 0.69. When the mean value of 1.3 m and 25% was used, it increased to 0.87, with ρ = 0.93. Adding the 50% as a 3rd position practically yielded a perfect correlation with the whole-tree value. For xylose, an important component of hemicellulose, a single sampling at 25% commercial height could moderately predict whole-tree with $R^{\wedge 2}$ = 0.67. The mean value of 1.3 m and 25% was 0.77. Insoluble lignin showed a similar trend, and the 25% position alone had R^2 = 0.76 and ρ = 0.84. For higher precision, the mean value of 1.3 m and 25% could raise the R^2 to 0.89.

Table 3.4. Reliabilities of simple linear regression models with different sets of positions to predict whole-tree level properties for screened kraft pulp yield (KPY), disc basic density (DBD), glucose (GLU), xylose (XYL) and insoluble lignin (INS).

Pos.	Linear model	KPY		DBD		GLU		XYL		INS	
		R^2	ρ	R^2	ρ	R^2	ρ	R^2	ρ	R^2	ρ
All	*y* ~ 0+1.3+25+50+75+100	1	1	1	1	1	1	1	1	1	1
Alone	*y* ~ 0	0.79	0.88	0.67	0.82	0.29	0.56	0.38	0.56	0.56	0.75
	***y* ~ 1.3**			0.84	0.91	0.69	0.86	0.30	0.55	0.62	0.71
	y ~ 25	0.94	0.97	0.85	0.91	0.65	0.80	0.67	0.82	0.76	0.84
	y ~ 50	0.80	0.88	0.79	0.85	0.59	0.75	0.36	0.62	0.54	0.73
	y ~ 75	0.61	0.77	0.76	0.86	0.33	0.53	0.15	0.32	0.40	0.66
	y ~ 100			0.30	0.39						
Pairwise	***y* ~ mean(0 ; 1.3)**			0.85	0.92	0.60	0.85	0.48	0.67	0.71	0.81
	y ~ mean(0 ; 25)	0.96	0.97	0.92	0.95	0.68	0.82	0.71	0.81	0.81	0.88
	y ~ mean(0 ; 50)	0.96	0.98	0.85	0.89	0.65	0.81	0.63	0.74	0.80	0.89
	***y* ~ mean(1.3 ; 25)**			0.95	0.97	0.87	0.93	0.77	0.86	0.89	0.93
	y ~ mean(1.3 ; 50)			0.94	0.96	0.90	0.94	0.69	0.84	0.86	0.91
	y ~ mean(25 ; 50)	0.94	0.96	0.94	0.96	0.83	0.88	0.75	0.84	0.83	0.88
Three-way	***y* ~ mean(0 ; 1.3 ; 25)**			0.95	0.97	0.80	0.93	0.77	0.87	0.87	0.92
	y ~ mean(0 ; 1.3 ; 50)			0.93	0.96	0.81	0.92	0.74	0.85	0.88	0.92
	y ~ mean(0 ; 25 ; 50)	1.00	1.00	0.96	0.98	0.87	0.93	0.84	0.87	0.91	0.94
	y ~ mean(1.3 ; 25; 50)			0.99	0.99	0.98	0.99	0.96	0.98	0.98	0.98

From a practical point of view, the positions that could eventually be sampled in standing trees are 0% of HTcom, 1.3 m height (breast height position), and 25% of HTcom, even though the sampling at 25% of HTcom may require some additional effort. The reliability comparisons of different possible practical sampling strategies with standing trees are presented in Figure 3.4. For DBD, the reliability of sampling at breast height position alone was above 0.8 and marginally improved with more positions sampled. For the other traits, adding the 25% of HTcom significantly improved the correlations, indicating its importance to have higher precision in predicting whole-tree properties. With any further positions being considered, no substantial improvement was observed.

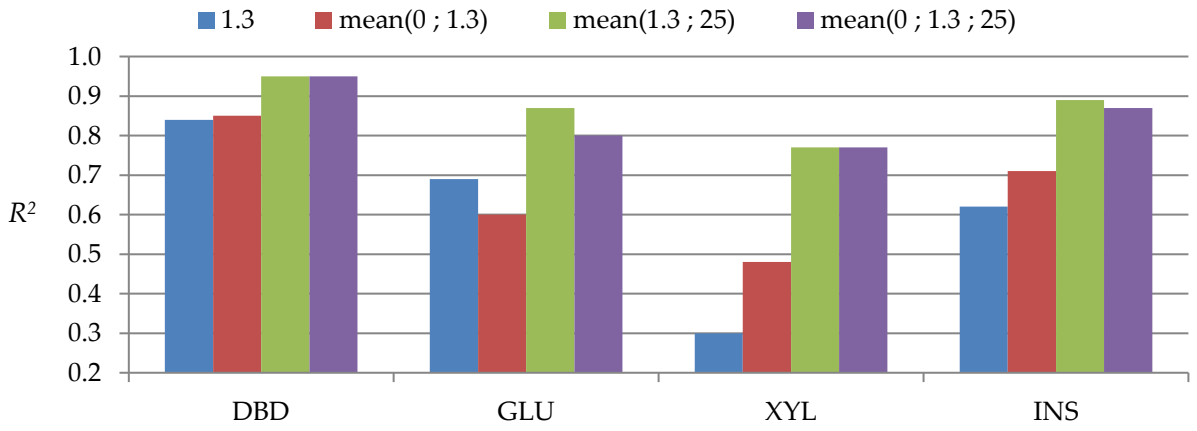


Figure 3.4. Bar plot with reliabilities of different practical standing tree sampling strategies. The position presented represents 0% of HTcom, 1.3 m (breast height position), and 25% of HTcom.

3.3.2.2 Whole-tree Properties Prediction with NIR Models

To examine the effect of indirect NIR measurements on whole-tree property prediction, a NIR prediction model was calibrated for disc basic density. The best model proved to be one where a Savitzky–Golay mathematical smoothing function was applied with a window size of seven

points (SG7) and 12 factors, resulting in a calibration R^2_c = 0.82, and the highest leave-one-out (LOO) cross-validation R^2_{cv} = 0.75. The root mean standard error of prediction was $RMSEP_{cv}$ = 22.3. This model was thereafter used for the DBD predictions with results presented at Table 3.5.

Table 3.5. Reliabilities of simple linear regression models with different sets of positions obtained by the NIR model (DBDnir) to predict whole-tree level disc basic density.

Positions	Linear model	DBD		DBD_{NIR}		RANKINGS		
		R^2	ρ	R^2	ρ	DBD_{WT}	$DBD_{1.3}$	$NIR_{1.3}$
All	*y* ~ 0+1.3+25+50+75+100	1	1	0.94	0.95	1	1	1
Alone	*y* ~ 0	0.67	0.82	0.52	0.69	2	6	2
	***y* ~ 1.3**	**0.84**	**0.91**	**0.72**	**0.80**	3	9	3
	y ~ 25	0.85	0.91	0.66	0.82	4	2	10
	y ~ 50	0.79	0.85	0.76	0.86	5	5	4
	y ~ 75	0.76	0.86	0.65	0.77	6	15	21
	y ~ 100	0.30	0.39	0.20	0.26	7	4	11
Pairwise	*y* ~ mean(0 ; 1.3)	0.85	0.92	0.71	0.79	8	12	12
	y ~ mean(0 ; 25)	0.92	0.95	0.73	0.84	9	3	6
	y ~ mean(0 ; 50)	0.85	0.89	0.75	0.83	10	11	7
	y ~ mean(1.3 ; 25)	0.95	0.97	0.83	0.88	11	7	5
	y ~ mean(1.3 ; 50)	0.94	0.96	0.89	0.92	12	13	28
	y ~ mean(25 ; 50)	0.94	0.96	0.84	0.90	13	19	17
Three-way	*y* ~ mean(0 ; 1.3 ; 25)	0.95	0.97	0.81	0.86	14	8	8
	y ~ mean(0 ; 1.3 ; 50)	0.93	0.96	0.82	0.87	15	26	31
	y ~ mean(0 ; 25 ; 50)	0.96	0.98	0.83	0.89	16	18	20
	y ~ mean(1.3 ; 25 ; 50)	0.99	0.99	0.91	0.94	17	14	23
						18	20	30
						19	10	19
						20	21	18
						21	16	9
						22	17	15
						23	30	24
						24	23	22
						25	34	36
						26	29	13
						27	22	16
						28	28	25
						29	36	32
						30	25	26
						31	32	27
						32	35	29
						33	27	14
						34	31	33
						35	24	38
						36	33	34
						37	37	39
						38	38	37
						39	39	35
						40	40	40

The reliabilities with the NIR predictions were lower than direct measurements, and the maximum efficacy obtained with all positions was R^2= 0.94. At the breast height position (1.3 m), the effect of the indirect measurement on the whole-tree prediction was a 0.12 reduction in the reliability and a 0.11 reduction in the Spearman's rank correlation. With the mean value of pairs of positions, the best combination was 1.3 m and 50%, generating an R^2= 0.89 and ρ = 0.92,

followed by 1.3 m and 25% which resulted in $R^2 = 0.83$ and $\rho = 0.88$. Reliability and ranking correlation above 0.90 were only obtained with the mean value of the three-way combination of positions 1.3 m, 25%, and 50%. The rankings obtained at whole-tree level (DBD_{WT}), breast height with direct measurements ($DBD_{1.3}$), and breast height with NIR predictions ($NIR_{1.3}$) are presented in Table 3.5.

3.4 Discussion

Despite the extensive use of *A. crassicarpa* in tropical forestry, information about the magnitude and variation of its wood properties and pulping traits is limited. In this study, a comprehensive set of wood and pulping properties estimates provided for a thorough investigation of a collection of wood properties. We destructively sampled 40 trees at multiple positions to assess the longitudinal patterns of variation in traits and to determine whole-tree property estimates. At the population level, basic density and lignin content showed a decreasing pattern along the bole from bottom to top, in contrast with carbohydrates, pulp yield, and S/G ratio which showed a more stable longitudinal pattern. However, at the individual tree level, we verified a range of longitudinal patterns following a normal distribution for all traits, estimated by the Slope100 statistic. For basic density and insoluble lignin, the slope100 varied only with negative values, indicating that, at the population mean and for all individual trees, there is a decreasing trend from the base to the top of the bole for these traits. This result is consistent with the longitudinal variation pattern for basic density reported for *A. mangium* (Chowdhury 2005) and contrary to *Eucalyptus nitens* (Kube and Raymond 2002), which showed an increasing longitudinal pattern for basic density. For kraft screened pulp yield, all individual tree slope100 values were larger than zero, indicating a consistent increasing longitudinal pattern. For the carbohydrates and S/G ratio,

slope100 varied from negative to positive values, even though the population mean longitudinal variation suggests a stable pattern with similar values across positions.

At the whole tree level, *A. crassicarpa* can be described as a tree species with medium-high wood density (DBD = 481 kg m^{-3}), high pulp yield (KPY = 0.538) and medium lignin content (LIG = 0.294). The basic density and lignin values found in this study are consistent with the results reported for the species by Laurila 1995 and Shukor *et al*. 1998. When comparing with other important acacia species, *A. crassicarpa* (AC) showed higher basic density than *A. mangium* (AM), similar basic density to *A. mangium* x *A. auriculiformis* hybrid (AMxAA), and lower basic density than *A. auriculiformis* (AA) (Moya and Muñoz 2010, Rafeadah and Rahim 2007, Jusoh *et al*. 2014, Yahya *et al*. 2010); the pulp yield was higher than AM (Griffin *at al*. 2014); the alpha cellulose was similar to AM and AMxAA, and significantly higher than AA (Yahya *et al*. 2010); lignin values were lower than AM, AA, and AMxAA reported by Yahya *et al*. 2010 , but similar to AMxAA reported by Rafaedah *et al*. 2007. Comparing AC with eucalypts, comparable values in lignin and cellulose contents were found for urograndis hybrids at similar age (Nickolas *et al*. 2020, Del-Río *et al*. 2005), and kraft pulp yield in the same ranges reported for *E. globulus* (Borralho *et al*. 1993, Nickolas *et al*. 2020, Stackpole *et al*. 2011, Del-Río *et al*. 2005) at ages ranging from 10-16 years and higher wood density. Regarding the S/G ratio, the low mean with a narrow range (1.5 – 1.8) found for AC in our study is somewhat different from typical S/G ranges reported for eucalyptus species and hybrids (Del-Río *et al*. 2005, Ohra-aho *et al*. 2013, Lima *et al*. 2019).

The phenotypic correlations among wood quality traits found in this study are, in general, similar to results reported for other woody species (Nickolas *et al*. 2020, Stackpole *et al*. 2011, Lima *et al*. 2019, Li *et al*. 2020). Kraft pulp yield was correlated with alpha cellulose, glucose, and

galactose, and negatively correlated with insoluble lignin. The S/G ratio was only significantly correlated with mannose, a minor carbohydrate, indicating that for the AC germplasm studied, S/G ratio did not greatly influence the pulp yield or its related traits. Basic density was positively correlated with pulp yield, alpha cellulose, arabinose, galactose, rhamnose, and mannose, showing its interdependency with cellulose contents in the xylem cell wall. This significant positive basic density correlation with pulp yield is not typically seen in eucalypts, with no correlation or lower values reported (Nickolas *et al*. 2020, Stackpole *et al*. 2011, Lima *et al*. 2019).

The reliability to predict the whole-tree property varied for different traits. For kraft pulp yield, the single position sampling at 25% was very reliable and should be preferred for direct measurement sampling. At the basal positions, less reliable estimates were found, but still with high ranking correlations with the whole-tree value. For basic density, sampling at breast height was the best single position and could explain 84% of the whole-tree variation, with a 0.91 rank correlation. For glucose and insoluble lignin, the two chemical traits with the largest impact on pulp yield, the reliability of single positions was lower than for wood density. For xylose, the second most abundant hemicellulose component, the single positions reliabilities were even lower. These results resemble reports for *E. globulus* and *E. nitens*, with reliabilities of breast height sampling for basic density of 0.82 and 0.89, respectively, higher than 0.60 found for cellulose content (Raymond and Muneri 2001, Kube and Raymond 2002). In hybrid poplars, Schimleck *et al*. 2005 found reliability of 0.65 for increment core cellulose content at breast height. For all traits, with two data points collected along the bole, the reliabilities were improved. The line traced with two estimates obtained at different positions in the bole allows for a crude sense of the individual tree level longitudinal variation pattern, and a more reliable estimate of the whole-tree property can be established. Looking for a practical multi-position sampling strategy, we evaluated the

reliabilities of mean values of combinations of positions because it allows predicting the whole-tree value without a statistical model by simply averaging the values obtained at each position considered. With the mean of 1.3 m (breast height) and 25%, the reliabilities were significantly improved, with ranking correlations above 0.90. The addition of subsequent positions marginally increased the statistics.

Several indirect measurement methods were developed for a range of physical, mechanical, and chemical wood properties in a collection of woody species (Schimleck *et al*. 2019). Regardless of the method, its efficiency depends on the correlation between predictions and the actual phenotype, typically expressed by the regression model R^2. We have chosen NIR models for disc basic density to quantify the effect of indirect measurements in whole-tree wood properties prediction because we have selected trees to maximize the DBD range, and it has the largest data set comprised of 240 direct measurements/NIR spectra data pairs, thus, satisfying two important requirements for calibrating good NIR models, sample size and variability for the trait (Meder *et al*. 2010, Schimleck 2008, Downes *et al*. 1997). Working with AM, Karlinasari *et al*. 2014 generated NIR models for α-cellulose and hemicelluloses with good calibration $R^2 \approx 0.80$, different of lignin and extractives content with poor quality and lower R^2 of 0.41 and 0.54, respectively. Hodge *et al*. 2018 developed global NIR models for five eucalyptus species and found higher cross-validation R^2 for lignin-related traits, with lignin, insoluble lignin, and syringyl–guaiacyl ratio R^2 of 0.95, 0.96, and 0.86, respectively. The global models for sugar content were slightly inferior, with R^2_{cv} of 0.74 for glucose, 0.89 for xylose, and from 0.72 to 0.91 for the minor sugars.

In our study, the selected disc basic density NIR model had calibration R^2 of 0.82 and cross-validation R^2_{cv} of 0.75, values in the typical range of good NIR models reported for wood properties. The predictions with this model slightly reduced the reliability for all sampling

positions scenarios. In our disc basic density case, starting with reliable breast height sampling, the precision of the indirect measurement did not greatly affect the overall whole-tree estimate. From a practical point, with the 0.80 ranking correlation obtained with NIR sampling at breast height, tree breeders would still efficiently select the best-ranked genotypes. If the purpose of the phenotyping requires even more precise estimates of the individual whole-tree level, then multiple positions sampling may be necessary. The mean value of the positions 1.3 m (breast height) and 25%, still executable in standing trees, could predict the whole-tree value with higher reliability and ranking correlation. For a near-perfect prediction, three-position sampling with NIR at positions 1.3 m, 25%, and 50% showed reliability and ranking correlation above 0.90, at the expense of felling the tree for the measurements. Generally, the overall precision was proportional to the product of the reliability of the positions set considered, and the Pearson correlation between the NIR prediction and the actual phenotype, i.e., the square root of the indirect method model R^2. Aiming for high accuracy whole-tree phenotyping, the two statistics must be considered. Sampling a tree with a precise indirect method in a single position with low reliability will inaccurately predict the whole-tree property; similarly, sampling a tree in a single position with high reliability using an imprecise indirect method will also provide inaccurate predictions of the whole-tree property.

3.5 Conclusion

The wood and pulping properties estimates obtained in the present study, with second-generation *A. crassicarpa* trees, show that the species has wood suitable for efficient pulp production, with lignin contents, carbohydrates contents, and kraft pulp yields in the range of the hardwoods commercially planted around the world. The within-tree longitudinal pattern of variation was described as a normally distributed numerical trait – the “Slope100”, and for basic

density and insoluble lignin content, showed a consistent decreasing trend from base to top of the bole. In contrast, for the carbohydrates, soluble lignin, and S/G ratio, no consistent pattern was observed. The reliability of sets of positions taken singly or combined to predict the whole-tree phenotype varied along with the different traits, and for pulp yield, basic density, glucose content, and lignin content reliable ground-level direct measurement sampling was found, with very high ranking correlations. With a NIR prediction model of basic density with observed cross-validation $R^2_{cv} = 0.75$, a 0.12 reduction in the reliability of breast height sampling was verified, but still with a 0.80 Spearman ranking correlation, which could efficiently rank the trees for selection in a breeding program. Multiple position sampling can be performed together with indirect measurements to achieve a near-perfect whole-tree property estimate. A strategy of sampling standing trees at breast height and 25% of the commercial height, and using the mean value of those positions will be a high degree of accuracy for individual whole-tree level phenotyping.

3.6 References

Acosta, J.J.; Castillo, M.S.; Hodge, G.R. Comparison of benchtop and handheld near infrared spectroscopy devices to determine forage nutritive value. Crop Science, 2020, 60, 3410-3422.

Atwood, R.A.; White, T.L.; Huber, D.A. Genetic parameters and gains for growth and wood properties in Florida source loblolly pine in the southeastern United States. Can. J. For. Res. 2002, 32:6, 1025-1038.

Bakri, M.K.B.; Jayamani, E.; Hamdan, S.; Rahman, R.; Kakar, A. Potential of Borneo acacia wood in fully biodegradable bio-composites' commercial production and application. Polym. Bull. 2018, 75, 5333–5354.

Barnett, J.R.; Jeronimidis, G. Wood quality and its biological basis. 1st ed.; Blackwell publishing Ltd: 9600 Garsington Road, Oxford OX4 2DQ, UK, 2003.

Barry, K.M.; Mihara, R; Davies, N.W.; Mitsunaga, T.; Mohammed, C.L. Polyphenols in Acacia mangium and Acacia auriculiformis heartwood with reference to heart rot susceptibility. J. Wood Sci. 2005, 51, 615-621.

Borralho, N.M.G; Cotterill, P.P.; Kanowski, P.J. Breeding objectives for pulp production of Eucalyptus globulus under different industrial cost structures. Can. J. For. Res. 1993, 23, 648–656.

Bouvet, J.; Ekomono, C.G.M.; Brendel, O.; Laclau, J.; Bouillet, J.; Epron, D. Selecting for water use efficiency, wood chemical traits and biomass with genomic selection in a Eucalyptus breeding program. For. Eco. Manag. 2020, 465, 1-10.

Breunig, M.; Hans-Peter, K.; Raymond, T.; Sander, J. LOF: identifying density-based local outliers. ACM Sigmod Record 2000, 29, 93–104.

Chen, C.; Duan, C.; Li, J.; Liu, Y.; Ma, X.; Zheng, L.; Stavik, J.; Ni, Y. Cellulose (dissolving pulp) manufacturing processes and properties: a mini-review. Bioresources 2016, 11, 5553–5564.

Chowdhury, M.Q.; Ishiguri, F.; Iizuka, K.; Hiraiwa, T.; Matsumoto, K.; Takashima, Y.; Yokota, S.; Yoshizawa, N. Wood property variation in Acacia auriculiformis growing in Bangladesh. Wood Fiber Sci. 2009, 41(4), 359–365.

Chowdhury, M.Q.; Shams, M.I.; Alam, M. Effects of age and height variation on physical properties of mangium (Acacia mangium Willd.) wood. Aust. For. 2005, 68(1), 17-19.

Clark, N.B.; Balodis, V.; Fang, G.; Wang, J. Pulping potential of acacias. In Australian tree species research in China. Proceedings of an international workshop held in Zhangzhou, China, 2-5 November 1992; Brown, A.G., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 48., Canberra, Australia, 1994; pp. 196-202.

Clark, N.B.; Balodis, V.; Fang, G.; Wang, J. Pulping properties of tropical acacias. In Advances in tropical acacia research, Proceedings of an international workshop held in Bangkok, Thailand, 11-15 February 1991; Turnbull, J.W., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 35., Canberra, Australia, 1991; pp. 138-144.

Cornelius, J. Heritabilities and additive genetic coefficients of variation in forest trees. Can. J. For. Res. 1994, 24, 372–379.

Lima, B.M.; Cappa, E.P.; Silva-Junior, O.B.; Garcia, C.; Mansfield, S.D.; Grattapaglia, D. Quantitative genetic parameters for growth and wood properties in Eucalyptus "urograndis" hybrid using near-infrared phenotyping and genome-wide SNP-based relationships. PLoS ONE 2019, 14 (6).

Del-Río, J. C.; Gutiérrez, A.; Hernando, M.; Landín, P.; Romero, J.; Martínez, A. T. Determining the influence of eucalypt lignin composition in paper pulp yield using Py-GC/MS. J. Anal. Appl. Pyro. 2005, 74, 110-115.

Dinh Kha, L.; Harwood, C.E.; Kien, N.D.; Baltunis, B.S.; Dinh Hai, N.; Thinh, H.H. Growth and wood basic density of acacia hybrid clones at three locations in Vietnam. New For. 2012, 43, 13–29.

Downes, G.M.; Hudson, I.L.; Raymond, C.A.; Dean, G.H.; Michell, A.J.; Schimleck, L.R.; Evans, R.; Muneri, A. Sampling plantation eucalypts for wood and fibre properties. CSIRO, Australia, 1997, 144 p.

Fundová, I. Quantitative genetics of wood quality traits in scots pine. Doctoral Thesis, Swedish University of Agricultural Sciences, Umeå, Sweden, 2020.

Gao, S.; Wang, X.; Wiemann, M.C.; Brashaw, B.K.; Ross, R.J.; Wang, L. A critical analysis of methods for rapid and non-destructive determination of wood density in standing trees. Ann. For. Sci. 2017, 74, 27.

Greaves, B.L.; Borralho, N.M.G.; Raymond, C.A. Breeding objective for plantation eucalypts grown for production of kraft pulp. For. Sci. 1997, 43:4, 465-472.

Griffin, A.R.; Midgley, S.J.; Bush, D.; Cunningham, P.J.; Rinaudo, A.T. Global uses of Australian acacias – recent trends and future prospects. Dive. and Dist. 2011, 17, 837–847.

Griffin, A.R.; Twayi, H.; Braunstein, R.; Downes, G.M.; Son, D.H.; Harwood, C.E. A comparison of fibre and pulp properties of diploid and tetraploid Acacia mangium grown in Vietnam. Appita J. 2014, 67(1), 43-49.

Hart, J.F.; de Araujo, F.; Thomas, B.R.; Mansfield, S.D. Wood quality and growth characterization across intra- and inter-Specific hybrid aspen clones. Forests 2013, 4, 786-807.

Harwood, C.E.; Hardiyanto, E.B.; Wong, C.Y. Genetic improvement of tropical acacias: achievements and challenges. South. For. 2015, 77, 11-18.

Hodge, G.R.; Acosta, J.J.; Unda, F.; Woodbridge, W.C.; Mansfield, S.D. Global near infrared spectroscopy models to predict wood chemical properties of Eucalyptus. J. Near Inf. Spectrosc. 2018, 26(2), 117-132.

Igartúa, D.V.; Monteoliva, S.E.; Monterubbianesi, M.G. ; Villegas, M.S. Basic density and fibre length at breast height of Eucalyptus globulus ssp. globulus for parameter prediction of the whole tree. IAWA J. 2003, 24(2),173–184.

Jin, J.; Zhao, X.; Liu, H.; Wang, S.; Song, Z.; Ma, X.; Li, K. Preliminary study on genetic variation of growth traits and wood properties and superior clones selection of Populus ussuriensis Kom. iForest – Biog. For. 2019, 12(5), 459-466.

Jones, P.D.; Schimleck, L.R.; Daniels, R.F.; Clark III, A.; Purnell, R.C. Comparison of Pinus taeda L. whole-tree wood property calibrations using diffuse reflectance near infrared spectra obtained using a variety of sampling options. Wood Sci. Tech. 2008, 42,385–400.

Jusoh, I.; Zaharin, F.A.; Adam, N.S. Wood quality of Acacia hybrid and second-generation Acacia mangium. Bioresources 2014, 9(1), 150-160.

Karlinasari, L.; Sabed, M.; Wistara, I.N.J.; Purwanto, Y.A. Near infrared (NIR) spectroscopy for estimating the chemical composition of (Acacia mangium Willd.) wood. J. Indi. Acad. Wood Sci. 2014, 11, 162–167.

Katz, J.L.; Spencer, P.; Wang, W.; Misra, A.; Marangos, O.; Friis, L. On the anisotropic elastic properties of woods. J. Mater. Sci. 2008, 43,139–145.

Kube, P.; Raymond, C.A. Prediction of whole-tree basic density and pulp yield using wood core samples in Eucalyptus nitens. Appita J. 2002, 55, 43-48.

Kumar, H.; Christopher, L.P. Recent trends and developments in dissolving pulp production and application. Cellulose 2017, 24, 2347-2365.

Laurila, R. Wood properties and utilization potential of eight fast-growing tropical plantation tree species. J. Trop. For. Prod. 1995, 1(2), 209-221.

Li, Y.; Ding, X.; Jiang, J.; Luan, Q. Inheritance and correlation analysis of pulpwood properties, wood density, and growth traits of slash pine. Forests 2020, 11, 493.

Lundberg, V.; Bood, J.; Nilsson, L.; Axelsson, E.; Berntsson, T.; Svensson, E. Converting a kraft pulp mill into a multi-product biorefinery: techno-economic analysis of a case mill. Clean Technol. Environ. 2014, 16,1411–1422.

Marinova, M.; Mateos-Espejel, E.; Paris, J. From kraft mills to forest biorefinery: an energy and water perspective II. Case study. Cell Chem. Technol. 2010, 44(1-3),21-26.

Marinova, M; Mateos-Espejel, E.; Jemaa, N.; Paris, J. Addressing the increased energy demand of a kraft mill biorefinery: the hemicellulose extraction case. Chem. Eng. Res. Des. 2009, 87, 1269–1275.

McKinnon, G.E.; Larcombe, M.J.; Griffin, A.R.; Vaillancourt, R.E. Development of microsatellites using next-generation sequencing for Acacia crassicarpa. J. Trop. For. Sci. 2018, 30:2, 252-58.

Meder, R.; Trung, T.; Schimleck, L.R. Seeing the wood in the trees: unleashing the secrets of wood via near infrared spectroscopy. J. Near Inf. Spectrosc. 2010, 18, v-vii.

Mendham, D.S.; White, D.A. A review of nutrient, water and organic matter dynamics of tropical acacias on mineral soils for improved management in Southeast Asia. Aust. For. 2019, 82:sup1, 45-56.

Mevik, B.H.; Wehrens, R.; Liland, K.H. PLS: partial least squares and principal component regression; R Package Version 2.4-3; R Foundation for Statistical Computing: Vienna, Austria, 2018.

Moya, R.; Muñoz, F. Physical and mechanical properties of eight fast-growing plantation species in Costa Rica. J. Trop. For. Sci. 2010, 22(3), 317–328.

Nambiar, E.K.S.; Harwood, C.E. Productivity of acacia and eucalypt plantations in Southeast Asia. 1. Bio-physical determinants of production: opportunities and challenges. Int. For. Rev. 2014, 16(2), 225–248.

Nambiar, E.K.S.; Harwood, C.E.; Mendham, D.S. Paths to sustainable wood supply to the pulp and paper industry in Indonesia after diseases have forced a change of species from acacia to eucalypts. Aust. For. 2018, 81:3, 148-161.

Neale, D.B.; Mitchell, M.S.; Brown, G.R. Molecular dissection of the quantitative inheritance of wood property traits in loblolly pine. Ann. For. Sci. 2002, 59, 595–605.

Nickolas, H.; Williams, D.; Downes, G.; Tilyard, P.; Harrison, P.A.; Vaillancourt, R.E.; Potts, B. Genetic correlations among pulpwood and solid-wood selection traits in Eucalyptus globulus. New For. 2020, 51, 137–158.

Niemczyk, M.; Przybysz, P.; Przybysz, K.; Kaliszewski, A.; Wojda, T.; Liesebach, M. Productivity, growth patterns, and cellulosic pulp properties of hybrid aspen clones. Forests 2019, 10, 450.

Nirsatmanto A.; Sunarti, S. Genetics and Breeding of Tropical Acacias for Forest Products: Acacia mangium, A. auriculiformis and A. crassicarpa. In: Advances in Plant Breeding Strategies: Industrial and Food Crops; Al-Khayri J., Jain S., Eds.; Springer, Cham: Springer Nature, Switzerland, 2019; Volume 6, pp. 3-28.

Nugroho, W.D.; Marsoem, S.N.; Yasue, K.; Fujiwara, T.; Nakajima, T.; Hayakawa, M.; Nakaba, S.; Yamagishi, Y.; Jin, H.O.; Kubo, T.; Funada, R. Radial variations in the anatomical characteristics and density of the wood of Acacia mangium of five different provenances in Indonesia. J. Wood Sci. 2012, 58,185–194.

Ohra-aho, T.; Gomes, F.J.B.; Colodette, J.L.; Tamminen, T. S/G ratio and lignin structure among Eucalyptus hybrids determined by Py-GC/MS and nitrobenzene oxidation. J. Anal. Appl. Pyro. 2013, 101, 166-171.

Ohshima, J.; Yokota, S.; Yoshizawa, N.; Ona, T. Examination of within-tree variations and the heights representing whole-tree values of derived wood properties for quasi-non-destructive breeding of Eucalyptus camaldulensis and Eucalyptus globulus as quality pulpwood. J. Wood Sci. 2005, 51,102–111.

Ona, T.; Sonoda, T.; Ito, K.; Shibata, M.; Tamai, Y.; Kojima, Y.; Ohshima, J.; Yokota, S.; Yoshizawa, N. Investigation of relationships between cell and pulp properties in Eucalyptus by examination of within-tree property variations. Wood Sci. Tech. 2001, 35, 229-243.

Plomion, C.; Leprovost, G.; Stokes, A. Wood formation in trees. Plant Physiol. 2001, 127, 1513–1523.

Porth, I.; Klápště, J.; Skyba, O.; Lai, B.S.K.; Geraldes, A.; Muchero, W.; Tuskan, G.A.; Douglas, C.J.; El-Kassaby, Y.A.; Mansfield, S.D. Populus trichocarpa cell wall chemistry and ultrastructure trait variation, genetic control, and genetic correlations. New Phyt. 2013, 197, 777-790.

R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing: https://www.R-project.org/ (accessed 24 Jun 2020).

Rafeadah, R.; Rahim, S. Chemical and physical properties of juvenile acacia hybrid and Azadiracta Excelsa. J. Inst. Wood Sci. 2007, 17(5), 290-294.

Raymond, C.A. Genetics of Eucalyptus wood properties. Ann. For. Sci. 2002, 59, 525-531.

Raymond, C.A.; Muneri, A. Non-destructive sampling of Eucalyptus globulus and E. nitens for wood properties. I. Basic Density. Wood Sci. Tech. 2001, 35, 27-39.

Raymond, C.A.; Schimleck, L.R.; Muneri, A.; Michell, A.J. Non-destructive sampling of Eucalyptus globulus and E. nitens for wood properties. III. Predicted pulp yield using near infrared reflectance analysis. Wood Sci. Tech. 2001, 35, 203-215.

Rezende, G.D.S.P.; Resende, M.D.V.; Assis, T.F. Eucalyptus Breeding for Clonal Forestry. In: Fenning Challenges and Opportunities for the World's Forests in the 21st Century. Fenning, T.; Springer, Dordrecht, 2014; Forestry Sciences, vol 81, pp. 393-424.

Robinson, A.R.; Mansfield, S.D. Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. The Plant J. 2009, 58, 706-714.

Rokeya, U.K.; Hossain, M.A.; Ali, R.A.; Paul, S.P. Physical and mechanical properties of (Acacia auriculiformis × A. Mangium) hybrid acacia. J. Bangladesh Acad. Sci. 2010, 34(2), 181-187.

Scandinavian pulp and paper board testing committee. Woodchips for pulp production, size distribution, SCAN-CM 40:01. Stockholm, Sweden, 2001.

Schimleck, L.; Dahlen, J.; Apiolaza, L.A.; Downes, G.; Emms, G.; Evans, R.; Moore, J.; Pâques, L.; Van den Bulcke, J.; Wang, X. Non-destructive evaluation techniques and what they tell us about wood property variation. Forests 2019, 10, 728.

Schimleck, L.R. Near infrared spectroscopy: a rapid, non-destructive method for measuring wood properties, and its application to tree breeding. New Zeal. J. For. Sci. 2008, 38(1), 14-35.

Schimleck, L.R.; Kube, P.D.; Raymond, C.A. Genetic improvement of kraft pulp yield in Eucalyptus nitens using cellulose content determined by near infrared spectroscopy. Can. J. For. Res. 2004, 34, 2363–2370.

Schimleck, L.R.; Kube, P.D.; Raymond, C.A.; Michell, A.J.; French, J. Estimation of whole-tree kraft pulp yield of Eucalyptus nitens using near-infrared spectra collected from increment cores. Can. J. For. Res. 2005, 35, 2797–2805.

Schimleck, L.R.; Payne, P.; Wearne, R.H. Determination of important pulp properties of hybrid poplar by near infrared spectroscopy. Wood and Fiber Sci. 2005, 37(3), 462 – 471.

Schimleck, L.R.; Rezende, G.D.S.P.; Demuner, B.J.; Downes, G.M. Estimation of whole-tree wood quality traits using near infrared spectra from increment cores. Appita J. 2006, 59, 231.

Schmidt, E.A. A practical model relating kraft pulping costs to hardwood chemical properties and morphology. Appita J 2005, 58, 218–224.

Shukor, N.A.A.; Nang, A.N.; Awang, K. Selected wood properties of Acacia auriculiformis and A. crassicarpa provenances in Malaysia. In Recent developments in acacia planting. Proceedings of an international workshop held in Hanoi, Vietnam, 27-30 October 1997; Turnbull, J.W., Crompton, H.R., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 82., Canberra, Australia, 1998; pp. 155-160.

Stackpole, D.J.; Vaillancourt, R.E.; Alves, A.; Rodrigues, J.; Potts, B.M. Genetic variation in the chemical components of Eucalyptus globulus wood. G3 (Bethesda) 2011, 1(2), 151-159.

Sykes, R.W.; Isik, F.; Li, B.; Kadla, J.; Chang, H-M. Genetic variation of juvenile wood properties in a loblolly pine progeny test. TAPPI J. 2003, 2(12), 3-8.

TAPPI - Technical Association of the Pulp and Paper Industry. TAPPI test methods T 236 om-99: Kappa number of pulp. Atlanta: Tappi Technology Park, 1999.

TAPPI - Technical Association of the Pulp and Paper Industry. TAPPI test methods T 257 cm-02: Sampling and preparing wood for analysis. Atlanta: Tappi Technology Park, 2002.

TAPPI - Technical Association of the Pulp and Paper Industry. TAPPI test methods T 258 om-16: Basic density and moisture content of pulpwood. Atlanta: Tappi Technology Park, 2016.

TAPPI - Technical Association of the Pulp and Paper Industry. TAPPI test methods T 550 om-08: Determination of equilibrium moisture in pulp, paper and paperboard for chemical analysis. Atlanta: Tappi Technology Park, 2008.

Tonouéwa, J.M.F.M.; Langbour, P.; Biaou, S.S.H.; Assèdé, E.S.P.; Guibal, D.; Kouchade, C.A.; Kounouhéwa, B.B. Anatomical and physico-mechanical properties of Acacia auriculiformis wood in relation to age and soil in Benin, West Africa. Euro. J. Wood Wood Prod. 2020, 78, 745-756.

Von Arx, G.; Crivellaro, A.; Prendin, A.L.; Čufar, K.; Carrer, M. Quantitative Wood Anatomy—Practical Guidelines. Front. Plant Sci. 2016, 7, 781.

White, T.L. A conceptual framework for tree improvement programs. New For. 1987, 1, 325–342.

Yahya, R.; Sugiyama, J.; Silsia, D.; Gril, J. Some anatomical features of an Acacia hybrid, A. mangium and A. auriculiformis grown in Indonesia with regard to pulp yield and paper strength. J. Trop. For. Sci. 2010, 22(3), 343-351.

Yao, S.; Wu, G.; Xing, M.; Zhou, S.; Pu, J. Determination of lignin content in Acacia spp. using near-infrared reflectance spectroscopy. Bioresources 2010, 5(2), 556-562.

Zobel, B.J.; Talbert, J. Applied Forest Tree Improvement; John Wiley & Sons: New York, NY, USA, 1984; pp. 375–413.

Zobel, B.J.; van Buijtenen, J.P. Wood variation: its causes and control. Springer-Verlag, Berlin, Heidelberg, New York. 1989.

CHAPTER 4

Genetic Control of Quantitative Traits in an *Acacia crassicarpa* Multi-Environment Progeny Trial

4.1 Introduction

Acacia crassicarpa A.Cunn. ex Benth. is an important tree species in Southeast Asia (Nambiar and Harwood 2014; Harwood and Nambiar 2014) and a valuable option for tropical industrial forestry (Turnbull *et al*. 1998; Midgley and Turnbull 2003) with plantations supported by advancements in silviculture (Mendham and White 2019), genetic improvement (Harwood *et al*. 2015), and molecular biology (Nirsatmanto and Sunarti 2019, McKinnon *et al*. 2018). The species is naturally distributed in tropical North-eastern Australia and New Guinea island, and provenance testing of wild collections in the 1990s showed the potential to select for good bole form, vigorous growth on poorly drained acidic soils, and resistance to pests and diseases. It is an important component of the wood supply of large integrated pulp and paper industries (Nambiar *et al.* 2018). The main breeding objectives are to improve the wood production per area unit and the wood quality to optimize the pulping process (Martins *et al*. 2020).

A. crassicarpa flowers are hermaphrodites with strictly protogynous flowers. Although technically feasible, controlled pollination of acacias is difficult, and the percentage of flowers that develop into pods is typically less than 5%. In the breeding program of April Asia, even with recommended methods such as “inflorescence pollination” (Griffin *et al*. 2010), a huge effort is required to produce more than a few crosses yearly. Therefore, controlled pollination is not practical for advancing breeding populations which require hundreds of crosses for each breeding cycle. Breeding populations have been bred by open pollination. Typically, progeny trials are established, and family and within-family selections are made to convert them into “seedling seed

orchards" (SSO). After thinning and a subsequent general flowering, the open-pollinated seed is collected from the best trees of the better-performing families to establish second-generation progeny trials. Clonal breeding orchard (CBO) establishment, which brings together elite selections captured by marcotting or grafting, is also practiced to increase selection intensity and genetic gain over that achievable from SSOs (Harwood *et al*. 2015).

In principle, clonal forestry should deliver the greatest genetic gain at any stage of a breeding program by deploying both additive and non-additive gain (Zobel 1993). It requires mass-production of juvenile propagules of selected trees as implemented for *Eucalyptus* (Rezende *et al*. 2014). For *A. crassicarpa*, vegetative propagation is cost-effective, but is possible only with juvenile ortets. With current vegetative propagation techniques, by the time phenotypes can be measured for assessment of genetic value, the selected tree can only be propagated by grafting. *Family forestry* can be defined as the commercial deployment of half-sib (open-pollinated) or full-sib (control-pollinated) family blocks. Family forestry can be done using either seedlings or vegetative propagules, such as rooted cuttings. In Indonesia, family forestry using vegetative propagules (sometimes called clonal family forestry) has been successfully employed with acacia (Wong and Yuliarto 2014). Selected female parents tested using progeny in open-pollination families will produce a limited amount of improved seeds in seed orchards. This seed can be sown and vegetatively propagated to bulk-up nursery hedges supporting the deployment of cuttings. The main limitation of this strategy with open-pollination families is that selection is based only on female additive effects. Still, it has been widely used due to the inefficiency of controlled crossing. The advantage of full-sib family deployment over open-pollinated family deployment is significant. The full parental control and potential capture of gain from specific combining ability

can capitalize on the full-sib family genetic value, delivering larger genetic gains and higher uniformity (White *et al*. 2007).

The successful application of molecular markers to perform parentage analysis using different genotyping platforms, statistical methods, and software has been reported for several animal and plant species (Jones and Ardren 2003; Jones *et al*. 2010). For *A. crassicarpa*, using molecular markers to reconstruct pedigrees could be of great value in determining full-sib families and managing kinship in breeding populations. With full-sib family models, breeders can better model the genetics of complex quantitative traits and their association with phenotypic variation. Quantitative genetics researchers are often interested in partitioning observed phenotypic variance into causal genetic and environmental components. Typically, this involves the use of linear mixed models that can handle unbalanced data and complex experimental designs to deal with a high level of field heterogeneity due to differences in soil type, fertility, water holding capacity, etc., and a high number of genetic treatments (Isik *et al.* 2017), common to tree breeding trials with large tree plot sizes.

Considering the biological constraints of the species, the objectives of this study were to 1) To use pedigree reconstruction to identify full-sib families of *A. crassicarpa* and establish a large full-sib family progeny trial to determine the genetic control of important traits; 2) To estimate additive and specific combining ability effects; 3) To determine the genetic correlations among important economic traits.

4.2 Material and Methods

4.2.1 Germplasm and Field Trials

The APRIL Asia breeding program in Riau province, Sumatra, Indonesia, collected twenty-four open-pollination families from elite clonal seed orchards. The open-pollination

families were sown and sampled at the nursery for parentage determination with a proprietary panel of 42 SNP markers developed from an internal genome sequencing effort to provide a high-throughput and cost-effective genotyping platform capable of accurately assigning *A. crassicarpa* genotypes. SNP markers were selected based on minor allele frequency (MAF), broad distribution across the genome, including representation of each of the 13 chromosomes, and uniqueness of sequence surrounding the SNP. Orion Biosains (Puchong, Malaysia) designed, optimized, and performed assays utilizing the LGC Array Tape genotyping platform.

The paternity was determined by exclusion, calculating the proportion of matching alleles between the male candidate and offspring in each open-pollination family with a known female genotype. There were 18 genotypes in the elite orchards, all considered candidate male parents. The parent-offspring combination with the highest average concordance across markers above 0.8 was assigned. This way, 93 putative full-sib families were organized in the nursery with varying numbers of individuals. The full-sib families with enough seedlings passing the nursery quality control were included in genetic trials established on three representative Histosol environments (E) in Riau province, Sumatra, Indonesia, being: E1) 80 families in Pelalawan location with hemic peat soil; E2) 93 families in Merenti location with sapric peat soil; E3) 69 families in Pelalawan with sapric peat soil. The trial design was a resolvable alpha-lattice with five replicates and 8, 11, and 9 incomplete blocks within replicate on E1, E2, and E3, respectively. Plots were 16 tree blocks composed of four rows and four trees per row, with a spacing of 2 m between trees within a row and 3 m between rows. Families were assigned to plots within incomplete blocks following the alpha-lattice design generated and optimized by the software CycDesign (Whitaker *et al*. 2001).

The phenotypic assessments of diameter at breast height (DBH) measured in centimeters, tree height (HT) measured in meters, and the binomial survival scores (SUR) were performed at

12, 18, 24, 30, and 36 months of age. A scoring system comprising six levels was used to assess tree straightness (STR) at 36 months, with STR=1 representing a very crooked tree and STR=6 representing a perfectly straight tree. To increase the phenotyping quality, the traits DBH, HT, SUR, and STR were independently measured twice at age 36m, and the phenotype was the average value of the two assessments. At the same age, resistograph measurements with an IML-RESI PD400 device were performed to predict wood basic density (BD), expressed as the dry mass in kilograms over wet volume in cubic meters (kg m^{-3}). Individual under-bark tree volume (VOL), expressed in m^3, was calculated with the DBH and HT measurements. Both basic density and individual tree volume predictions were obtained with statistical models developed and validated for *A. crassicarpa* and routinely used by the company. The mean annual increment (MAI), measured in cubic meters of wood per hectare per year (m^3 ha^{-1} $year^{-1}$), was calculated for individual trees by adjusting tree volume for the plot mean survival. For each plot, dead trees were treated as missing values, and live trees had their volume penalized by the plot's survival. For example, for each live tree in a given plot with survival 0.75, $MAI = \frac{Volume * (10{,}000/spacing) * Plot\ Survival}{Age} = \frac{Volume * 1666.6 * 0.75}{3}$. The MAI Ton ($MAI_T$), measured in metric tons of wood per hectare per year (ton ha^{-1} $year^{-1}$), was calculated as $MAI_T = MAI * \frac{BD}{1000}$.

4.2.2 SNP Genotyping and Parentage Reconstruction

At 12 months, 16,899 live trees were sampled for DNA extraction and SNP genotyping with a mid-density panel of 700 SNP markers. The panel was developed using a multiplexed PCR amplicon/sequencing-based approach (AmpSeq). The panel validation consisted of 16 dried leaf samples. For each sample, DNA was extracted in three technical replicates performed by Orion Genomics (St. Louis, MO) using a proprietary extraction protocol optimized for AmpSeq robustness. Each extraction was then genotyped by the AmpSeq method three independent times.

Seedlings were sampled from known mothers and assigned to candidate fathers using the paternity analysis of the software CERVUS version 3.0.7 (Marshall *et al*. 1998; Kalinowski *et al*. 2007). The company had previously genotyped all its clonal breeding orchards, and a collection of 800 genotypes determined from this campaign were considered candidate fathers. The simulation of paternity analysis, utilized by the software to estimate the resolving power of the markers given their allele frequencies and to estimate critical values of the log-likelihood statistics, was performed with the following parameters: 10,000 individuals; 800 candidate parents; 0.95 proportion of sampled parents; 0.01 proportion of mistyped loci; and the option to test for self-fertilization. All other parameters of the paternity analysis followed the software default. Offspring assigned to a candidate father, given the known mother, with a trio confidence level equal to or above 99% were considered for downstream analysis.

4.2.3 Quantitative Genetic Analysis

Initially, an exploratory analysis was performed for all traits at all ages to assess the population's phenotypic variability. Next, at 36 months, single-environment mixed models (Model 1) were fit for all traits to assess the genetic variances in each environment separately.

$$\boldsymbol{y_{ijk} = \mu + T_i + F_j + B(R)_k + e_{ijk}} \qquad \text{(Model 1)}$$

y_{ijk} = phenotype measured in the i^{th} tree of family j in the incomplete block k nested within replication. $y_{ijk} \sim N\,[E(y), \sigma_y^2]$;

μ = intercept;

T_i = random effect of the individual tree. $T_i \sim N\,(0, \sigma_{tree}^2)$.

F_j = random effect of full-sib family. $F_j \sim N\,(0, \sigma_{fam}^2)$.

$B(R)_k$ = random effect of block k nested within replication. $B_k \sim N\,(0, \sigma_B^2)$.

e_{ijk} = residual effect of observation $y_{ijk} \sim N\,(0, \sigma_e^2)$.

The binary response variable survival was modeled here and in all downstream analyses with Generalized Linear Mixed Models (GLMM) with a logit link function, assuming underlying residuals to have a standard logistic distribution with variance $\pi^2/3$ (Rodriguez 2003). The narrow-sense heritability (h_a^2) and the reliability of dominance effects (h_d^2) were calculated as:

$$h_a^2 = \frac{\sigma_A^2}{\sigma_{phenotypic}^2} = \frac{\sigma_{tree}^2}{\sigma_{tree}^2 + \sigma_{fam}^2 + \sigma_e^2} \quad \text{(Equation 1)}$$

$$h_d^2 = \frac{\sigma_D^2}{\sigma_{phenotypic}^2} = \frac{4 * \sigma_{fam}^2}{\sigma_{tree}^2 + \sigma_{fam}^2 + \sigma_e^2} \quad \text{(Equation 2)}$$

The multi-environment trial analysis was performed following model 2.

$$\boldsymbol{y_{sijk} = \mu + E_s + T_i + ET_{si} + F_j + EF_{sj} + B(E)_{sk} + e_{sijk}} \quad \text{(Model 2)}$$

y_{sijk} = phenotype measured in the i^{th} tree of family j in the incomplete block k at environment s.

$y_{sijk} \sim N\ [E(y), \sigma_y^2]$;

μ = intercept;

E_s = fixed effect of environment s.

T_i = random effect of the individual tree. $T_i \sim N\ (0, \sigma_{tree}^2)$.

ET_{si} = random interaction between tree i and environment s. $ET_{si} \sim N\ (0, \sigma_{E.tree}^2)$;

F_j = random effect of full-sib family j. $F_j \sim N\ (0, \sigma_{fam}^2)$.

EF_{sj} = random interaction between family j and environment s. $EF_{sj} \sim N\ (0, \sigma_{E.fam}^2)$;

$B(E)_{sk}$ = random effect of block k nested within environment s. $B_k \sim N\ (0, \sigma_B^2)$.

e_{sijk} = residual effect of observation $y_{sijk} \sim N\ (0, \sigma_e^2)$;

From model 2 variance components, the following genetic parameters were obtained:

Additive genetic variance: $\sigma_A^2 = \sigma_{tree}^2$ (Equation 3)

Dominance genetic variance: $\sigma_D^2 = 4 * \sigma_{fam}^2$ (Equation 4)

Additive by environment interaction variance: $\sigma_{E.A}^2 = \sigma_{E.tree}^2$ (Equation 5)

Dominance by environment interaction variance: $\sigma^2_{E.D} = 4 * \sigma^2_{E.fam}$ (Equation 6)

Residual variance: σ^2_e (Equation 7)

Individual phenotypic variance: $\sigma^2_P = \sigma^2_A + \sigma^2_{E.A} + \sigma^2_{fam} + \sigma^2_{E.fam} + \sigma^2_e$ (Equation 8)

Narrow-sense heritability: $h^2_a = \frac{\sigma^2_A}{\sigma^2_P}$ (Equation 9)

Reliability of dominance effects: $h^2_d = \frac{\sigma^2_D}{\sigma^2_P}$ (Equation 10)

Type B genetic correlation of additive effects: $r_bA = \frac{\sigma^2_A}{\sigma^2_A + \sigma^2_{E.A}}$ (Equation 11)

Type B genetic correlation of dominance effects: $r_bD = \frac{\sigma^2_D}{\sigma^2_D + \sigma^2_{E.D}}$ (Equation 12)

Next, bivariate mixed models of all 2-way combinations of traits were performed to estimate genetic correlations following model 3. The models were fit without the GxE term because including genotype-by-environment interaction terms caused singularities in the average information matrix.

$$\boldsymbol{y_{tsijk} = \mu_t + E_{ts} + T_{ti} + F_{tj} + B(E)_{tsk} + e_{tsijk}} \quad \text{(Model 3)}$$

y_{tsijk} = phenotype of trait t measured in the i^{th} tree of family j in the incomplete block k at environment s. $y_{tsijk} \sim N\,[E(y), \sigma^2_y]$.

μ_t = intercept for trait t.

E_{ts} = fixed effect of environment s for trait t.

T_{ti} = random effect of individual tree i for trait t. $T_{ti} \sim N\,[0, \sigma^2_A(T)_t]$.

F_{tj} = random effect of full-sib family j for trait t. $F_{tj} \sim N\,[0, \sigma^2_{fam}(T)_t]$.

$B(E)_{tsk}$ = random effect of block k nested within environment s for trait t. $B_k \sim N\,[0, \sigma^2_B(T)_t]$.

e_{tsijk} = residual effect of observation $y_{tsijk} \sim N\,[0, \sigma^2_e(T)_t]$;

For all random effects, i.e., individual tree effects, family effects, incomplete block effects, and the residual, a general unstructured variance-covariance structure (Gilmour *et al*. 2014) was fit for random effect within trait. The genetic correlations between traits was estimated by

$$rG = \frac{\sigma^2_{R_{12}}}{\sqrt{\sigma^2_{R_1}\sigma^2_{R_2}}} \quad \text{(Equation 13)}$$

where $\sigma^2_{R_1}$ is the variance of random genetic effect for trait one, $\sigma^2_{R_2}$ the variance of random effect for trait two, and $\sigma^2_{R_{12}}$ is the covariance between traits.

The variance structure for the individual tree effect for all models is a product of the variance component explained by the tree effect and numerator relationship matrix A derived from the pedigree. Confidence intervals of estimates presented were calculated as $\pm\, t\, SE$, with $t = 1.96$ representing the student-t' probability for a large number of degrees of freedom at 95% confidence level and SE the standard error. All statistical analyses were performed with customized scripts in the software R version 4.1.1 (R Core Team 2021) with the package ASReml-R version 4 (Butler *et al*. 2017) to fit mixed models.

4.3 Results

4.3.1 Pedigree Reconstruction and Family Structure

The estimated genotyping reproducibility was above 99% among AmpSeq genotyping rounds per DNA extraction (intra-DNA) and independent DNA extractions of the same sample (inter-DNA). The overall call rate was 94.8%, and the accuracy of genotype calls was 99.5%.

The parentage analysis identified 19,344 individuals to 136 breeding parents. Thirty-two thousand five hundred seven (32,507) parental assignments were found to be the 18 ESO parents, and 2,562 assignments were to 118 external male parents (Figure 4.1). The average number of progeny per ESO breeding parent was 1,806, with a minimum of 165 and a maximum of 3612. The average for external male parents was 22, with a minimum of one and a maximum of 240.

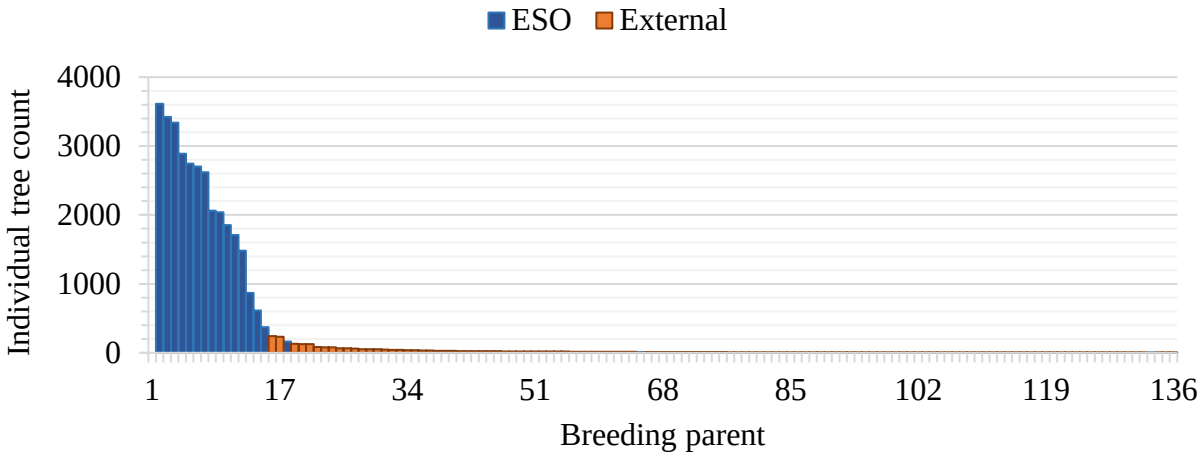


Figure 4.1. Distribution of parental assignments over 136 breeding parents. Elite seed orchard (ESO) male parents (blue) account for 93% of parental assignments. External male parents (orange) account for 7%.

With a half-diallel structure ignoring selfings and reciprocal crosses, 15,649 full-siblings were assigned to 623 full-sib families, with 84% assigned to 105 full-sib families between two of the 18 ESO parents, with an average family size of 126 with range 1-857. The remaining 16% were assigned to 518 full-sib families between an ESO female and external pollen contamination, with an average family size of 4.7 with a range of 1-80 (Figure 4.2).

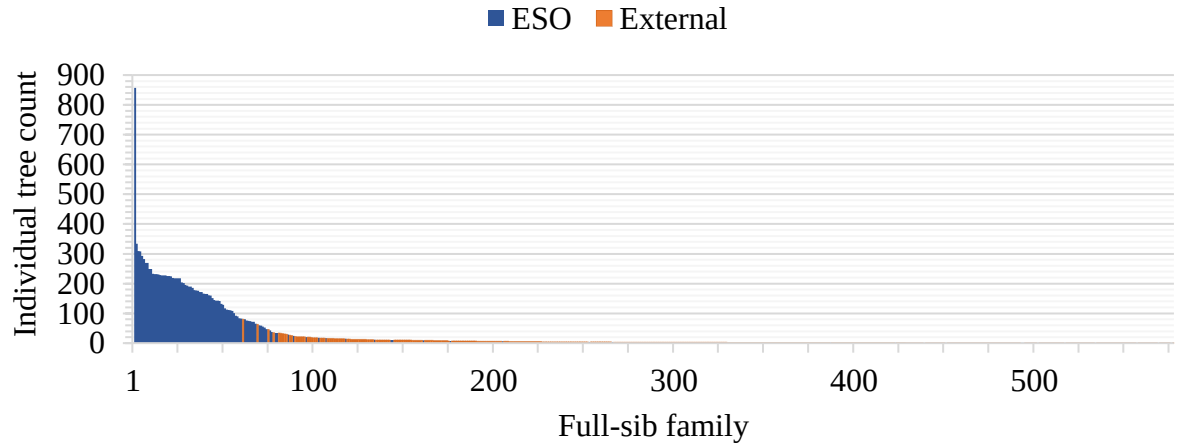


Figure 4.2. Distribution of 15,649 individuals over 623 full-sib families. Families from elite seed orchard (ESO) parents (blue) account for 84% of the individuals. Families with an external male parent (orange) account for 16%.

4.3.2 Quantitative Genetic Analysis

The graphical exploratory analysis for growth traits is presented in Figure 4.3. The complete exploratory analysis results are presented in Appendix A6.

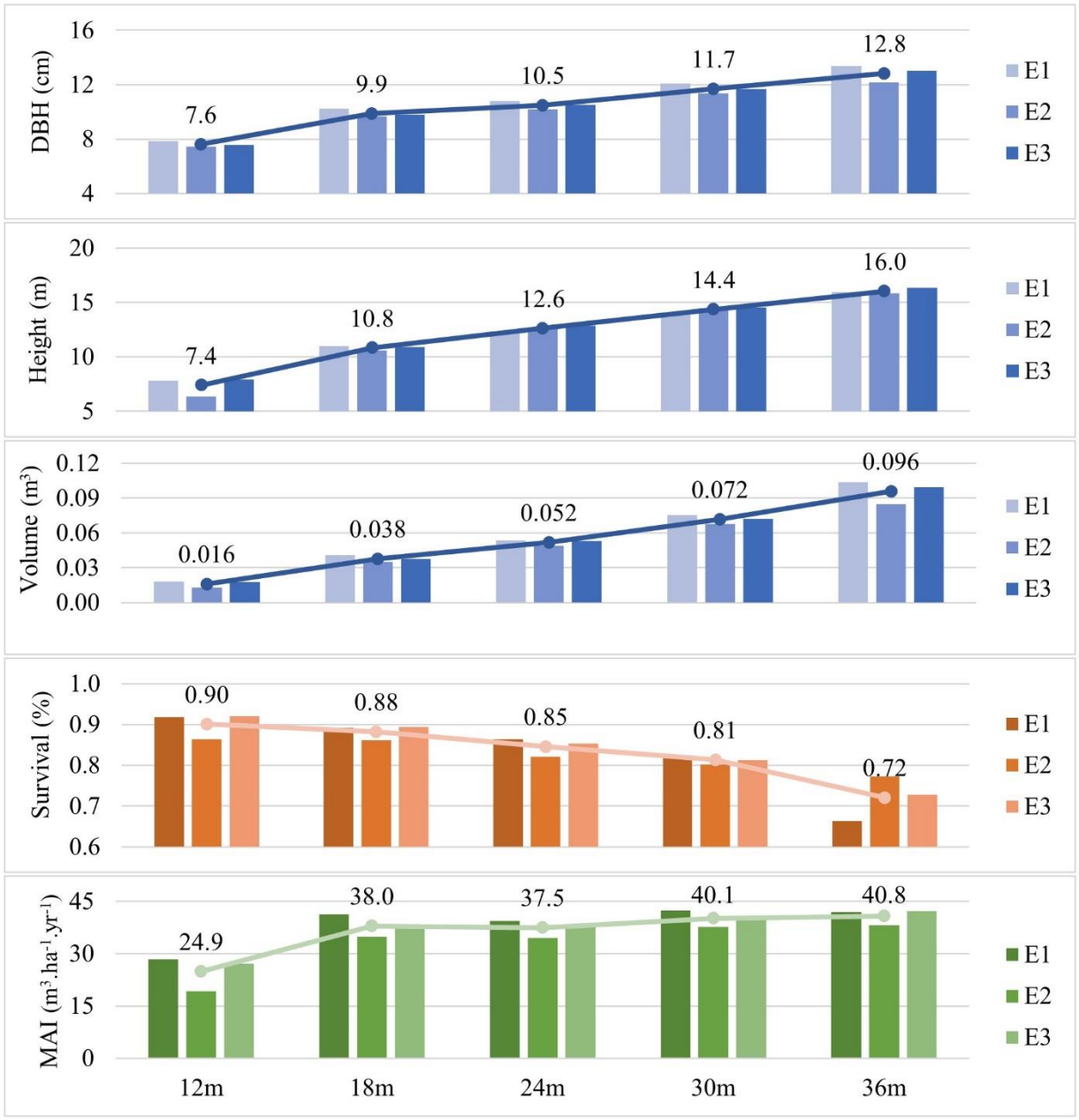


Figure 4.3. Phenotypic exploratory analysis by age and environment for DBH, HT, VOL, SUR, and MAI. The line shows the evolution of the phenotypic means over time.

The summary of the single-environment individual mixed models (model 1) fitted for all traits at 36 months is presented in Table 4.1.

Table 4.1. Variance components for random effects of Model 1 single-environment genetic analysis. The narrow-sense heritability (h^2_a) and reliability of dominance effects (h^2_d) were obtained by equations 1 and 2.

	Env	Effect	STR	SUR	HT	DBH	VOL	MAI	BD	MAI_T
Random Effects	E1	Block	0.01	0.20	0.29	0	0	15.0	115.3	3.4
	E2		0.04	0.10	0.06	0	0	3.6	23.6	0.6
	E3		0	0.07	0.09	0.10	0.02	6.8	33.3	1.2
	E1	Tree	0.05	0.34	2.18	2.16	0.85	192.9	667.9	42.4
	E2		0.37	0.53	1.11	1.61	0.37	90.7	533.4	19.4
	E3		0.04	0.40	1.75	2.43	0.79	168.4	513.8	37.8
	E1	SCA	0	0.06	0.22	0.13	0.03	22.0	2.2	4.3
	E2		0.02	0.08	0.06	0.09	0.02	6.4	19.3	1.2
	E3		0	0.07	0.02	0.10	0.01	13.1	0.0	2.6
	E1	Residual	0.22	$\pi^2/3$	8.92	9.22	2.83	505.8	657.4	110.8
	E2		0.77	$\pi^2/3$	4.36	5.74	1.47	311.7	723.8	73.6
	E3		0.25	$\pi^2/3$	5.52	7.55	2.05	386.8	791.5	88.4
	E1	$\mathbf{h^2_a}$	0.20	0.09	0.19	0.19	0.23	0.27	0.50	0.27
	E2		0.32	0.14	0.20	0.22	0.20	0.22	0.42	0.21
	E3		0.13	0.11	0.24	0.24	0.28	0.30	0.39	0.29
	E1	$\mathbf{h^2_d}$	0.02	0.06	0.08	0.05	0.04	0.12	0.01	0.11
	E2		0.08	0.08	0.05	0.05	0.04	0.06	0.06	0.05
	E3		0.02	0.07	0.01	0.04	0.02	0.09	0.00	0.08

The magnitude of incomplete block effects compared to genetic and residual effects suggests a homogeneous environment. Furthermore, there were no great discrepancies over estimates of genetic variance components and parameters, encouraging a joint analysis of the trial series. The genetic parameters of the multi-environment trial analysis obtained from model 2 are presented in Table 4.2.

Table 4.2. Genetic parameters with 95% confidence intervals of the multi-environment analysis (Model 2) variance components.

	STR	SUR	HT	DBH	VOL	MAI	BD	MAI_T
σ^2_A	**0.22** +-0.09	**0.34** +-0.12	**1.18** +-0.60	**2.05** +-0.98	**0.56** +-0.26	**103.0** +-46.1	**585.4** +-180.3	**23.3** +-10.7
σ^2_D	**0.00** +-0.02	**0.08** +-0.13	**0.16** +-0.21	**0.38** +-0.26	**0.08** +-0.06	**26.1** +-20.8	**16.21** +-24.73	**5.93** +-4.50
$\sigma^2_{E.A}$	**0.02** +-0.01	**0.04** +-0.05	**0.24** +-0.18	**0.15** +-0.13	**0.07** +-0.05	**15.3** +-12.6	**17.23** +-15.61	**3.26** +-2.70
$\sigma^2_{E.D}$	**0.04** +-0.02	**0.19** +-0.15	**0.25** +-0.25	**0.06** +-0.20	**0.02** +-0.06	**32.3** +-20.6	**13.23** +-30.04	**5.50** +-4.18
σ^2_B	**0.02** +-0.01	**0.13** +-0.04	**0.14** +-0.05	**0.02** +-0.03	**0.01** +-0.01	**7.91** +-3.12	**53.30** +-15.31	**1.63** +-0.67
σ^2_e	**0.43** +-0.05	**3.29**	**6.19** +-0.37	**7.21** +-0.57	**2.05** +-0.15	**405.8** +-27.7	**704.4** +-100.5	**92.2** +-6.44
σ^2_P	**0.67** +-0.04	**3.74** +-0.12	**7.71** +-0.32	**9.52** +-0.48	**2.71** +-0.13	**538.7** +-24.0	**1314.4** +-86.45	**121.5** +-5.54
h^2_A	**0.32** +-0.11	**0.09** +-0.03	**0.15** +-0.07	**0.22** +-0.09	**0.21** +-0.09	**0.19** +-0.08	**0.45** +-0.11	**0.19** +-0.08
h^2_D	**0.01** +-0.03	**0.02** +-0.03	**0.02** +-0.03	**0.04** +-0.03	**0.03** +-0.02	**0.05** +-0.04	**0.01** +-0.02	**0.05** +-0.04
r_bA	**0.92** +-0.07	**0.88** +-0.12	**0.83** +-0.14	**0.93** +-0.07	**0.89** +-0.09	**0.87** +-0.11	**0.97** +-0.03	**0.88** +-0.11
r_bD	**0.10** +-0.42	**0.31** +-0.43	**0.40** +-0.46	**0.86** +-0.43	**0.82** +-0.54	**0.45** +-0.29	**0.55** +-0.82	**0.52** +-0.31

Straightness, survival, tree height, and basic density were genetically controlled by additive effects. Their dominance reliabilities' confidence intervals contained zero exposing the uncertainty about significant dominance effects. On the other hand, DBH was predominantly additively controlled with smaller but significant dominance in its phenotypic expression. Consequently, VOL and MAI showed a similar genetic control with small but significant dominance effects driven by DBH, an explanatory variable of VOL and MAI calculations. The additive genetic correlations across sites were above r_bA = 0.83 for all traits. The correlation for dominance effects was above r_bD = 0.82 for DBH and VOL and slightly lower around r_bD = 0.5, for MAI and MAI_T.

The trait-trait genetic correlation matrix obtained with model 3 is presented in Figure 4.4.

Additive

Dominance								
	STR	**-0.22** ns +-.29	**0.08** ns +-.31	**-0.08** ns +-0.29	**-0.04** ns +-.029	**-0.04** ns +-.029	**-0.09** ns +-.024	**-0.03** ns +-.29
	0.22 ns +-.45	**SUR**	**-0.10** ns +-.33	**-0.17** ns +-.31	**-0.18** ns +-.31	**0.74** +-.14	**-0.10** ns +-.27	**0.70** +-.16
	-0.03 ns +-.47	**0.49** +-.41	**HT**	**0.90** +-.08	**0.95** +-.04	**0.83** +-.10	**0.57** +-.20	**0.88** +-.08
	0.25 ns +-.43	**0.25** ns +-.43	**0.64** +.25	**DBH**	**0.99** +-.02	**0.86** +-.08	**0.57** +-.18	**0.88** +-.08
	0.09 ns +-.47	**0.30** ns +-.45	**0.85** +-.14	**0.90** +-.08	**VOL**	**0.88** +-.06	**0.57** +-.18	**0.89** +-.06
	0.06 ns +-.41	**0.80** +-.27	**0.76** +-.18	**0.51** +-.25	**0.57** +-.24	**MAI**	**0.39** +-.22	**0.99** +-.00
	-0.06 ns +-.57	**0.46** ns +-.51	**0.24** ns +-.51	**0.44** +-.41	**0.29** ns +-.49	**0.15** ns +-.45	**BD**	**0.51** +-.20
	0.02 ns +-.43	**0.80** +-.27	**0.74** +-.20	**0.48** +-.27	**0.55** +-.25	**0.99** +- .00	**0.13** ns +-.47	**MAI$_T$**

Figure 4.4. Matrix of genetic correlations with 95% confidence intervals. Additive correlations are on the upper diagonal, and dominance correlations are on the lower-diagonal cells.

Additive correlations were, in general, more precise than dominance correlations. Straightness was not correlated with any other trait. Survival was only correlated with MAI. For the growth traits DBH, HT, and VOL, very high positive additive genetic correlations were verified among themselves, and they were moderately correlated with BD with a coefficient rG = 0.57. For the index MAI, high additive correlations were verified with growth traits and survival, even though they were not correlated between themselves.

4.4 Discussion

The genotyping validation results were excellent in terms of call rate and accuracy. Furthermore, the number of SNP (700) allowed for setting a threshold for parentage assignment at a 99% confidence level, providing a high-quality validated pedigree for the genetic analysis.

Most of the full-sib progeny was derived from elite seed orchard parents, the primary founders selected for the breeding population, and the target for backward selection and full-sib family deployment. Overall for full-sib families, on the female side, we ensured that all seedlings were derived from a single mother correcting for 0.5% maternal pedigree errors found. On the male side, it increased the proportion of full siblings between orchard parents to 93%, contrasting with the approximate 32% pollen contamination found in the orchard where the original open-pollination families were harvested (Chapter 2). Even for the 7% seedlings with an external male parent, it was ensured that it is a half-sib on the maternal side. This pattern was expected because the full-sib families were organized in the nursery to enrich the multi-environment trial plots with full siblings.

The rationale behind the trial design was to assess full siblings in a growing/competitive condition similar to commercial plantations, where families are deployed with individuals within the family competing with their siblings. The square plot was thought to represent this scenario by providing phenotypes that reflect the growth under an intra-family competitive environment. Competition effects can influence the phenotypic performance of trees. Even though no studies deal with this aspect with *A. crassicarpa* plantations, their influence on the phenotype was shown for a similar industrial forestry model with *Eucalyptus* (Martins *et al*. 2014, Rezende *et al*. 2019).

With parentage resolved, we could establish a trial series with a full-sib family structure never tested before for the species, with the capability to estimate additive and dominance effects

involved in the phenotypic expression of breeding parents, individual trees, and families. At three years of age, the exploratory analysis of phenotypes showed rapid and linear volumetric growth in contrast to a decrease in survival. However, MAI was not influenced by the drop in survival with age development and was stable after 18 months. This pattern suggests that as stocking decreases, live trees could compensate for the loss by increasing their volume, exhibiting resilience to sustain growth per hectare with an average of 28% stocking loss as verified at three years of age.

This apparent resilience of *A. crassicarpa* is more evident in environment 1. The stocking loss was the most prominent at this site, with an average survival of 66%. On the other hand, the tree volume was larger, resulting in an MAI on the average of the other two sites with higher stockings. At three years of age and starting with a planting density of one tree per 6 m^2, MAI is stable and hasn't decreased up to 36 months.

4.4.1 Quantitative Genetic Analysis

Single-site quantitative genetic analysis at age three showed adequate phenotypic variation levels amenable for selection for all traits. The objective of the single-site analysis was to check the data quality and evaluate the variance components before performing a multi-environment trial analysis. The results encouraged a joint analysis without larger concerns about the heterogeneity of variances. Survival showed a weak genetic signal, with a statistically significant narrow-sense heritability of $h^2_a = 0.09$. Basic density was additively controlled with a heritability of $h^2_a = 0.45$. Estimates around $h^2_a = 0.20$ were found for growth traits, and for straightness, $h^2_a = 0.32$. The MET analysis narrow-sense heritability estimates were typical of tree species planted commercially worldwide reported by Hanchor *et al*. 2016 studying *A. crassicarpa* at 12-13 years and other studies with *A. mangium* (Hai *et al.* 2015), loblolly pine (Walker *et al*. 2021) and Eucalyptus (Tan *et al*. 2017).

The reliability of dominance estimates around $h^2_d = 0.04$ and 0.05 were significant for DBH and MAI, respectively. For these traits, the ratio of dominance to the total genetic variance ($\frac{h^2_d}{h^2_a+h^2_d}$) were 0.15 and 0.20, respectively, showing their primarily additive control. Still, dominance can be important for family selection. If positive specific combining ability effects are captured into the family mean with clonal family forestry or individual clonal selection, yield could increase its value beyond the average breeding value of the parents. Similar estimates of the relative importance of dominance effects were reported by Gwaze *et* al. 2002 with linear models for *Pinus taeda* L. However, dominance effects were small compared to hybrid *Eucalyptus* (Resende *et al*. 2017, Tan *et al*. 2018, Lima *et al*. 2019), where it has a much larger importance for the phenotypic expression of growth traits. To my knowledge, no other studies have estimated dominance effects for *A. crassicarpa,* and a direct comparison was not possible.

The type B additive genetic correlation (r_bA) for all traits was very high, indicating a small effect of genotype by environment interaction among the sites studied. In this scenario, selecting generally adaptable germplasm with the best overall performance across sites is recommended.

The bivariate models estimated the genetic correlation between traits without the additive and dominance genotype-by-environment interaction terms. Singularities in the average information matrix made obtaining a solution to the mixed model equations impossible when these terms were included in the model. However, the type B genetic correlation of additive effects was high, and the importance of dominance and dominance-by-environment effects was secondary to the genetic control for all traits. Thus, the results are valid for examining the genetic correlation between traits in the population studied, revealing some interesting patterns from a breeding perspective. Straightness was not genetically correlated with other traits and had a high narrow-sense heritability $h^2_a = 0.32$. There are excellent opportunities for selection and improvement, with

straighter trees preventing wood losses during debarking and optimizing harvesting and log stacking during wood transportation. Similarly, there was a very low correlation between growth traits with survival. As expected, high correlations between growth traits and MAI were verified.

Basic density was positively correlated with growth traits, with a surprising additive genetic correlation of rG = 0.57 with HT, DBH, and VOL. This pattern is different from almost all reports for other tree species. For the hybrid *Eucalyptus grandis x E. urophylla* (urograndis), no significant genetic correlation between growth and wood density has been found (Lima *et al*. 2019). For the hard pines (e.g., *Pinus taeda, P. radiata, P.elliottii, P. sylvestris*), Zobel and Buitjen (1989) have listed 59 references about the relationship between growth rate and wood density. Of these, 35 showed no relationship, 9 showed only a small correlation, and 11 showed a negative relationship between growth and density. Zobel and Jett (1995) listed 38 more studies and separated the effect of growth rate on wood density into four categories: a) Most of the conifers with dense wood show little or no meaningful relationship; b) A negative relationship between growth and density has been reported in several genera such as spruce (*Picea* spp.) and fir (*Abies* spp.); c) Contradictory reports on Douglas Fir (*Pseudotsuga menziesii*); d) Contradictory reports about the relationship between growth rate and wood properties in the hardwoods. Usually, the diffuse-porous hardwoods (i.e., *Populus, Eucalyptus*) show little or no relationship (Saranpää 2003). Further wood technology studies are necessary to help explain the biological reasons for the positive correlation reported between density and tree volume. In the current study with *A. crassicarpa*, the positive additive genetic correlation between MAI and density contributed to a genetic correlation of rG = 0.99 between MAI and MAIton, indicating that wood production on a mass basis can be efficiently improved indirectly by selecting for higher growth and survival.

4.5 Conclusion

Parentage analysis using an informative SNP panel was used to reconstruct pedigree and allow the use of a full-sib family model to estimate additive and dominance effects and genetic correlations across sites and among important traits. The genetic control of all traits assessed in this study was primarily additive. In this scenario, the simple recurrent selection is recommended as a breeding strategy targeting genetic gains at individual and family levels, with recombination of forward selections for generation advancing and backward selection of parental combinations of the best full-sib families to produce seed for deployment via family forestry. Small but significant dominance effects for DBH and MAI can be explored, looking forward to the positive specific combining ability effects and heterosis that can increase the genetic value of families and, ultimately, the volume of wood produced.

The additive Type B genetic correlation across sites was very high for all traits. Consequently, the genotype-by-environment interaction will likely have a minor influence on the genotypic performance. A single breeding population can be developed with the potential to provide genotypes and families with broad adaptation and stability over the environmental range represented by the sites studied.

The pattern of genetic correlations among traits simultaneously favors the genetic progress for all traits. Straightness and survival were independent of growth traits, and tree volume was correlated with the basic density and mean annual increment. Thus, trees with the highest possible volume and straightness scores should be the target for individual tree selection, advancing the breeding population and promoting genetic gains simultaneously for all traits.

4.6 References

Butler DG, Cullis BR, Gilmour AR, Gogel BG, Thompson R (2017) ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Gilmour A, Gogel B, Cullis B, Welham S, Thompson R (2014) ASReml user guide. Release 4.1 structural specification. VSN International Ltd, Hemel Hempstead, UK.

Griffin AR, Vuong TD, Harbard JL, Wong CY, Brooker C, Vaillancourt RE (2010) Improving controlled pollination methodology for breeding *Acacia mangium* Willd. New Forests, 2010, 40:2, 131–142. https://doi.org/10.1007/s11056-010-9188-x

Gwaze DP, Bridgwater FE, Williams CG (2002) Genetic analysis of growth curves for a woody perennial species, *Pinus taeda* L. Theoretical and Applied Genetics, 105, 526–531. https://doi.org/10.1007/s00122-002-0892-6

Hai PH, La AD, Toan NQ, Trieu TTH (2015) Genetic variation in growth, stem straightness, pilodyn and dynamic modulus of elasticity in second-generation progeny tests of *Acacia mangium* at three sites in Vietnam. New Forests, 46, 577–591. https://doi.org/10.1007/s11056-015-9484-6

Hanchor U, Maelim S, Suanpaga W, Park JM, Kang KS (2016) Growth performance and heritability estimation of *Acacia crassicarpa* in a progeny trial in eastern Thailand. Silvae Genetica 65:2, 58–64. https://doi.org/10.1515/sg-2016-0017

Harwood CE, Nambiar EKS (2014) Sustainable plantation forestry in South East Asia. ACIAR Technical reports No. 84. Australian Centre for International Agricultural Research: Canberra. 100 pp.

Harwood CE, Hardiyanto EB, Wong CY (2015) Genetic improvement of tropical acacias: achievements and challenges. Southern Forests, 77:1, 11-18. https://doi.org/10.2989/20702620.2014.999302

Isik F, Holland J, Maltecca C (2017) Genetic data analysis for plant and animal breeding. Cham, Switzerland: Springer International Publishing.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. Molecular Ecology, 12, 2511–2523. https://doi.org/10.1046/j.1365-294X.2003.01928.x

Jones AG, Clayton MS, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. Molecular Ecology Resources, 10, 6–30. https://doi.org/10.1111/j.1755-0998.2009.02778.x

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology, 16, 1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x

Lima BM, Cappa EP, Silva-Junior OB, Garcia C, Mansfield SD, 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, e0218747. https://doi.org/10.1371/journal.pone.0218747

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology, 7, 639-655. https://doi.org/10.1046/j.1365-294x.1998.00374.x

Martins GS, Moura GPL, Ramalho MAP, Gonçalves FMA (2014) Performance of Eucalyptus Clones in Auto and Allocompetition. Silvae Genetica, 63:1, 9-14. https://doi.org/10.1515/sg-2014-0002

Martins GS, Yuliarto M, Antes R, Sabki, Prasetyo A, Unda F, Mansfield SD *et al*. (2020) Wood and Pulping Properties Variation of *Acacia crassicarpa* A.Cunn. ex Benth. and Sampling Strategies for Accurate Phenotyping. Forests, 11, 1043. https://doi.org/10.3390/f11101043

McKinnon GE, Larcombe MJ, Griffin AR, Vaillancourt RE (2018) Development of microsatellites using next-generation sequencing for *Acacia crassicarpa*. Journal of Tropical Forest Science, 30:2, 252–258. https://doi.org/10.26525/jtfs2018.30.2.252258

Mendham DS, White DA (2019) A review of nutrient, water and organic matter dynamics of tropical acacias on mineral soils for improved management in Southeast Asia. Australian Forestry, 82:sup1, 45-56. https://doi.org/10.1080/00049158.2019.1611991

Midgley SJ, Turnbull JW (2003) Domestication and use of Australian acacias: case study of five important species. Australian Systematic Botany, 16:1, 89-102. https://doi.org/10.1071/SB01038

Nambiar EKS, Harwood CE (2014) Productivity of acacia and eucalypt plantations in Southeast Asia. 1. Bio-physical determinants of production: opportunities and challenges. International Forestry Review, 16:2, 225–248. https://doi.org/10.1505/146554814811724757

Nambiar EKS, Harwood CE, Mendham DS (2018) Paths to sustainable wood supply to the pulp and paper industry in Indonesia after diseases have forced a change of species from acacia to eucalypts. Australian Forestry, 81:3, 148-161. https://doi.org/10.1080/00049158.2018.1482798

Nirsatmanto A, Sunarti S (2019) Genetics and Breeding of Tropical Acacias for Forest Products: *Acacia mangium*, *A. auriculiformis* and *A. crassicarpa*. In Advances in Plant Breeding Strategies: Industrial and Food Crops; Al-Khayri, J., Jain, S., Eds.; Springer: Cham, Switzerland, 2019; Volume 6, pp. 3–28.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, Silva-Junior OB, Grattapaglia D (2017) Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity, 119, 245–255. https://doi.org/10.1038/hdy.2017.37

Rezende GDSP, Resende, MDV, Assis TF (2014) *Eucalyptus* Breeding for Clonal Forestry. In Fenning Challenges and Opportunities for the World's Forests in the 21st Century; Fenning, T., Ed.; Springer: Dordrecht, The Netherlands, 2014; Forestry Sciences; Volume 81, 393–424.

Rezende GDSP, Lima JL, da Costa Dias D, de Lima BM, Aguiar AM *et al*. (2019) Clonal composites: An alternative to improve the sustainability of production in eucalypt forests. Forest Ecology and Management, 449, 117445. https://doi.org/10.1016/j.foreco.2019.06.042

Rodriguez G, Elo I (2003) Intra-class correlation in random-effects models for binary data. The Stata Journal, 3(1), 32-46. https://doi.org/10.1177/1536867X0300300102

Saranpää P (2003) Wood Density and Growth. In Wood Quality and Its Biological Basis; Barnett, J.R., Jeronimidis, G., Eds.; Blackwell Publishing Ltd.: Oxford, UK, pp. 87–117.

Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK (2017) Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus*

species and their F1 hybrids. BMC Plant Biology, 17, 110. https://doi.org/10.1186/s12870-017-1059-6

Tan B, Grattapaglia D, Wu HX, Ingvarsson PK. (2018) Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. Plant Science, 267, 84-93. https://doi.org/10.1016/j.plantsci.2017.11.011

Turnbull JW, Midgley SJ, Cossalter C (1998) Tropical acacias planted in Asia: an overview. In: Recent developments in acacia planting. Proceedings of an international workshop held in Hanoi, Vietnam, 27-30 October 1997; Turnbull, J.W., Crompton, H.R., Eds.; ACIAR (Australian Centre for International Agriculture Research) proceedings no. 82., Canberra, Australia, pp. 155-160.

Walker TD, Cumbie WP, Isik F (2021) Single-step genomic analysis increases the accuracy of within-family selection in a clonally replicated population of *Pinus taeda* L. Forest Science, 68:1, 37–52. https://doi.org/10.1093/forsci/fxab054

Whitaker D, Williams E R, John J A (2001) CycDesigN: A Package for the Computer Generation of Experimental Designs. CSIRO Forestry and Forest Products, Canberra.

White TL, Adams WT, Neale DB. (2007). Forest genetics. Wallingford: CAB International.

Wong CY, Yuliarto M (2014) Deployment of acacias in short rotation pulpwood plantation. In: Acacia 2014 “Sustaining the Future of Acacia Plantation Forestry” International Conference, IUFRO Working Party 2.08.07: Genetics and Silviculture of Acacias, Hue, Vietnam, 18–21 March 2014, Compendium of Abstracts.

Zobel BJ, Van Buijtenen JP (1989) Wood variation, its causes and control. Springer-Verlag, Berlin, 363 p.

Zobel BJ (1993) Clonal Forestry in the Eucalypts. In: Ahuja, MR., Libby, W.J. (eds) Clonal Forestry II. Springer, Berlin, Heidelberg.

Zobel BJ, Jett JB (1995) The Genetics of Wood Density. In: Genetics of Wood Production. Springer Series in Wood Science. Springer, Berlin, Heidelberg.

CHAPTER 5

Applied genomics to overcome biological constraints and elucidate quantitative genetic architecture: A case study in *Acacia crassicarpa* for accelerating gains in wood production

5.1 Introduction

Tree breeding programs are usually based on recurrent selection methods and individual tree selection among and within families for recombination and generation advancement (Zobel and Talbert 1984). In principle, at any stage of a breeding program, clonal forestry should deliver the greatest genetic gain into commercial production through the capture of both additive and non-additive gain, potentially including hybrid vigor (Isik *et al.* 2017, Holland *et* al. 2003, Wu 2018). Clonal forestry requires the ability to mass-produce juvenile propagules of selected trees, as is commonly implemented for hybrid *Eucalyptus*, the world-class benchmark for forest productivity (Zobel 1993, Rezende *et al*. 2014). Hybrid eucalypt breeding programs often adopt recurrent selection with multispecies synthetic populations as a breeding strategy (Assis and Resende, 2011), as this approach arguably provides cost-effective production of new elite clones and higher speed of generation advancement (Kerr *et al*. 2004). The strategy is based on a relatively small effective population size of elite parents crossed in diallel mating designs. Furthermore, it has been shown by deterministic simulations (Grattapaglia and Resende, 2011) and empirical reports (Resende *et al*., 2012, Resende *et al*., 2017) that there is immediate potential for the application of operational genomic selection (GS) in the framework of such small, fast-moving specialized breeding populations.

Some tree species, however, present challenges due to reproductive biology and amenability of vegetative propagation that can effectively limit cloning capacity and/or the

efficiency of controlled pollination. Although technically feasible, controlled pollination of acacias is difficult, and the percentage of flowers that develop into pods is typically less than 5%. In the breeding program of April Asia, even with recommended methods such as "inflorescence pollination" (Griffin *et al*. 2010), a huge effort is required to produce more than a few crosses per year. Therefore, controlled pollination is not practical for advancing breeding populations which require hundreds of crosses for each breeding cycle. For acacia, breeding populations have typically been developed using open pollination (OP) (Harwood *et al*. 2015). Vegetative propagation is cost-effective but is possible only with juvenile ortets, leading to the employment of family forestry in Indonesia (Griffin 2014, Wong and Yuliarto 2014). Due to the inefficiency of controlled pollinations, OP seed is collected from selected parent trees based on progeny performance. The OP seed is sown and vegetatively propagated to bulk-up nursery hedges supporting the deployment of rooted cuttings. The main limitation of this strategy is that selection is based only on female additive effects. Developing methods to produce and test full-sib families could deliver larger genetic gains and uniformity captured and deployed by full-sib family forestry. The advantages of full-sibling genetic models over open-pollinated models are significant (White *et al*. 2007).

The past decades have seen considerable progress in forest tree genomics research, and cost-effective genotyping platforms of single nucleotide polymorphism markers (SNP) are now available for many mainstream plantation forest trees (Grattapaglia 2022). A variety of tools and methods have been successfully employed in applied forest tree improvement programs (Whetten *et al*. 2023), for example, marker-assisted selection for resistance to fusiform rust disease in loblolly pine (Cumbie *et al*. 2020, Lauer and Isik 2021), genetic quality control of planting stock (White *et al*. 2014), parentage reconstruction (El-Kassaby and Lstibůrek 2009, Klápště *et al.* 2017)

and genomic prediction of genetic values (Lima *et al*. 2019, Ratcliffe *et al*. 2017, Durán *et al*. 2017, Tan *et al*. 2017). A valuable use of parentage reconstruction is to allow full pedigree determination of open-pollinated seed lots. It may be the only way full parental control can be applied to estimate genetic parameters for species where controlled crossing is inefficient or impossible. Additionally, mid-density panels utilized for parentage reconstruction can serve as an imputation panel allowing large-scale utilization of high-density genotyping with significant cost reduction (Whalen *et al*. 2020). Several published forest tree studies show that utilizing genomic information to inform selection decisions in a breeding program matches or surpasses the performance of phenotypic selection for growth and wood properties traits (Grattapaglia 2022). Genomic selection can increase genetic gain per unit of time, for example, by allowing a very early selection of non-phenotyped seedlings in the nursery while simultaneously increasing selection intensity (*i*) and the accuracy of breeding values compared to pedigree models (Muñoz *et al*. 2014).

Furthermore, realized relationship matrices constructed from genome-wide SNP markers can accurately measure relatedness (Hayes *et al*. 2009), capturing the deviation from the expected value due to the Mendelian sampling of alleles during sexual recombination. Within-family genomic selection is feasible, allowing the breeder to explore the genetic variance present within a full-sib family (½ of additive variance + ¾ of dominance variance) (Falconer and MacKay 1996) and deliver additional genetic gains beyond family mean selection (Lynch and Walsh 1998). This strategy would be useful with tree species with difficult vegetative propagation, as demonstrated for loblolly pine, with accurate within-family genomic selection (Walker *et al*. 2021). In addition, upon the breakthrough development of techniques for mass-scale production of juvenile propagules, superior individual genotypes can be sourced with genomic selection, maximizing the genetic gain obtained per breeding generation with clonal forestry.

The success of a genomic selection program is determined by several factors that can be divided into two main groups (Grattapaglia 2017). The first group comprises the four key factors from population and quantitative genetics theory that affect the accuracy of genomic prediction: (1) the effective size of the population (N_e); (2) the composition and size of the training population; (3) the trait heritability; and (4) the genetic architecture of the target trait. The influence of these factors on the success of genomic prediction was extensively examined, initially in simulation studies (Grattapaglia and Resende 2011) and later empirically, as reviewed by Isik 2022 and Lebedev *et al.* 2020. The level of relatedness between the training population and the candidates for selection is the main factor impacting its success. The closer the genetic relationships, the higher the predictive ability. The second group of factors deals with aspects that define costs and resource allocation in the breeding program, and among these are (1) the quality of the phenotyping; (2) multi-environment testing for estimating GxE interaction; and (3) the genotyping system, especially regarding data quality and cost (Grattapaglia 2022).

This study evaluated how genomic solutions can improve selection and increase the genetic gains in an *A. crassicarpa* breeding program. The genotyped population with 9,973 offspring from 28 breeding parents consists of large families with hundreds of progeny, which makes inferences applicable to different selection targets. The study objectives included (1) comparing models with and without genomic relationships to evaluate their impact on fit and precision of breeding value predictions, (2) quantifying the predictive ability of genomic selection for individual trees, breeding parents, families, and progeny within-family with varying levels of relatedness between the training and validation populations, and (3) determining the response to within-family genomic selection and quantify potential additional gains that can be captured and deployed via family forestry.

5.2 Material and Methods

5.2.1 Germplasm and Phenotyping

The APRIL Asia breeding program collected twenty-four open-pollinated families from elite clonal seed orchards. Eighty-four thousand seedlings from the open-pollinated families were initially sampled at the nursery for paternal determination with a low-density panel of 42 SNP markers. The paternity was determined by exclusion, calculating the proportion of matching alleles between the male candidate and offspring in each open-pollinated family derived from a known female genotype. There were 18 genotypes in the orchard, all considered candidate male parents. The parent-offspring combination with the highest average concordance across SNP above 0.8 was assigned. Nineteen thousand three hundred sixty siblings were organized into full-sib families in the nursery with varying numbers of individuals per family.

The full-sib families with enough seedlings passing the nursery quality control were included in genetic trials established on three representative Histosol environments (E) in Riau province, Sumatra, Indonesia, being E1) 80 families in Pelalawan location with hemic peat soil; E2) 93 families in Merenti location with sapric peat soil; E3) 69 families in Pelalawan with sapric peat soil. The trial design was a resolvable alpha-lattice with five replicates and 8, 11, and 9 incomplete blocks within replicate on E1, E2, and E3, respectively. Plots were 16 tree blocks composed of four rows and four trees per row, with a spacing of 2 m between trees within a row and 3 m between rows. Families were assigned to incomplete blocks following the alpha-lattice design generated and optimized by the software CycDesign (Whitaker *et al*. 2001).

At 36 months of age, phenotypes were measured: diameter at breast height (DBH), in centimeters; tree height (HT), in meters; and straightness (STR), with a scoring system comprising six levels with STR=1 representing a very crooked tree and STR=6 representing a perfectly straight

tree. These traits were independently measured twice to improve the phenotyping quality, and the final phenotypic value was the average of the two assessments. At the same age, resistograph measurements with an IML-RESI PD400 device were performed to predict wood basic density (BD), expressed as the dry mass in kilograms over wet volume in cubic meters (kg m^{-3}). Individual under-bark tree volume (VOL), expressed in dm^3, was calculated with the DBH and HT measurements. Both wood basic density and individual tree volume predictions were obtained with statistical models developed and validated for *A. crassicarpa* and routinely used by the company.

5.2.2 SNP genotyping

A custom SNP microarray was designed by Orion Genomics (St. Louis, MO, USA) for the proprietary *A. crassicarpa* genome utilizing the Thermo Fisher (Affymetrix) Axiom myDesign GW platform. A set of 610,000 markers was selected for high performance on the Axiom platform. SNPs were prioritized based on the uniqueness of the surrounding sequence, even physical distribution across the genome, and minor allele frequency (MAF).

All 16,899 live trees in the three field trials at age 12 months were sampled for DNA extraction and SNP genotyping with a panel of 700 SNP markers. The set of markers is a subset of the HD panel. The mid-density panel was developed by Orion Genomics using a multiplexed PCR amplicon/sequencing-based approach.

The low-density panel with 42 SNP markers was developed to provide a high-throughput and cost-effective genotyping platform capable of accurately assigning *A. crassicarpa* genotypes. SNP markers were selected based on minor allele frequency (MAF) and broad distribution across the genome. Orion Biosains (Puchong, Malaysia) designed, optimized, and performed assays utilizing the LGC Array Tape genotyping platform.

5.2.3 SNP imputation

The parentage of trial-established trees was validated with mid-density genotyping. Seedlings were sampled from known mothers and assigned to candidate fathers using the paternity analysis of the software CERVUS version 3.0.7 (Marshall *et al*. 1998; Kalinowski *et al*. 2007). The company has previously genotyped all its clonal breeding orchards, and a collection of 800 genotypes determined from this campaign were considered candidate fathers. The simulation of paternity analysis, utilized by the software to estimate the resolving power of the markers given their allele frequencies and to estimate critical values of the log-likelihood statistics, was performed with the following parameters: 10,000 individuals; 800 candidate parents; 0.95 proportion of sampled parents; 0.01 proportion of mistyped loci; and the option to test for self-fertilization. All other parameters of the paternity analysis followed the software default. Offspring assigned to a candidate father, given the known mother, with a trio confidence level equal to or above 99% were considered for downstream analysis.

With the pedigree update, breeding parents were ranked by their number of offspring, and 28 parents along with 328 progeny were selected for HD genotyping with the axiom array. The parents were genotyped in duplicate and had at least seven progeny with HD data. An analysis pipeline was developed by Orion Genomics to systematically test parental phasing and imputation of 267,000 high-performing SNP markers for a selection of 9,973 individuals derived from the 28 breeding parents and previously genotyped with the mid-density panel. The pipeline consists of a set of interlinked R (R Core Team 2021) scripts for pre-processing and conversion of data and parsing these to the software packages AlphaSimR (Gaynor *et al*. 2020), AlphaPhase (Hickey *et al*. 2011), AlphaImpute 2 (Whalen and Hickey 2020) and AlphaFamImpute (Whalen *et al*. 2020).

After these programs completed the main analysis, additional pipeline scripts evaluated and summarized results.

5.2.4 Construction of the genomic relationship matrix

Initially, out of the 267,000 imputed SNP, redundant markers with a correlation r = 1 were removed, resulting in 152,892 non-redundant SNP markers. A realized genomic relationship matrix (G) was generated with the method proposed by VanRaden (2008) that is scaled to the numerator relationship matrix A, according to equation 1.

$$G = \frac{(M-P)(M-P)'}{2\sum p_i(1-p_i)} \quad \text{(Equation 1)}$$

M is a matrix of individuals by SNP markers, and P is a matrix that contains twice the MAF for each locus. Dimensions of M and P are 10,001 by 152,892. The G matrix was aligned based on information on the expected relationships in the numerator relationship matrix (A) obtained from the validated pedigree. The G matrix generation and alignment were performed in R with the package ASRgenomics (Gezan *et al.* 2022).

5.2.5 Linear Mixed Models

For the genomic analysis, phenotypes were adjusted for environment and blocks with solutions obtained with Model 1.

$$\boldsymbol{y_{sij} = \mu + E_s + T_i + B(E)_j + e_{sij}} \quad \text{(Model 1)}$$

where y_{sij} = phenotype measured in the i^{th} tree in the incomplete block j at environment s.
$y_{sij} \sim N\,[E(y), \sigma_y^2]$;
μ = intercept;
E_s = fixed effect of environment s.
T_i = random effect of individual tree i. $T_i \sim N\,(0, \sigma_{tree}^2)$.
$B(E)_j$ = random effect of incomplete block j nested in environment s. $B_j \sim N\,(0, \sigma_B^2)$.
e_{sij} = residual effect of observation y_{sij}. $e_{sij} \sim N\,(0, \sigma_e^2)$.

The adjusted phenotypes ($\boldsymbol{y_i^*}$), corrected for fixed effect of environment and random effect of block were obtained following $\boldsymbol{y_i^* = \mu + \widehat{T}_i + \hat{e}_{sij}}$. With that, linear mixed models with different variance-covariance structures for individual trees were fitted. Model 2 (A-BLUP) had a numerator relationship matrix A, constructed with the SNP-validated pedigree to model the genetic variance and covariance among individual trees.

$$\boldsymbol{y_i^* = \mu + T_i + e_i} \quad \text{(Model 2)}$$

where y_i^* = adjusted phenotype measured in the i^{th} tree. $y_i \sim N\,[E(y), \sigma_y^2]$;
μ = intercept;
T_i = random effect of individual tree i. $T_i \sim N\,(0, \sigma_{tree}^2)$.
e_i = residual effect of observation y_i^*. $e_i \sim N\,(0, \sigma_e^2)$.

Model 3 (G-BLUP model) is an alternative to model 2, with the same specification but employing the realized genomic relationship matrix G to model individual covariances and the genetic variance.

Genetic parameters were calculated from variance components estimated by restricted maximum likelihood (REML), and BLUP of genetic entries were obtained with R utilizing the package "ASReml-R 4" (Butler *et al.* 2017). The narrow-sense heritability (h_a^2) of model 2 and the genomic heritability (h_G^2) of model 3 were calculated following $h^2 = \frac{\sigma_{tree}^2}{\sigma_{tree}^2 + \sigma_e^2}$. BLUP's accuracy ($r$), which estimates the correlation between true and predicted genetic value, has been recommended as an appropriate measure to compare A-BLUP and G-BLUP models (Putz *et al*. 2018). It was calculated using the prediction error variance (PEV) and the genetic variance component from the model (Gilmour *et al*. 2014) according to equation 2.

$$r = \sqrt{1 - \frac{PEV}{(1+F)\sigma_{tree}^2}} = \sqrt{1 - \frac{PEV}{\sigma_{tree}^2}} \quad \text{(Equation 2)}$$

where F is the coefficient of inbreeding, assumed to be 0 for all individuals in this study.

5.2.6 Genomic Prediction Validation

A leave-one-environment-out validation scheme was performed with varying model training and validation set composition. There were three trials in this study. So, there were three independent combinations of trials to train a model with two of them and validate with the third independent trial. In each scenario, phenotypes of one environment were masked, and genomic estimated breeding values (GEBV) were predicted with a G-BLUP model that still contained all individuals in the G matrix.

The predictive ability (PA) was calculated as the Pearson correlation between the GEBV of the masked individuals in each scenario and their corresponding BLUP obtained with Model 3, fitted with all phenotypes. Predictive abilities were obtained for three different selection targets in each validation scenario: (1) for individual tree selection, calculating a single PA_{ind} by the correlation between the vector of GEBV and model 3 BLUP for all individuals; (2) for family mean selection, calculating a single PA_{fam} by the correlation between the vector of family mean predictions $\left(\frac{BV_{P1}+BV_{P2}}{2}\right)$ obtained with parental GEBV and BLUP from model 3; (3) for individual selection within family, calculating a $PA_{w.fam}$ for each family in the testing set of each scenario by the correlation of the vectors of GEBV and model 3 BLUP for the individuals within each family separately.

5.2.7 Within-family genomic selection performance

The predicted response to within-family genomic selection (RGS_w) was evaluated for the top five ranked families with at least 40 individuals for the trait volume (VOL) in each validation scenario, according to equation 3.

$$RGS_w(\%) = \left(\frac{\overline{GEBV}_{10\%} - \overline{GEBV}}{\overline{GEBV}}\right) x\ 0.833\ x\ 100 \qquad \text{(Equation 3)}$$

where $\overline{GEBV}_{10\%}$ is the average GEBV of the top 10% of progeny selected within a family, and $\overline{GEBV}$ is the average GEBV of the whole family. RGS_w is equivalent to response to selection as defined in Bernardo (2020) because the GEBV predicts the genetic gain due to the shrinkage properties of the REML/BLUP methodology.

Within-family genomic selection can be practiced in *A. crassicarpa* at the seedling stage with leaf sampling followed by DNA extraction, SNP genotyping, GEBV predictions, and bulk-up of selected individuals to achieve nursery requirements for deployment at scale. This process could take a year longer than setting nursery hedges directly from seedlings without using genomic selection. To account for this extra time, we adjusted RGS_w by a factor of 0.833, which corresponds to the ratio between a breeding cycle of 5 years with the additional year, i.e., 5/6 = 0.833.

5.3 Results

5.3.1 Population structure

With imputed high-density genotyping and parentage reconstruction, the family structure in the population was updated. Including a larger set of known candidate male parents resulted in the assignment of small full-sib families of female orchard trees with pollen contaminants. The overall number of families increased. The validated family connectivity across the three test environments is presented in Figure 5.1. The diallel crossing matrix is presented in Appendix A7. There were 16 orchard parents in the population. In addition, there were 12 external genotypes that contributed as pollen contaminants had sufficient offspring that allowed parental phasing and imputation and were included in the genomic studies.

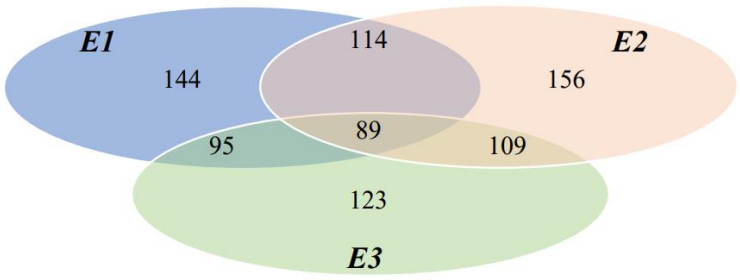


Figure 5.1. Venn diagram showing the family connectivity between environments (E).

5.3.2 Genomic data

The graphical representation of the realized relationship matrix G is presented in Figure 5.2. The diallel with 9,973 progeny trees derived from 28 breeding parents comprises large full-sib and half-sib blocks, with several breeding parents with over a thousand individual progeny.

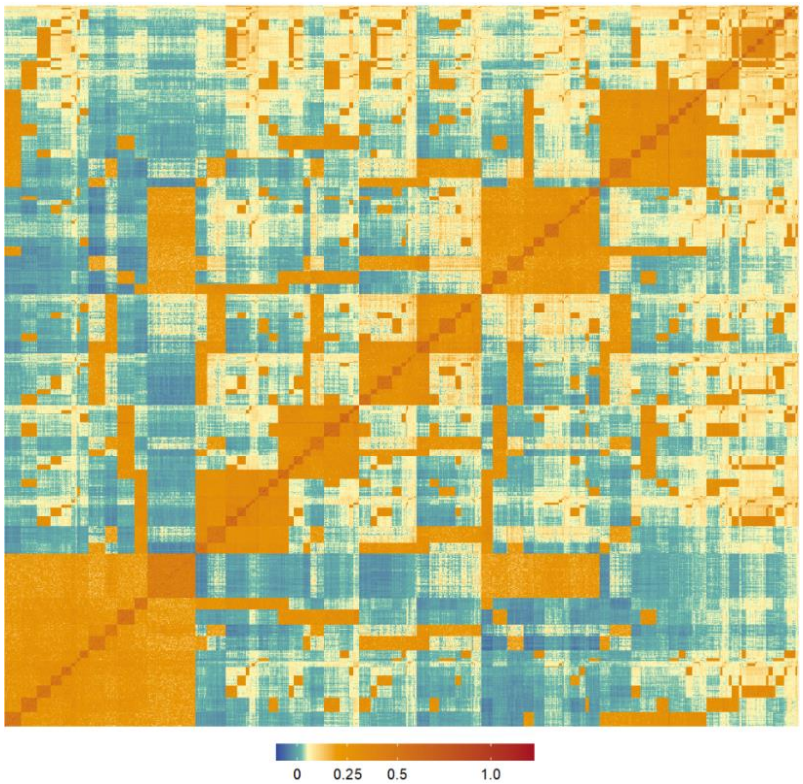


Figure 5.2. G-matrix heatmap. Off-diagonal values show full-sib blocks in dark orange along the diagonal and larger half-sib blocks in lighter orange spread across the matrix.

As expected, realized relationship coefficients exposed the deviation from expected values between unrelated (0), half-sibs (0.25), full-sibs (0.5), and selfings (1.0) due to mendelian sampling of alleles. The 10,001 diagonal values were centered at 1.0 with a range of 0.88-1.23, with low levels of inbreeding corroborating an outcrossed segregation pattern in the population (Figure 5.3).

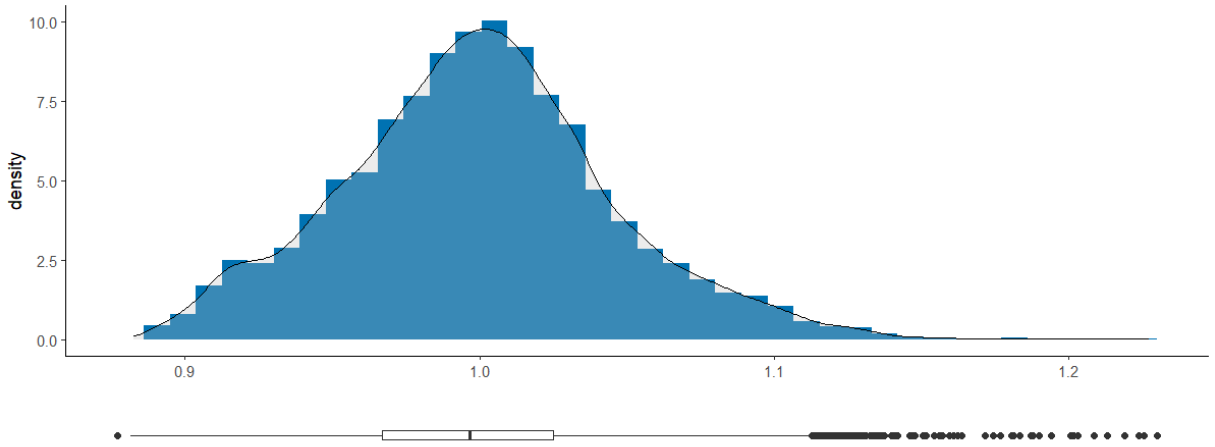


Figure 5.3. Distribution of the diagonal coefficients of the G matrix.

The distribution of the 50,005,000 off-diagonal elements shows the relationships among individuals with peaks in the expected values of unrelated, half-sibs, and full-sibs (Figure 5.4).

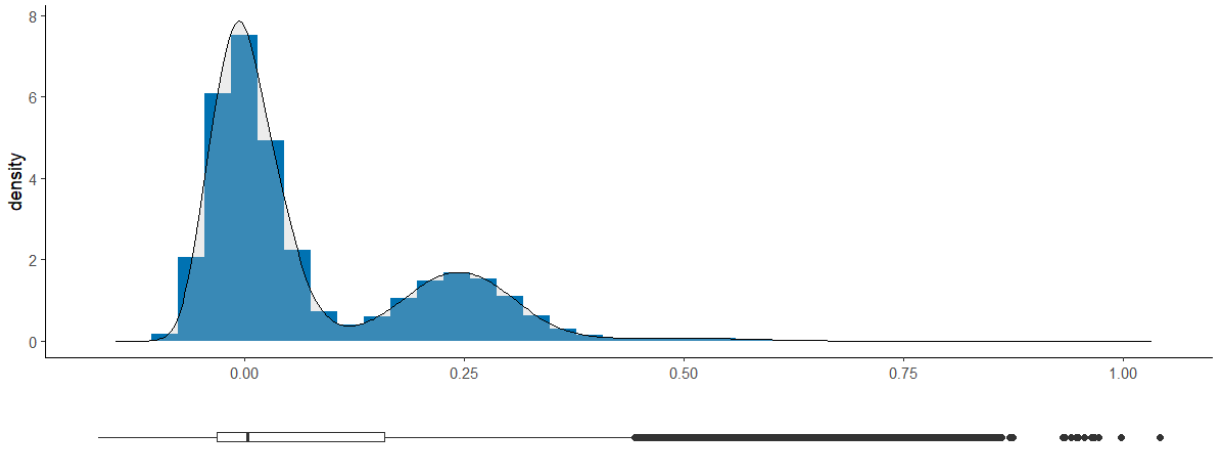


Figure 5.4. Distribution of the off-diagonal coefficients of the G matrix.

5.3.3 Genomic model performance

The performances of the mixed models are presented in table 5.1. There was considerable phenotypic variation for tree volume, with a coefficient of variation (CV) of 0.49. Smaller phenotypic variations were seen for basic density and straightness, with a CV of 0.08 and 0.25, respectively. The G-BLUP model outperformed the A-BLUP model for all traits, with heritability estimates with smaller standard deviations and significantly higher prediction accuracy (r) of individual trees across traits.

Table 5.1. Models' performance for VOL, BD, and STR. Estimated variances (σ^2) and heritability (h^2) are presented with their standard deviation. Accuracy (r) of predictions is presented with its mean and range.

Attribute	***VOL***	***BD***	***STR***
Phenotype (y*)			
Mean	105.6	421.6	3.11
Standard deviation	51.6	35.2	0.76
Coefficient of variation	0.49	0.08	0.25
A-BLUP			
σ^2_{tree}	585.6 ± 180.6	613.6 ± 178.7	0.155 ± 0.049
σ^2_e	2056. ± 96.41	681.2 ± 90.53	0.471 ± 0.026
h^2_a	0.222 ± 0.061	0.474 ± 0.105	0.248 ± 0.068
Parental r	0.92 (0.77-0.98)	0.95 (0.88-0.98)	0.92 (0.78-0.98)
Individual r	0.72 (0.68-0.72)	0.78 (0.65-0.79)	0.72 (0.62-0.78)
G-BLUP			
σ^2_{tree}	763.9 ± 66.41	334.0 ± 28.47	0.102 ± 0.011
σ^2_e	1996. ± 29.94	843.2 ± 12.77	0.502 ± 0.007
h^2_G	0.277 ± 0.018	0.284 ± 0.018	0.169 ± 0.016
Parental r	0.90 (0.78-0.95)	0.91 (0.78-0.95)	0.87 (0.68-0.95)
Individual r	0.88 (0.77-0.92)	0.89 (0.78-0.93)	0.85 (0.71-0.91)

Not surprisingly, the most successful reproductive breeding parents are orchard parents, not pollen contaminants. Orchard parents had many individuals in the progeny trial driving their prediction accuracy with the G-BLUP model above 0.9 (Figure 5.5).

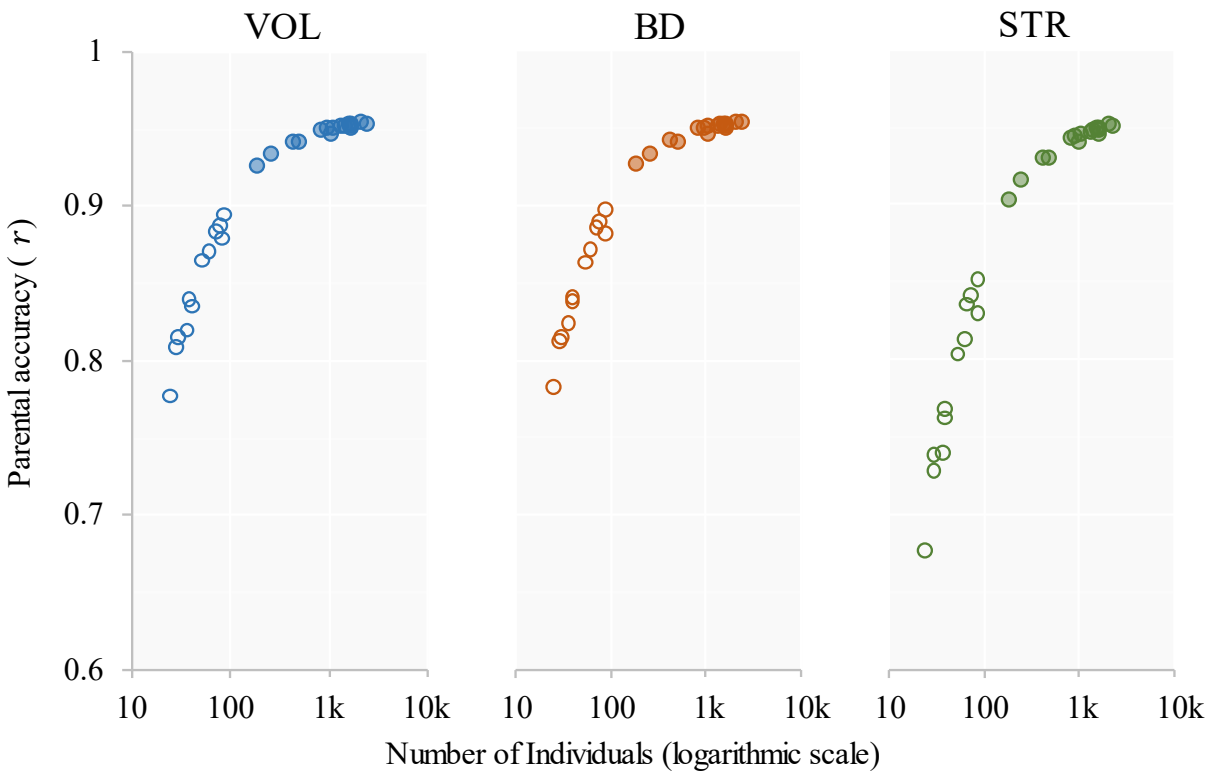


Figure 5.5. BLUP accuracies of the 28 breeding parents for individual tree volume (VOL), wood basic density (BD), and straightness (STR). Orchard parents are presented in full circles, and pollen-contaminant parents in empty circles.

Regarding individual trees, the average accuracy of additive values obtained with the A-BLUP model was around 0.7. With the G-BLUP model, higher accuracies were found, with an increase of almost 0.2, resulting in an average of around 0.9. The better modeling of genetic relationships among individuals improved individual accuracies to a level as precise as found for parental genotypes (Figure 5.6).

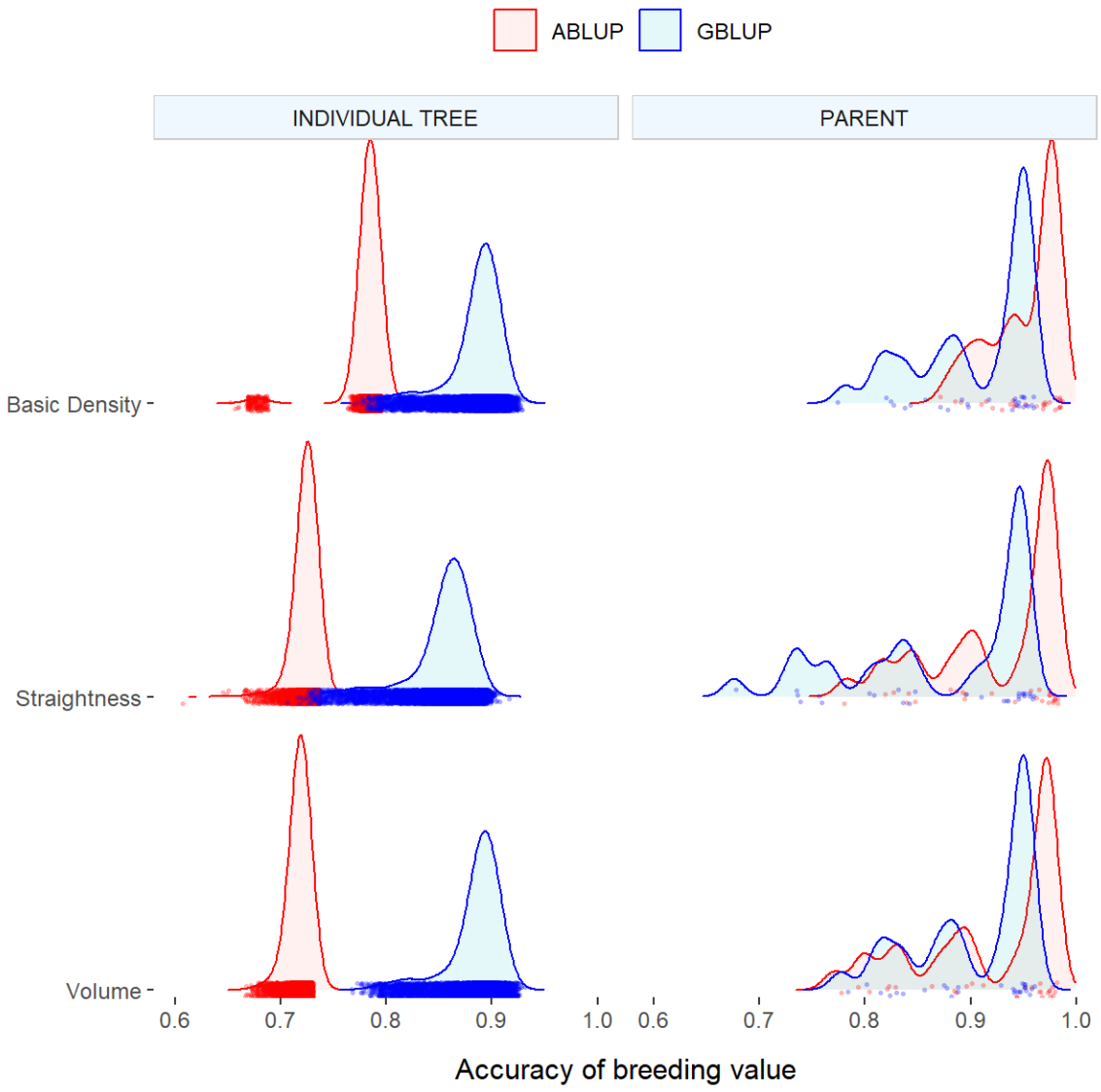


Figure 5.6. Distributions of prediction accuracy (r) of individual and parental trees with the A-BLUP and G-BLUP models for VOL, BD, and STR. The y-axis (frequency of individuals) was omitted.

5.3.4 Genomic selection validation

The results of genomic selection validation are presented in table 5.2.

Table 5.2. Predictive abilities (PA) of validation scenarios for VOL, BD, and STR. Variance (σ^2) and heritability (h^2) estimates are presented with their standard deviation.

Validation scenario	***VOL***	***BD***	***STR***
Environment 1 out			
σ^2_{tree}	592.0 ± 59.80	296.5 ± 29.61	0.128 ± 0.015
σ^2_e	1707. ± 31.01	862.16 ± 15.84	0.618 ± 0.011
h^2_G	0.258 ± 0.020	0.256 ± 0.020	0.172 ± 0.018
PA_{ind}	0.953	0.972	0.978
PA_{fam}	0.961	0.980	0.981
$\overline{PA}_{w.fam}$	0.851	0.863	0.910
Environment 2 out			
σ^2_{tree}	984.7 ± 99.85	360.1 ± 35.64	0.045 ± 0.006
σ^2_e	2322. ± 46.09	838.1 ± 16.63	0.248 ± 0.005
h^2_G	0.298 ± 0.022	0.301 ± 0.022	0.153 ± 0.018
PA_{ind}	0.966	0.971	0.836
PA_{fam}	0.968	0.991	0.911
$\overline{PA}_{w.fam}$	0.883	0.847	0.690
Environment 3 out			
σ^2_{tree}	713.2 ± 70.03	343.6 ± 32.29	0.137 ± 0.016
σ^2_e	1983. ± 35.95	825.3 ± 15.19	0.592 ± 0.011
h^2_G	0.265 ± 0.020	0.294 ± 0.021	0.188 ± 0.018
PA_{ind}	0.971	0.981	0.964
PA_{fam}	0.991	0.996	0.992
$\overline{PA}_{w.fam}$	0.920	0.890	0.907
Average PA_{ind}	0.964	0.975	0.919
Average PA_{fam}	0.973	0.989	0.961
Average $\overline{PA}_{w.fam}$	0.885	0.867	0.836

The rationale behind the leave-one-site-out validation was that by masking an entire trial, we would simulate the prediction of an independent testing set, with phenotypes reflecting the performance of trees grown in a different environment than the training population. The results show high, consistent PA across scenarios above 0.90 for individuals and family means.

The distributions of within-family predictive abilities ($PA_{w.fam}$) in each scenario for the trait VOL are presented in Figure 5.7. High predictive abilities for full-sib families of orchard

parents were verified, even for small families, regardless of presence or absence in the training population. Conversely, small families of male parents with low accuracy (contaminants) exhibited lower $PA_{w.fam}$.

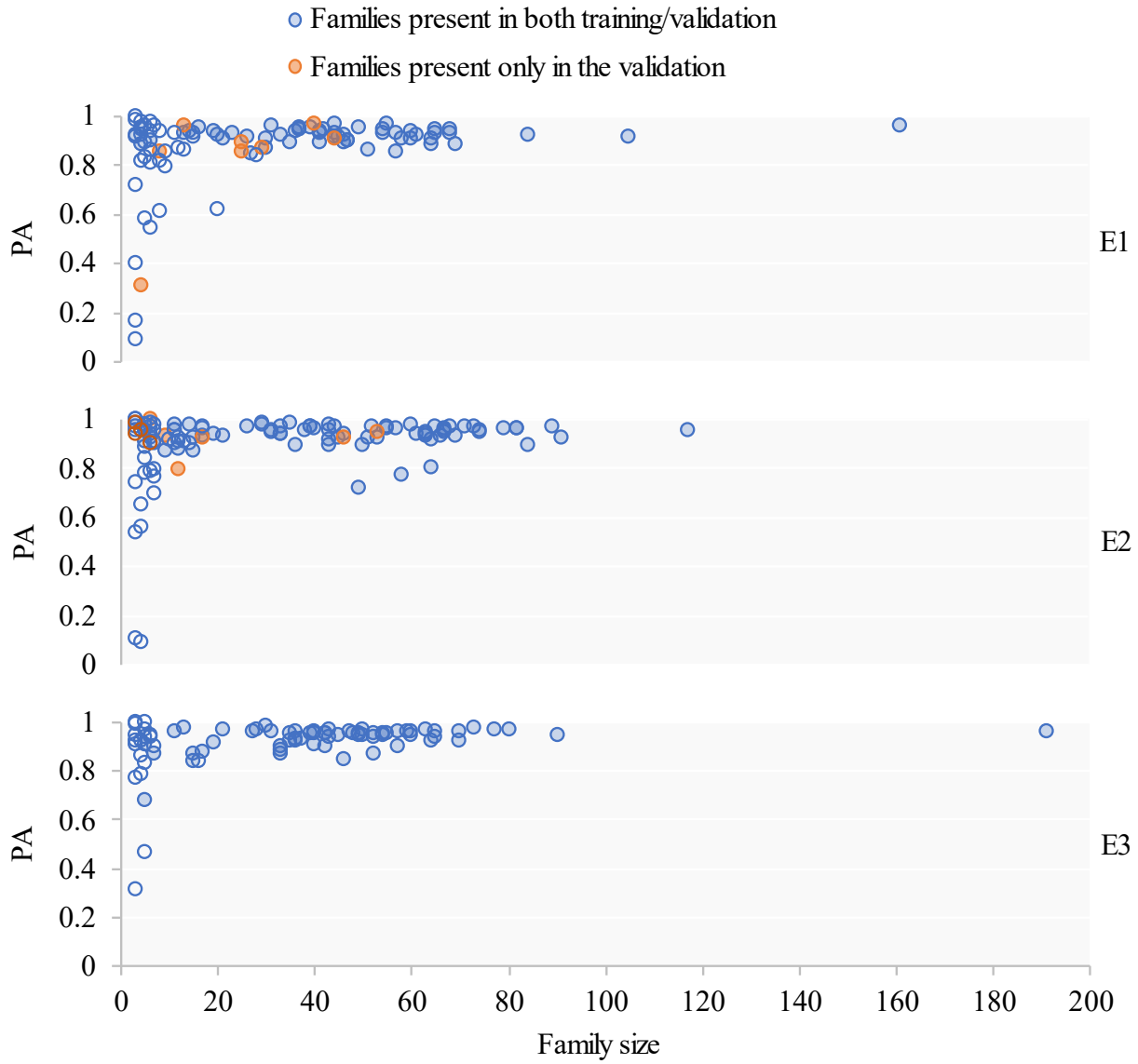


Figure 5.7. Distribution of $PA_{w.fam}$ over family size for the trait VOL. Full-sib families of orchard parents are presented in full circles and external male parents in empty circles. Families present in both training/validation are colored in blue, and those only present in the validation set in orange.

5.3.5 Predicted response to genomic selection

Even though the construction of indices for the optimal selection of multiple traits at once is a relevant aspect of a breeding strategy for wood production (White and Hodge 1989), it is not in the scope of this study. The focus here is on the results of individual selection on tree volume. The response to within-family genomic selection (RGS_w) for volume is presented in Table 5.3, with the indirect responses on BD and STR also present.

Table 5.3. Predicted response to genomic selection within family (RGS_w) for tree volume (VOL) for the five top-ranked families in each validation scenario. The indirect responses to the volume selection on basic density (BD) and straightness (STR) are also presented.

Validation Scenario	Family Rank	Family Size	VOL			BD	STR
			GEBV	$GEBV_{10\%}$	RGS_w	RGS_w	RGS_w
E1	1	42	140.1	160.6	12.2%	1.0%	0.8%
	2	45	135.4	162.2	16.5%	3.3%	4.1%
	3	60	135.1	159.6	15.1%	2.3%	-2.3%
	4	55	130.4	160.0	18.9%	2.0%	-0.8%
	5	68	129.7	157.3	17.7%	1.7%	3.4%
Mean		54	134.1	159.9	16.1%	2.1%	1.0%
E2	1	69	145.1	171.1	14.9%	2.3%	0.8%
	2	82	140.0	175.3	21.0%	1.6%	-1.6%
	3	68	139.8	176.5	21.9%	2.5%	1.6%
	4	63	136.8	167.9	18.9%	3.3%	1.9%
	5	74	136.3	172.5	22.1%	3.1%	1.4%
Mean		71	139.6	172.7	19.8%	2.5%	0.8%
E3	1	73	141.4	171.9	18.0%	3.5%	0.6%
	2	65	135.8	168.7	20.2%	2.5%	-2.0%
	3	54	132.3	160.8	18.0%	3.3%	-2.2%
	4	77	130.3	160.7	19.4%	2.0%	1.5%
	5	57	120.3	147.9	19.1%	2.4%	-0.3%
Mean		65	132.0	162.0	18.9%	2.7%	-0.5%
Grand Mean		63	135.3	164.9	18.3%	2.4%	0.5%

Within-family genomic selection, practiced with an intensity of 10%, delivered additional gains of 18.3% for volume, on average. Consistent positive gains were verified in all validation scenarios. The family with the lowest additional gain was the top family in scenario 1, with an RGS_w = 12.2%. Conversely, the family with the highest additional gain was the fifth-ranked family in scenario 2, with a RGS_w = 22.1%. Basic density responded positively and consistently to the selection for volume in all scenarios, with an average gain of 2.4%. Straightness was independent of the selection for volume, with varying small positive and negative gains across families and validation scenarios, with an average 0.5% gain.

5.4 Discussion

This study assessed the potential of genomic resources to improve progeny testing and selection in a population of *A. crassicarpa* tested in a multi-environmental trial. Like several *Acacia* sp., it is insect-pollinated, with reproductive biology that makes controlled crossing difficult. Breeding was based on testing open-pollinated families to identify seed parents with desirable progeny phenotypes. Vegetative propagation is cost-effective but is possible only with juvenile ortets. The deployment region consists of hundreds of thousands of hectares of relatively uniform climate and soil conditions, so the ability to identify elite germplasm during the juvenile stage would allow the deployment over large areas. Identifying parents of progeny in open-pollinated families would permit the establishment of full-sibling genetic trials to estimate additive and non-additive effects and specific and general combining abilities. Given these advantages, the strategic use of genomic tools for marker discovery can push forward breeding efforts accelerating genetic gains in wood productivity.

The first step in this process was to develop a proprietary chromosome-scale genome assembly, followed by a diversity survey to identify SNP variants and estimate allele frequencies

(results not shown). Following that, three genotyping panels were synthesized. A high-throughput and cost-effective low-density genotyping platform, capable of accurately determining *A. crassicarpa* genotypes and with the capability of paternity assignments; a mid-density panel for imputation and full parentage reconstruction; and a higher-density panel containing hundreds of thousands of SNPs for constructing high-resolution genetic maps and realized genomic relationship matrices.

An elite seed orchard with 18 high-ranked genotypes provided a large set of open-pollinated families from where a population of 84,315 seedlings was genotyped with the low-density SNP panel in the nursery. It allowed paternity assignments for about two-thirds of the progeny and identified 93 putative full-sib families with enough size to establish a multi-environmental replicated field trial. This number of crosses is more than half of the combinations possible in a half-diallel mating design. The diallel was obtained in one year, including flowering, open-pollination, pod harvesting, seed processing and germination, leaf sampling, parentage determination, full-sib family reconstruction and trial establishment. Using molecular markers allowed trial testing of a large collection of full-sib families for the first time for the species.

All progeny established in the trials were further genotyped with the mid-density panel for imputation of the high-density genotypes, allowing the construction of realized relationship matrices. The superiority of genomic-based models over pedigree-based models has been reported for growth traits in several tree species. The first advantage of molecular markers is to allow full-parental control revealing paternal coancestry in open-pollinated populations. Its direct impact on the model fit was reported for growth traits in white spruce and eucalypts studies with open-pollinated populations (Ratcliffe *et al*. 2017, Klápště *et al*. 2018). Another aspect that contributed to better model fit when using genomic relationships over pedigree-based relationships was the

inclusion of non-additive effects. Genomic models accounting for dominance effects performed better for growth traits in scots pine (Calleja-Rodriguez *et al.* 2020), loblolly pine (Walker *et al.* 2021), and hybrid eucalypts (Tan *et al.* 2018). In this study, however, the pedigree was validated with a mid-density panel setting an elevated threshold of 99% confidence for paternity assignment, leaving little room for improvements in model fit by revealing hidden relationships in the open-pollinated family. In addition, non-significant dominance effects were reported for tree height, straightness, and basic density in our population, minimizing the importance of dominance in the genetic control of growth traits for the species (chapter 4). Simulations have indicated that including nonadditive effects in the model only improves prediction ability when they are prominent (Denis and Bouvet, 2013; de Almeida Filho *et al*. 2016). Empirical studies supported this finding for growth traits in eucalypts (Bouvet *et al*. 2016, Resende *et al.* 2017) and loblolly pine (Muñoz *et al*. 2014; de Almeida Filho *et al*. 2016), showing that despite the documented importance of dominance, adding it or epistatic effects did not improve genomic prediction.

Variance component and genetic parameter estimates were variable when comparing models with and without genomic information, suggesting different proportions of phenotypic variance attributed to genetic effects for A-BLUP and G-BLUP models. For tree volume, the heritability of the genomic model $h^2_G = 0.28$ was larger than the pedigree model $h^2_a = 0.22$. The opposite was verified for basic density and straightness, with larger estimates for the ABLUP model. However, regardless of the magnitude fluctuation of the heritability estimates, their standard deviation was consistently smaller with the G-BLUP models, with at least a 3-fold reduction compared with the A-BLUP models, providing strong evidence that the inclusion of genomic information yielded more accurate heritability estimates. The forest research literature shows an inconsistent pattern of heritability estimates comparing pedigree models with genomic

models. Studies reported stable heritability estimates between A-BLUP and G-BLUP models (Ukrainetz and Mansfield 2020, Walker *et al*. 2021), while other studies reported less stable variance component estimates (Muñoz *et al.* 2014). Some studies with open-pollinated populations observed changes in additive genetic variance with genomic information (Gamal El-Dien *et al.* 2016). Additive and nonadditive variance can be confounded for certain population structures, with reduced additive variance estimates when genomic additive and dominance relationships are included in the model (Nazarian and Gezan 2016, Cappa *et al.* 2019).

Several previous studies also reported more precise breeding value predictions when including genomic relationships for breeding populations of forest trees (Muñoz *et al*. 2014, Ratcliffe *et al*. 2017, Cappa *et al*. 2017, Cappa *et al*. 2019, Jurcic *et al*. 2021), livestock (Legarra *et al*. 2014), and aquaculture (Vallejo *et al*. 2017, Yoshida *et al*. 2019). Traditionally, in forest trees, increased accuracy of genetic values can be achieved for parental genotypes with many offspring and individual genotypes that have been clonally replicated (Isik *et al*. 2005). The use of genomic relationships produces a similar result for unreplicated individual genotypes. The unit of replication is no longer the individual, but rather alleles and haplotypes shared among individuals (Walker *et al*. 2021). In this study's additive genomic model framework, genomic relationships consistently improved model fit beyond pedigree-based models for all traits evaluated. Given the size and structure of the genotyped population, most of the gain in accuracy arises from the dense identity-by-state based realized relationship matrix explaining the Mendelian sampling term, improving model fit by modeling relationships among individuals precisely (de los Campos *et al*. 2013). With the G-BLUP model, the accuracy of predictions of individual genotypes was nearly identical to those of parental genotypes, with estimates around r = 0.9. This excellent accuracy facilitates accurate selection among individuals, families, and within families.

Another potential advantage of genomic selection is the opportunity to reduce the breeding cycle through early selection. In forest trees, it is common to make selections around one-half of the commercial rotation age (Osorio *et al*. 2003). Genomic selection could identify next-generation breeding parents with high genetic value during the juvenile phase, selecting as early as three months in the nursery. Genomic selection can also provide additional genetic gain in deployment populations. Typically, seed orchards have a useful production life of many years. So knowledge about the genetic value of parents in the orchard can be exploited for multiple rotations with a family deployment strategy that should increase the yield and quality of the wood harvested from plantations within 5 to 10 years. With the within-family genomic selection, the genetic gain of family forestry can be increased by capturing the genetic variance within the family. To achieve these objectives and accelerate the rate of genetic gain in breeding populations and commercial germplasm, the efficacy of genomic selection must be high. The predictive ability results of the genomic models tested in this study were very satisfactory, with PAs above 0.9 for individual selection and family mean selection for all traits.

For within-family selection, families derived from parents with lower breeding value accuracies showed lower within-family PA. Conversely, full-sib families of elite seed orchard parents, each with thousands of progeny and very high breeding values accuracies, resulted in within-family PA above 0.8, regardless of the family size and if the family was present or not in the model training set. Given our study population structure and marker density, 100 progeny per parent and 50 individuals per family seems adequate to achieve high predictive abilities for all selection targets. Finally, for tree volume, the average response to within-family genomic selection practiced at 10% selection intensity on the top five ranked families in each validation scenario was 18%. This additional gain delivered by the within-family genomic selection was obtained for the

top five ranked families, with an average gain over the population mean of 28% already. It is gain over gain, maximizing the performance of family forestry in one generation.

5.5 Conclusion

Integrating genomic solutions into the breeding program of *A. crassicarpa*, a tree species with biological limitations to produce controlled crosses and vegetative propagation, allowed full parental control and the construction of genomic models with excellent breeding value prediction ability. Genomic models enable the selection of superior individual genotypes and families at a very early age, providing genetically improved trees in a juvenile state that can be vegetatively propagated for deployment over large areas. The genotyped population size of ten thousand individuals and the marker density with more than one hundred fifty-two thousand non-redundant SNP markers were the drivers of high genomic predictive ability in this population derived from 28 founders. There was considerable phenotypic variation in the population that could be explored with accurate genomic models that outperformed pedigree-based models for all traits based on their average reliability of individual tree predictions. The accurate individual tree selection resulted in valuable gains for all units of selection: individual trees for generation advancement or within-family genomic selection for deployment with family forestry. The predicted response to selection on the top-ranked full-sib families in the population showed the potential of within-family genomic selection to maximize the genetic gains achieved within a generation, doubling the gain compared to the deployment based on family means.

5.6 References

Assis TF, Resende MDV (2011) Genetic improvement of forest tree species. Crop Breeding and Applied Biotechnology, 11, 44–49. https://doi.org/10.1590/S1984-70332011000500007

Bernardo R (2020) Breeding for Quantitative Traits in Plants (3rd edition). Stemma Press: Woodbury, MN, USA, p. 301.

Bouvet JM, Makouanzi G, Cros D, Vigneron P (2016) Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. Heredity, 116, 146–157. https://doi.org/10.1038/hdy.2015.78

Butler DG, Cullis BR, Gilmour AR, Gogel BG, Thompson R (2017) ASReml-R Reference Manual Version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Calleja-Rodriguez A, Pan J, Funda T, Chen Z, Baison J, Isik F *et al*. (2020) Evaluation of the efficiency of genomic versus pedigree predictions for growth and wood quality traits in Scots pine. BMC Genomics, 21, 796. https://doi.org/10.1186/s12864-020-07188-4

Cappa EP, El-Kassaby YA, Muñoz F, Garcia MN, Villalba PV, Klápště P, Poltri SNM (2017) Improving accuracy of breeding values by incorporating genomic information in spatial-competition mixed models. Molecular Breeding, 37, 125. https://doi.org/10.1007/s11032-017-0725-6

Cappa EP, Lima BM, Silva-Junior OB, Garcia CC, Mansfield SD, Grattapaglia D (2019) Improving genomic prediction of growth and wood traits in Eucalyptus using phenotypes from nongenotyped trees by single-step GBLUP. Plant Science, 284, 9–15. https://doi.org/10.1016/j.plantsci.2019.03.017

Cumbie WP, Huber DA, Steel VC, Rottmann W *et al*. (2020) Marker associations for fusiform rust resistance in a clonal population of loblolly pine (Pinus taeda, L.). Tree Genetics & Genomes, 16, 86. https://doi.org/10.1007/s11295-020-01478-4

de Almeida Filho J, Guimarães J, e Silva F *et al*. (2016) The contribution of dominance to phenotype prediction in a pine breeding and simulated population. Heredity, 117, 33–41. https://doi.org/10.1038/hdy.2016.23

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics, 193, 327–345. https://doi.org/10.1534/genetics.112.143313

Denis M, Bouvet JM (2013) Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. Tree Genetics & Genomes, 9, 37–51. https://doi.org/10.1007/s11295-012-0528-1

Durán R, Isik F, Zapata-Valenzuela J, Balocchi C, Valenzuela S (2017) Genomic predictions of breeding values in a cloned Eucalyptus globulus population in Chile. Tree Genetics & Genomes, 13, 74. https://doi.org/10.1007/s11295-017-1158-4

El-Kassaby YA, Lstibůrek M (2009) Breeding without breeding. Genetics Research, 91:2, 111–120. http://dx.doi.org/10.1017/S001667230900007X

Falconer DS, Mackay TF (1996) Introduction to Quantitative Genetics (4th edition). 480 p. Longman Group Ltd.

Gamal El-Dien O, Ratcliffe B, Klápště J, Porth I, Chen C, El-Kassaby YA (2016) Implementation of the realized genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from nonadditive genetic effects. G3, 6, 743–753. https://doi.org/10.1534%2Fg3.115.025957

Gaynor RC, Gorjanc G, Hickey JM (2020) AlphaSimR: An R-package for Breeding Program Simulations. G3, 11:2, jkaa017. https://doi.org/10.1093/g3journal/jkaa017

Gezan SA, de Oliveira AA, Galli G, Murray D (2022) ASRgenomics: An R package with Complementary Genomic Functions. Version 1.1.0. VSN International, Hemel Hempstead, United Kingdom.

Gilmour A, Gogel B, Cullis B, Welham S, Thompson R (2014) ASReml user guide. Release 4.1 structural specification. VSN International Ltd, Hemel Hempstead, UK.

Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. Tree Genetics & Genomes, 7, 241–255. https://doi.org/10.1007/s11295-010-0328-4

Grattapaglia D (2017) Status and Perspectives of Genomic Selection in Forest Tree Breeding. In: Genomic Selection for Crop Improvement: New Molecular Breeding Strategies for Crop Improvement; Varshney, R.K., Roorkiwal, M., Sorrells, M.E., Eds.; Springer International Publishing: Cham, Switzerland, 199–249.

Grattapaglia D (2022) Twelve Years into Genomic Selection in Forest Trees: Climbing the Slope of Enlightenment of Marker Assisted Tree Breeding. Forests, 13, 1554. https://doi.org/10.3390/f13101554

Griffin AR, Vuong TD, Harbard JL, Wong CY, Brooker C, Vaillancourt RE (2010) Improving controlled pollination methodology for breeding *Acacia mangium* Willd. New Forests, 2010, 40:2, 131–142. https://doi.org/10.1007/s11056-010-9188-x

Griffin AR (2014) Clones or improved seedlings of Eucalyptus? Not a simple choice. The International Forestry Review, 16:2, 216-224. https://doi.org/10.1505/146554814811724793

Harwood CE, Hardiyanto EB, Wong CY (2015) Genetic improvement of tropical acacias: achievements and challenges. Southern Forests, 77, 11-18. https://doi.org/10.2989/20702620.2014.999302

Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research, 91:1, 47–60. https://doi.org/10.1017/s0016672308009981

Hickey JM, Kinghorn BP, Tier B *et al*. (2011) A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genetics Selection Evolution, 43, 12. https://doi.org/10.1186/1297-9686-43-12

Holland JB, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding: An update. Plant Breeding Reviews, 22, 9–111. https://doi.org/10.1002/9780470650202.ch2

Isik F, Goldfarb B, LeBude A, Li B, McKeand S (2005) Predicted genetic gains and testing efficiency from two loblolly pine clonal trials. Canadian Journal of Forest Research, 35, 1754–1766. https://doi.org/10.1139/x05-064

Isik F, Holland J, Maltecca C (2017) Genetic data analysis for plant and animal breeding. Cham, Switzerland: Springer International Publishing.

Isik F (2022) Genomic Prediction of Complex Traits in Perennial Plants: A Case for Forest Trees. In Complex Trait Prediction: Methods and Protocols; Ahmadi, N., Bartholomé, J., Eds.; Humana: New York, NY, USA, pp. 493–520.

Jurcic EJ, Villalba PV, Pathauer PS *et al*. (2021) Single-step genomic prediction of *Eucalyptus dunnii* using different identity-by-descent and identity-by-state relationship matrices. Heredity, 127, 176–189. https://doi.org/10.1038/s41437-021-00450-9

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology, 16:5, 1099-1106. https://doi.org/10.1111/j.1365-294x.2007.03089.x

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

Klápště J, Suontama M, Telfer E *et al*. (2017) Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of Eucalyptus nitens. PLoS ONE, 12(9), e0185137. https://doi.org/10.1371/journal.pone.0185137

Klápště J, Suontama M, Dungey HS, Telfer EJ, Graham NJ, Low CB, Stovold GT (2018) Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. Journal of Heredity, 109:7, 802–810. https://doi.org/10.1093/jhered/esy051

Lauer E, Isik F (2021) Major QTL confer race-nonspecific resistance in the co-evolved *Cronartium quercuum f. sp. Fusiforme-Pinus* taeda pathosystem. Heredity, 127, 288–299. https://doi.org/10.1038/s41437-021-00451-8

Lebedev VG, Lebedeva TN, Chernodubov AI, Shestibratov KA (2020) Genomic Selection for Forest Tree Improvement: Methods, Achievements and Perspectives. Forests, 11, 1190. https://doi.org/10.3390/f11111190

Legarra A, Christensen OF, Aguilar I, Misztal I (2014) Single step, a general approach for genomic selection. Livestock Science, 166, 54–65. https://doi.org/10.1016/j.livsci.2014.04.029

Lima BM, Cappa EP, Silva-Junior OB, Garcia C, Mansfield SD, 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, e0218747. https://doi.org/10.1371/journal.pone.0218747

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits (1st ed.). Sunderland: Sinauer Associates.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology, 7:5, 639-655. https://doi.org/10.1046/j.1365-294x.1998.00374.x

Muñoz PR, Resende MFR, Gezan SA, Resende MDV, Campos GDL, Kirst M, Huber D, Peter GF (2014) Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices. Genetics, 198:4, 1759–1768. https://doi.org/10.1534%2Fgenetics.114.171322

Nazarian A, Gezan SA (2016) Integrating nonadditive genomic relationship matrices into the study of genetic architecture of complex traits. Journal of Heredity, 107, 153–162. https://doi.org/10.1093/jhered/esv096

Osorio L, White T, Huber D (2003) Age-age and trait-trait correlations for *Eucalyptus grandis* Hill ex Maiden and their implications for optimal selection age and design of clonal trials. Theoretical and Applied Genetics, 106, 735–743. https://doi.org/10.1007/s00122-002-1124-9

Putz AM, Tiezzi F, Maltecca C, Gray KA, Knauer MT (2018) A comparison of accuracy validation methods for genomic and pedigree-based predictions of swine litter size traits using Large White and simulated data. Journal of animal breeding and genetics, 135:1, 5-13. https://doi.org/10.1111/jbg.12302

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratcliffe R, El-Dien OG, Cappa EP, Porth I, Klápště J, Chen C, El-Kassaby YA (2017) Single-Step BLUP with varying genotyping effort in *Picea glauca*. G3, 7, 935–942. https://doi.org/10.1534/g3.116.037895

Resende MDV, Resende MFR, Sansaloni CP *et al*. (2012) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytologist, 194, 116–128. https://doi.org/10.1111/j.1469-8137.2011.04038.x

Resende RT, Resende MDV, Silva FF *et al*. (2017) Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity, 119, 245–255. https://doi.org/10.1038/hdy.2017.37

Rezende GDSP, Resende MDV, Assis TF (2014) Eucalyptus breeding for clonal forestry. In: Fenning T (ed). Challenges and Opportunities for the World's Forests in the 21st Century. Springer Netherlands: Dordrecht, pp 393–424.

Tan B, Grattapaglia D, Martins GS, Ferreira KZ, Sundberg B, Ingvarsson PK (2017) Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. BMC Plant Biology, 17, 110. https://doi.org/10.1186/s12870-017-1059-6

Tan B, Grattapaglia D, Wu HX, Ingvarsson PK (2018) Genomic relationships reveal significant dominance effects for growth in hybrid Eucalyptus. Plant Science, 267, 84-93. https://doi.org/10.1016/j.plantsci.2017.11.011

Ukrainetz NK, Mansfield SD (2020) Prediction accuracy of single-step BLUP for growth and wood quality traits in the lodgepole pine breeding program in British Columbia. Tree Genetics & Genomes, 16, 64. https://doi.org/10.1007/s11295-020-01456-w

Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, Fragomeni BO *et al*. (2017) Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. Genetics Selection Evolution, 49:17. https://doi.org/10.1186/s12711-017-0293-6

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science, 91:11, 4414-4423. https://doi.org/10.3168/jds.2007-0980

Walker TD, Cumbie WP, Isik F (2021) Single-step genomic analysis increases the accuracy of within-family selection in a clonally replicated population of *Pinus taeda* L. Forest Science, 68:37–52. https://doi.org/10.1093/forsci/fxab054

Whalen A, Hickey JM (2020) AlphaImpute2: Fast and accurate pedigree and population based imputation for hundreds of thousands of individuals in livestock populations. BioRxiv https://doi.org/10.1101/2020.09.16.299677

Whalen A, Gorjanc G, Hickey JM (2020) AlphaFamImpute: high-accuracy imputation in full-sib families from genotype-by-sequencing data. Bioinformatics, 36:15, 4369–4371. https://doi.org/10.1093/bioinformatics/btaa499

Whetten RW, Jayawickrama KJS, Cumbie WP, Martins GS (2023) Genomic Tools in Applied Tree Breeding Programs: Factors to Consider. Forests, 14, 169. https://doi.org/10.3390/f14020169

Whitaker D, Williams E R, John J A (2001) CycDesigN: A Package for the Computer Generation of Experimental Designs. CSIRO Forestry and Forest Products, Canberra.

White TL, Adams WT, Neale DB (2007) Forest genetics. Wallingford: CAB International.

White TL, Davis J, Gezan S *et al*. (2014) Breeding for value in a changing world: Past achievements and future prospects. New Forests, 45, 301–309. https://doi.org/10.1007/s11056-013-9400-x

White TL, Hodge G (1989) Predicting breeding values with application in forest tree improvement. Kluwer, UK, p 208.

Wong CY, Yuliarto M (2014) Deployment of acacias in short rotation pulpwood plantation. In: Acacia 2014 "Sustaining the Future of Acacia Plantation Forestry" International Conference, IUFRO Working Party 2.08.07: Genetics and Silviculture of Acacias, Hue, Vietnam, 18–21 March 2014, Compendium of Abstracts.

Wu HX (2018) Benefits and risks of using clones in forestry – a review. Scandinavian Journal of Forest Research, 34, 352-359. https://doi.org/10.1080/02827581.2018.1487579

Yoshida GM, Carvalheiro R, Rodríguez FH, Lhorente JP, Yáñez JM (2019) Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. Genomics, 111, 127–132. https://doi.org/10.1016/j.ygeno.2018.01.008

Zobel BJ, Talbert J (1984) Applied Forest Tree Improvement. John Wiley & Sons: New York, NY, USA, p. 139.

Zobel BJ (1993) Clonal Forestry in the Eucalypts. In: Ahuja, MR., Libby, W.J. (eds) Clonal Forestry II. Springer, Berlin, Heidelber

CONCLUSION

Parentage analysis using SNP markers successfully characterized a reproductive season of a breeding orchard of *A. crassicarpa*. The species has a generalist entomophilous pollination syndrome with a mating dynamic favoring a rich male composition of open-pollination families. In general, most families showed many male parents without a dominant one. On average, the male that produced the most progeny in a given family had a 23% contribution. If seed orchard managers and breeders target orchard designs that provide optimal mating, open pollination should result in good admixture minimizing preferential mating. Our study showed a low selfing rate of 0.3%, corroborating an outcrossed breeding system for the species. From an applied tree breeding perspective, the parentage reconstruction efficiently generated many full-sib families, allowing progeny testing and overcoming the controlled pollination limitation to produce the crosses.

The wood and pulping properties estimates obtained in the present study showed the suitability of the species' wood for efficient pulp production, with lignin contents, carbohydrates contents, and kraft pulp yields in the range of the hardwoods commercially planted around the world. The within-tree longitudinal pattern of variation showed a consistent decreasing trend from base to top of the bole for basic density and insoluble lignin. In contrast, no consistent pattern was observed for the carbohydrates, soluble lignin, and S/G ratio. The reliability of sets of positions taken singly or combined to predict the whole-tree phenotype varied along with the different traits. For pulp yield, basic density, glucose content, and lignin content, reliable ground-level direct measurement sampling was found, with very high correlations. With a NIR prediction model of basic density with observed cross-validation $R^2_{cv} = 0.75$, a 0.12 reduction in the reliability of breast height sampling was verified, but still with a 0.80 Spearman ranking correlation, which could efficiently rank the trees for selection in a breeding program.

The genetic control of all traits assessed in this study was mainly additive. In this scenario, the simple recurrent selection is recommended as a breeding strategy targeting genetic gains at individual and family levels, with recombination of forward selections for generation advancing and backward selection of parental combinations of the best full-sib families to produce seed for deployment via family forestry. Small but significant dominance effects for DBH and MAI can be explored, looking forward to the positive specific combining ability effects and heterosis that can increase the genetic value of families and, ultimately, the volume of wood produced. Furthermore, the additive Type B genetic correlation across sites was very high for all traits. Consequently, the genotype-by-environment interaction will likely have a minor influence on the genotypic performance. A single breeding population can be developed with the potential to provide genotypes and families with broad adaptation and stability over the environmental range represented by the sites studied.

The pattern of genetic correlations among traits simultaneously favors the genetic progress for all traits. Straightness and survival were independent of growth traits, and tree volume was correlated with the mean annual increment and wood basic density. Thus, trees with the highest possible volume and straightness scores should be the target for individual tree selection, advancing the breeding population and promoting genetic gains simultaneously for all traits.

Integrating genomic solutions into the breeding program of *Acacia crassicarpa*, a tree species with biological limitations to produce controlled crosses and vegetative propagation, made possible full parental control and the construction of genomic models with excellent breeding value prediction ability. Genomic models enable the selection of superior individual genotypes and families at a very early age, providing genetically improved trees in a juvenile state that can be vegetatively propagated for deployment over large areas. The genotyped population size of ten

thousand individuals and the marker density with more than one hundred fifty-two thousand SNP markers were the drivers of high genomic predictive ability in this population derived from 28 founders. There was considerable phenotypic variation in the population that could be explored with accurate genomic models that outperformed pedigree-based models for all traits based on their average reliability of individual tree predictions. The accurate individual tree selection resulted in valuable gains for all units of selection: individual trees for generation advancement or within-family genomic selection for deployment with family forestry. The predicted response to selection on the top-ranked full-sib families in the population showed the potential of within-family genomic selection to maximize the genetic gains achieved within a generation, doubling the gain compared to the deployment based on family means.

APPENDICES

Appendix A1. Summary of the SNP panel validation.

Sample Set	Samples	Assay Replicates	SNP			Call Rate	Call Accuracy
			Total	Called	Concordant		
Orchard trees	77	3	9,702	9,575	9,445	98.7%	98.6%
Seedlings	96	4	16,128	15,834	15,772	98.2%	99.6%
Total	173		25,830	25,409	25,217	98.4%	99.2%

Appendix A2. Allele frequency analysis of the 42 SNP markers used for parentage analysis. N: number of individuals genotyped; MAF: minor allele frequency; H_O: observed heterozygosity; H_E: expected heterozygosity; PIC: Polymorphic information content; NE-1P: Average non-exclusion probability (NEP) for one candidate parent; NE-2P: Average NEP for one candidate parent given the genotype of a known parent of the opposite sex; NE-PP: Average NEP for a candidate parent pair; NE-I: Average NEP for the identity of two unrelated individuals; NE-SI: Average NEP for the identity of two siblings.

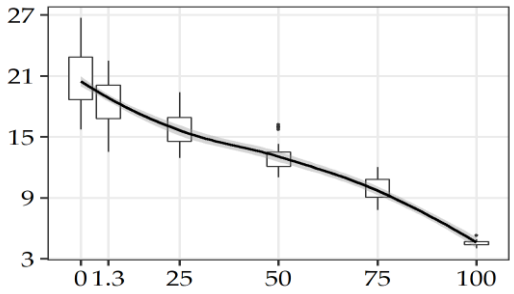
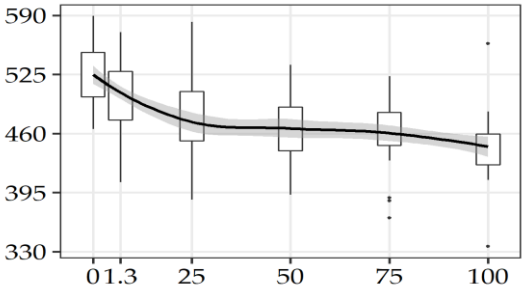
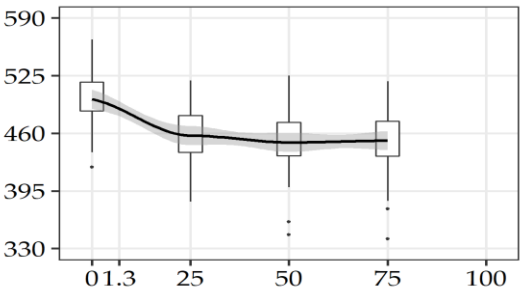
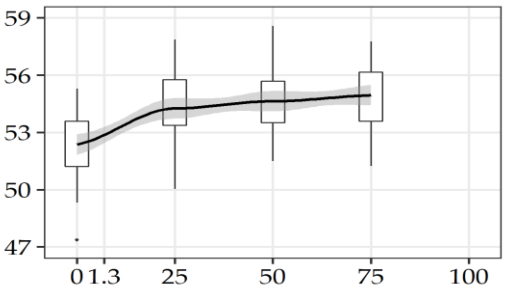
Locus	N	MAF	H_O	H_E	PIC	NE-1P	NE-2P	NE-PP	NE-I	NE-SI
SNP01	82273	0.433	0.504	0.491	0.371	0.879	0.815	0.722	0.380	0.599
SNP02	81088	0.485	0.566	0.500	0.375	0.875	0.813	0.719	0.375	0.594
SNP03	80816	0.421	0.507	0.487	0.369	0.881	0.816	0.723	0.382	0.602
SNP04	79763	0.483	0.523	0.499	0.375	0.875	0.813	0.719	0.375	0.594
SNP05	80218	0.396	0.511	0.478	0.364	0.886	0.818	0.726	0.386	0.607
SNP06	81215	0.484	0.614	0.499	0.375	0.875	0.813	0.719	0.375	0.594
SNP07	83295	0.452	0.511	0.495	0.373	0.877	0.814	0.720	0.377	0.597
SNP08	81423	0.461	0.469	0.497	0.373	0.877	0.813	0.720	0.377	0.596
SNP09	80182	0.326	0.476	0.439	0.343	0.904	0.829	0.738	0.411	0.633
SNP10	82465	0.387	0.465	0.474	0.362	0.887	0.819	0.727	0.389	0.610
SNP11	82478	0.380	0.510	0.471	0.360	0.889	0.820	0.728	0.391	0.612
SNP12	80444	0.353	0.488	0.457	0.352	0.896	0.824	0.733	0.400	0.622
SNP13	81562	0.422	0.506	0.488	0.369	0.881	0.816	0.723	0.381	0.602
SNP14	83244	0.406	0.485	0.482	0.366	0.884	0.817	0.724	0.384	0.605

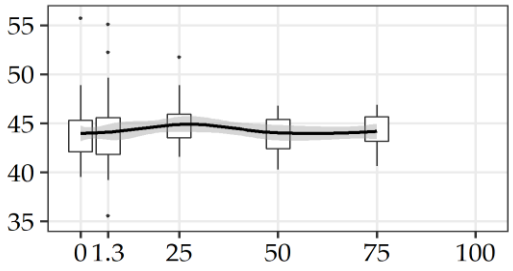
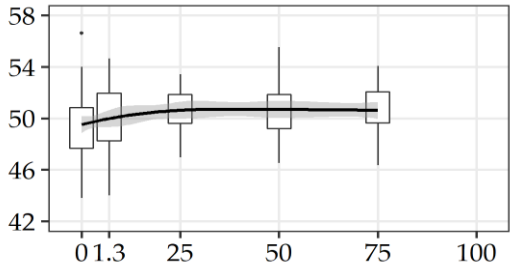
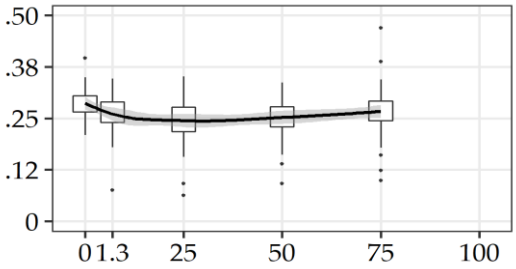
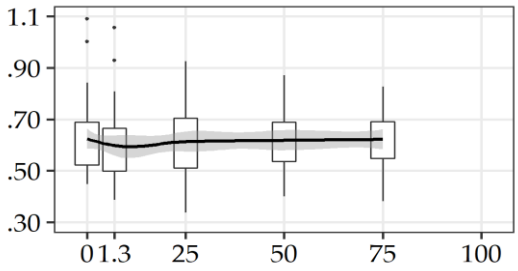
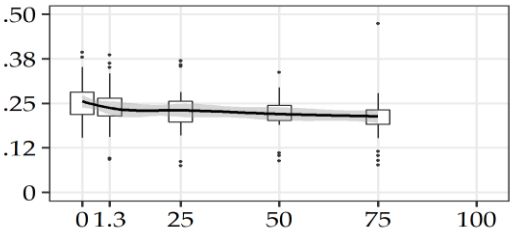
Locus	N	MAF	H_O	H_E	PIC	NE-1P	NE-2P	NE-PP	NE-I	NE-SI
SNP15	80509	0.431	0.515	0.491	0.370	0.880	0.815	0.722	0.380	0.600
SNP16	82648	0.473	0.525	0.499	0.374	0.876	0.813	0.719	0.376	0.595
SNP17	82518	0.500	0.594	0.500	0.375	0.875	0.813	0.719	0.375	0.594
SNP18	79341	0.457	0.536	0.496	0.373	0.877	0.813	0.720	0.377	0.596
SNP19	77935	0.394	0.484	0.478	0.364	0.886	0.818	0.726	0.387	0.608
SNP20	82878	0.301	0.435	0.421	0.332	0.911	0.834	0.745	0.424	0.645
SNP21	82938	0.485	0.567	0.500	0.375	0.875	0.813	0.719	0.375	0.594
SNP22	82069	0.458	0.498	0.497	0.373	0.877	0.813	0.720	0.377	0.596
SNP23	83064	0.497	0.518	0.500	0.375	0.875	0.813	0.719	0.375	0.594
SNP24	80607	0.319	0.435	0.434	0.340	0.906	0.830	0.740	0.414	0.637
SNP25	83201	0.455	0.523	0.496	0.373	0.877	0.814	0.720	0.377	0.596
SNP26	82425	0.496	0.539	0.500	0.375	0.875	0.813	0.719	0.375	0.594
SNP27	83148	0.461	0.516	0.497	0.373	0.877	0.813	0.720	0.377	0.596
SNP28	83238	0.481	0.531	0.499	0.375	0.875	0.813	0.719	0.375	0.594
SNP29	82802	0.339	0.459	0.448	0.348	0.899	0.826	0.735	0.405	0.627
SNP30	82722	0.469	0.521	0.498	0.374	0.876	0.813	0.719	0.376	0.595
SNP31	83125	0.419	0.533	0.487	0.368	0.882	0.816	0.723	0.382	0.602
SNP32	82603	0.496	0.529	0.500	0.375	0.875	0.813	0.719	0.375	0.594
SNP33	82523	0.439	0.519	0.493	0.371	0.879	0.814	0.721	0.379	0.598
SNP34	83419	0.355	0.474	0.458	0.353	0.895	0.823	0.732	0.399	0.621
SNP35	83078	0.434	0.528	0.491	0.371	0.879	0.815	0.721	0.379	0.599
SNP36	83062	0.344	0.479	0.451	0.349	0.898	0.825	0.734	0.403	0.625
SNP37	82766	0.490	0.533	0.500	0.375	0.875	0.813	0.719	0.375	0.594
SNP38	83404	0.286	0.386	0.409	0.325	0.916	0.837	0.749	0.433	0.654
SNP39	82674	0.357	0.526	0.459	0.354	0.895	0.823	0.732	0.398	0.620
SNP40	82593	0.472	0.528	0.498	0.374	0.876	0.813	0.719	0.376	0.595
SNP41	83205	0.400	0.490	0.480	0.365	0.885	0.818	0.725	0.386	0.606
SNP42	83223	0.439	0.499	0.493	0.371	0.879	0.814	0.721	0.379	0.598
Mean	82100	0.425	0.508	0.482	0.365	0.884	0.817	0.725	0.386	0.606

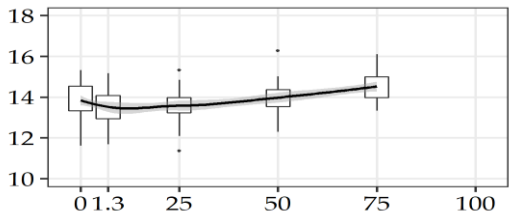
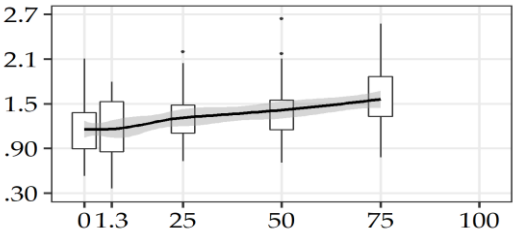
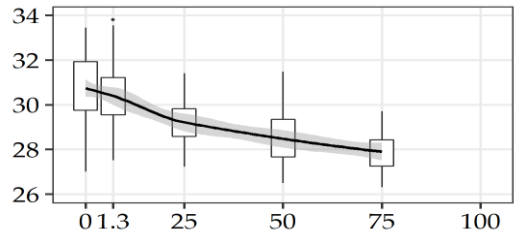
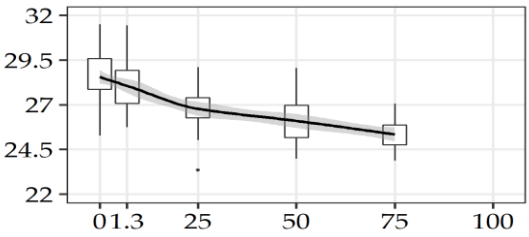
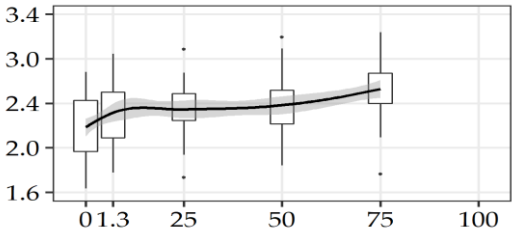
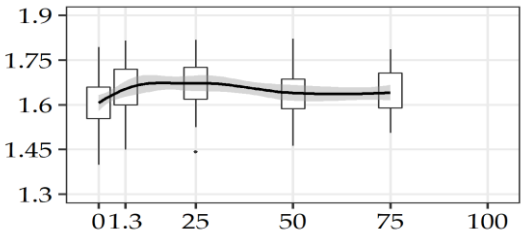
Appendix A3. The pedigree reconstruction revealed the full diallel of the 18 breeding parents present in the ESO. Female parents are in rows, and male parents are in columns. Diagonal cells show the number of self-pollinated progeny per parent.

F/M	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	Total
P01	3	137	19		10		234	9		2	12	2	6	39	1	14	64	28	580
P02	256	48	991	4	852	23	1864	951	10	541	368	532	255	401	61	426	631	34	8248
P03	128	239			125	118	484	30	2	46	54	36	15	79	10	114	140	62	1682
P04																			0
P05	131	531	310	4	72	539	1885	520	4	469	274	790	33	47	28	309	77	129	6152
P06	3	26	279	5	577	164	1479	346	1	470	154	62	77	21	92	626	6	8	4396
P07	138	1643	255	1	461	321	11	150	6	59	406	90	86	6	7	404	38	166	4248
P08																			0
P09																			0
P10	169	550	269	2	535	2391	555	866	2	20	1283	67	149	67	29	5	14	220	7193
P11	127	333	329		412	689	422	362	1	643	2	75	71	51	10	706	48	175	4456
P12	29	416	280	1	614	335	418	596	1	392	75	4	22	23	5	2	51	39	3303
P13	13	231	8	1		12	42			4	2	5				6	1	3	328
P14	23	229	43		6	37	2	4		10	18	11	4	2	2	16	52	16	475
P15																			0
P16	222	678	1801	2	602	3522	1263	448	17	14	391	34	150	32	46	42	147	750	10161
P17	171	306	121		94	62	206	14		5	41	12	16	116	3	48	2	59	1276
P18	149	57	249	5	173	23	1366	36	5	68	259	54	248	73	49	1494	79	53	4440
Total	1562	5424	4954	25	4533	8236	10231	4332	49	2743	3339	1774	1132	957	343	4212	1350	1742	56938

Appendix A4. Descriptive statistics number of observations (N), mean, standard deviation (SD), coefficient of variation (CV) and box plots of the 40 trees selected for the destructive sampling by position for diameter, disc basic density (DBD), chips basic density (CBD), screened kraft pulp yield (KPY), alpha cellulose (αCEL), glucose (GLU), arabinose (ARA), galactose (GAL), rhamnose (RHA), xylose (XYL), mannose (MAN), total lignin (LIG), insoluble lignin (INS), acid-soluble lignin (SOL) and syringyl-guaiacyl ratio (S/G).

Variable	Pos	N	Mean	SD	CV	Mean longitudinal variation
Diameter	0	40	20.9	2.8	13.5	
	1.3	40	18.2	2.2	11.8	
	25	40	15.9	1.7	10.8	
	50	40	13.1	1.3	10.0	
	75	40	9.9	1.1	11.4	
	100	40	4.5	0.3	5.6	
DBD	0	40	527.9	31.5	6.0	
	1.3	40	499.8	39.8	8.0	
	25	40	475.0	43.8	9.2	
	50	40	466.6	36.6	7.8	
	75	40	462.8	32.2	7.0	
	100	40	445.1	32.8	7.4	
CBD	0	40	498.4	32.9	6.6	
	1.3	0	477.9			
	25	40	457.4	31.6	6.9	
	50	40	449.7	36.0	8.0	
	75	40	451.8	36.4	8.1	
	100	0				
KPY	0	40	52.4	1.8	3.4	
	1.3	0	53.3			
	25	40	54.3	1.8	3.4	
	50	40	54.6	1.6	2.9	
	75	40	54.9	1.6	2.9	
	100	0				

Variable	Pos	N	Mean	SD	CV	Mean longitudinal variation
αCEL	0	40	44.0	3.0	6.8	
	1.3	40	44.1	3.5	8.0	
	25	40	44.9	2.2	4.8	
	50	40	44.0	1.8	4.0	
	75	40	44.2	1.7	3.9	
GLU	0	40	49.5	2.5	5.0	
	1.3	40	50.0	2.4	4.9	
	25	40	50.6	1.7	3.4	
	50	40	50.7	2.0	4.0	
	75	40	50.6	1.8	3.5	
ARA	0	40	0.29	0.04	12.8	
	1.3	40	0.26	0.05	18.0	
	25	40	0.24	0.06	23.0	
	50	40	0.25	0.05	19.4	
	75	40	0.27	0.06	23.7	
GAL	0	40	0.63	0.14	22.1	
	1.3	40	0.60	0.15	24.8	
	25	40	0.61	0.12	20.1	
	50	40	0.62	0.11	18.2	
	75	40	0.62	0.10	16.0	
RHA	0	40	0.26	0.06	22.5	
	1.3	40	0.24	0.06	25.6	
	25	40	0.23	0.07	28.9	
	50	40	0.22	0.05	21.6	
	75	40	0.21	0.06	30.1	

Variable	Pos	N	Mean	SD	CV	Mean longitudinal variation
XYL	0	40	13.8	0.9	6.4	
	1.3	40	13.5	0.8	6.0	
	25	40	13.6	0.8	5.8	
	50	40	14.0	0.7	5.1	
	75	40	14.5	0.6	4.5	
MAN	0	40	1.16	0.37	32.2	
	1.3	40	1.16	0.38	32.8	
	25	40	1.31	0.32	24.1	
	50	40	1.42	0.37	26.5	
	75	40	1.56	0.40	25.8	
LIG	0	40	30.7	1.6	5.1	
	1.3	40	30.4	1.5	5.0	
	25	40	29.2	1.1	3.7	
	50	40	28.5	1.2	4.1	
	75	40	27.9	0.8	3.0	
INS	0	40	28.6	1.5	5.2	
	1.3	40	28.1	1.5	5.2	
	25	40	26.8	1.2	4.5	
	50	40	26.1	1.2	4.5	
	75	40	25.3	0.8	3.3	
SOL	0	40	2.18	0.26	12.1	
	1.3	40	2.34	0.30	12.6	
	25	40	2.35	0.22	9.5	
	50	40	2.38	0.25	10.4	
	75	40	2.53	0.28	10.9	
S/G	0	40	1.61	0.09	5.5	
	1.3	40	1.65	0.09	5.4	
	25	40	1.67	0.08	5.1	
	50	40	1.65	0.09	5.7	
	75	40	1.64	0.07	4.5	

Appendix A5. Whole-tree level phenotypic correlation matrix between the wood property traits disc basic density (DBD), chips basic density (CBD), composite chips basic density (CBDc), screened kraft pulp yield (KPY), composite screened kraft pulp yield (KPYc), alpha cellulose (αCEL), glucose (GLU), arabinose (ARA), galactose (GAL), rhamnose (RHA), xylose (XYL), mannose (MAN), total lignin (LIG), insoluble lignin (INS), acid-soluble lignin (SOL) and syringyl-guaiacyl ratio (S/G).

	DBD	**CBD**	**CBDc**	**KPY**	**KPYc**	**αCEL**	**GLU**	**ARA**	**GAL**	**RHA**	**XYL**	**MAN**	**LIG**	**INS**	**SOL**
CBD	**0.92****														
CBDc	**0.91****	**0.96****													
KPY	**0.52****	**0.43****	**0.36***												
KPYc	**0.47****	**0.36***	0.30	**0.93****											
αCEL	**0.44****	**0.35***	0.29	**0.36***	**0.35***										
GLU	0.28	0.18	0.10	**0.49****	**0.48****	**0.70****									
ARA	**0.47****	**0.45****	**0.42****	0.08	0.07	0.30	0.09								
GAL	**0.59****	**0.49****	**0.44****	**0.43****	**0.37***	**0.44****	**0.37***	**0.68****							
RHA	**0.50****	**0.47****	**0.42****	0.16	0.13	**0.40***	0.16	**0.88****	**0.68****						
XYL	-0.10	-0.17	-0.12	-0.21	-0.17	0.02	0.06	-0.30	**-0.37***	-0.27					
MAN	**-0.37***	**-0.49****	**-0.40***	-0.24	-0.11	0.02	0.13	-0.17	**-0.31***	-0.02	0.24				
LIG	-0.21	-0.09	-0.10	**-0.46****	**-0.49****	**-0.48****	**-0.60****	-0.02	**-0.43****	-0.15	0.10	-0.15			
INS	-0.20	-0.09	-0.07	**-0.47****	**-0.50****	**-0.41****	**-0.55****	0.06	**-0.35***	-0.08	0.08	-0.12	**0.97****		
SOL	**-0.41****	-0.30	**-0.40***	-0.20	-0.20	-0.28	-0.24	**-0.39***	**-0.58****	**-0.35***	0.05	-0.03	0.31	0.13	
S/G	0.16	0.17	0.09	0.18	0.06	-0.05	-0.09	0.03	0.11	-0.01	0.13	**-0.44****	-0.10	-0.14	0.18

Appendix A6. Exploratory analysis of phenotypic data for the traits survival (SUR), straightness (STR), tree height (HT), diameter at breast height (DBH), tree volume (VOL), mean annual increment (MAI), basic density (BD) and ton mean annual increment (MAI_T) at ages 12, 18, 24, 30 and 36 months of age.

Variable	Environment	n	mean	sd	median	min	max	range
sur.12	E1	6400	0.92	0.28	1	0	1	1
sur.12	E2	7440	0.86	0.34	1	0	1	1
sur.12	E3	5520	0.92	0.27	1	0	1	1
sur.18	E1	6400	0.89	0.31	1	0	1	1
sur.18	E2	7440	0.86	0.35	1	0	1	1
sur.18	E3	5520	0.89	0.31	1	0	1	1
sur.24	E1	6400	0.86	0.34	1	0	1	1
sur.24	E2	7440	0.82	0.38	1	0	1	1
sur.24	E3	5520	0.85	0.35	1	0	1	1
sur.30	E1	6400	0.82	0.38	1	0	1	1
sur.30	E2	7440	0.80	0.40	1	0	1	1
sur.30	E3	5520	0.81	0.39	1	0	1	1
sur.36	E1	6400	0.66	0.47	1	0	1	1
sur.36	E2	7440	0.77	0.42	1	0	1	1
sur.36	E3	5520	0.73	0.45	1	0	1	1
str.36	E1	4244	3.1	0.5	3.0	2.0	5.0	3.0
str.36	E2	5719	3.8	1.1	4.0	1.0	6.0	5.0
str.36	E3	3951	3.2	0.5	3.0	1.0	6.0	5.0
ht.12	E1	5872	7.8	0.8	8.0	1.0	9.6	8.6
ht.12	E2	6431	6.3	0.5	6.4	3.6	7.5	3.9
ht.12	E3	5081	7.9	0.9	8.0	1.9	10.9	9.0
ht.18	E1	5709	11.0	1.1	11.1	4.4	13.7	9.3
ht.18	E2	6407	10.6	1.0	10.6	3.5	13.0	9.5
ht.18	E3	4937	10.9	1.1	11.0	3.5	19.9	16.4
ht.24	E1	5533	12.4	2.4	12.8	3.2	16.9	13.7
ht.24	E2	6109	12.6	2.3	12.5	2.8	18.0	15.2
ht.24	E3	4707	12.9	2.3	13.5	1.8	16.7	14.9
ht.30	E1	5275	14.1	2.6	14.4	2.6	20.3	17.7
ht.30	E2	5969	14.5	2.1	14.8	4.8	21.5	16.7
ht.30	E3	4491	14.5	2.3	15.1	3.6	18.8	15.2
ht.36	E1	4244	15.9	3.4	15.9	2.8	23.9	21.1
ht.36	E2	5751	15.9	2.4	15.8	4.8	22.6	17.8
ht.36	E3	4016	16.3	2.7	16.5	3.4	22.5	19.1
dbh.12	E1	5937	7.9	1.4	8.0	1.8	11.7	9.9
dbh.12	E2	6434	7.4	1.2	7.5	2.2	11.9	9.7
dbh.12	E3	5077	7.6	1.3	7.8	0.8	11.8	11.0
dbh.18	E1	5523	10.2	1.6	10.4	2.8	15.5	12.7
dbh.18	E2	6405	9.6	1.6	9.8	1.6	15.2	13.6
dbh.18	E3	4927	9.8	1.7	10.0	2.8	21.6	18.8

Variable	Environment	n	mean	sd	median	min	max	range
dbh.24	E1	5533	10.8	2.6	11.0	1.3	18.3	17.0
dbh.24	E2	6109	10.2	2.3	10.3	2.5	18.0	15.5
dbh.24	E3	4707	10.5	2.4	10.7	1.8	17.6	15.8
dbh.30	E1	5275	12.1	2.9	12.1	3.0	21.5	18.5
dbh.30	E2	5969	11.4	2.5	11.2	3.0	21.3	18.3
dbh.30	E3	4491	11.7	2.7	11.6	3.0	20.6	17.6
dbh.36	E1	4244	13.4	3.4	13.1	3.5	23.8	20.3
dbh.36	E2	5751	12.2	2.7	11.9	3.3	23.4	20.1
dbh.36	E3	4016	13.0	3.1	12.7	4.0	25.5	21.5
vol.12	E1	5839	0.018	0.006	0.019	0.001	0.040	0.039
vol.12	E2	6430	0.013	0.004	0.013	0.001	0.033	0.032
vol.12	E3	5077	0.017	0.006	0.018	0.000	0.040	0.040
vol.18	E1	5395	0.041	0.012	0.041	0.002	0.097	0.094
vol.18	E2	6401	0.035	0.011	0.035	0.001	0.083	0.082
vol.18	E3	4927	0.037	0.012	0.038	0.002	0.110	0.108
vol.24	E1	5533	0.054	0.027	0.052	0.001	0.157	0.156
vol.24	E2	6109	0.049	0.024	0.045	0.002	0.157	0.156
vol.24	E3	4707	0.053	0.025	0.052	0.001	0.145	0.145
vol.30	E1	5275	0.075	0.040	0.069	0.002	0.246	0.244
vol.30	E2	5969	0.068	0.032	0.063	0.003	0.250	0.247
vol.30	E3	4491	0.072	0.035	0.068	0.003	0.213	0.210
vol.36	E1	4244	0.104	0.061	0.089	0.002	0.348	0.346
vol.36	E2	5751	0.085	0.043	0.076	0.004	0.333	0.329
vol.36	E3	4016	0.100	0.053	0.088	0.005	0.361	0.357
mai.12	E1	5839	28.3	9.6	28.7	1.2	63.4	62.2
mai.12	E2	6430	19.3	6.6	19.2	1.2	54.8	53.6
mai.12	E3	5077	27.1	9.4	27.3	0.2	67.2	67
mai.18	E1	5395	41.3	12.8	40.9	2.4	93.5	91.1
mai.18	E2	6401	34.8	12.0	34.6	0.8	86.2	85.4
mai.18	E3	4927	37.9	13.0	37.6	1.6	94.6	93
mai.24	E1	5533	39.5	20.6	37.6	0.6	117.1	116.5
mai.24	E2	6109	34.6	17.8	31.7	1.6	123	121.4
mai.24	E3	4707	38.5	18.8	37.2	0.5	106.7	106.2
mai.30	E1	5275	42.4	23.2	38.3	0.9	153.7	152.8
mai.30	E2	5969	37.7	18.9	34.4	1.5	146.7	145.2
mai.30	E3	4491	40.3	20.4	37.2	1.7	142.3	140.6
mai.36	E1	4244	42.0	27.4	34.9	0.8	167.7	166.9
mai.36	E2	5751	38.2	20.7	33.6	1.7	173.3	171.6
mai.36	E3	4016	42.3	24.1	36.3	1.3	158.7	157.4
bd.36	E1	4242	422.5	37.0	422.3	302.0	556.4	254.4
bd.36	E2	5542	427.7	35.4	426.1	300.8	561.1	260.3
bd.36	E3	3968	429.3	36.6	429.1	315.6	569.1	253.5
maiT.36	E1	4242	18.3	12.8	14.8	0.3	79.3	79.0
maiT.36	E2	5542	16.6	9.9	14.3	0.6	79.7	79.1
maiT.36	E3	3968	18.7	11.4	15.6	0.5	73.8	73.3

Appendix A7. Diallel structure of the population. Orchard parents are presented in blue and pollen contaminants are presented in green.

	P01	P02	P03	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P16	P17	P18	X Sum	Total
P01																	0	509
P02	69	1															70	2096
P03	54	182															236	1619
P05	8	200	96														304	1657
P06		8	172	163													343	1653
P07	127	611	214	206	153	1											1312	2402
P08		134		122	210	40											506	1060
P09																	0	86
P10	44	137	134	146	232	159	126										978	1471
P11	25	44	196	144	115	195	160		113								992	1366
P12		120	116	190	72	134	109		112	37							890	955
P13		45			7	46			44	1	9						152	260
P14	17	186	64	17		1			29	25							339	435
P16	80	131	197	129	268	206	121	4	6	159		11	7				1319	1688
P17	34	128	119	53	131	106				13	12	1	59	93			749	850
P18	51	4	67	75		175	38	82	143	89	28	96	20	116	55		1039	1095
P101															25	12	37	37
P102		1	2	1		1				2	2			6		10	25	25
P103		25	1	10	2	3			1	1	1		1	8	2		55	55
P104		1			26	1				10				2		1	41	41
P105		4		12	9				8	1	3		1	10	6	8	62	62
P106		1		8	16	3			2	3	1		1	30	1	4	70	70
P107		9	1	4	7				2		1			7			31	31
P108		19	3	10	19	1			6	12	2		5	8	4		89	89
P109		8		15	11	5			10		5		1	21		1	77	77
P110		16		43	18	11			13	15			1	51	3	16	187	187
P111		6	1	3	7	1								13	5	4	40	40
P112		5		2	7	1			4	6	1			4			30	30
Y Sum	509	2026	1383	1353	1310	1090	554	86	493	374	65	108	96	369	101	56	9973	19946