GENOMIC BREEDING FOR ACCELERATED IMPROVEMENT OF GROWTH, WOOD PROPERTIES AND PLANT DEFENCE IN *EUCALYPTUS GRANDIS*

by

MAKOBATJATJI MMOLEDI MPHAHLELE

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor

Faculty of Natural and Agricultural Sciences

Department of Biochemistry, Genetics and Microbiology

University of Pretoria

Pretoria

March 2022

Under the supervision of Prof Alexander A. Myburg,

and co-supervision of Prof Fikret Isik and Prof Gary R. Hodge

DECLARATION

I, **MAKOBATJATJI MMOLEDI MPHAHLELE,** student number 23101271, declare that the dissertation I hereby submit for the PhD: Genetics degree at the University of Pretoria is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: _____

Date: 10th July 2022

THESIS SUMMARY

GENOMIC BREEDING FOR ACCELERATED IMPROVEMENT OF GROWTH, WOOD PROPERTIES AND PLANT DEFENCE IN *EUCALYPTUS GRANDIS*

Makobatjatji Mmoledi Mphahlele

Supervised by ***Prof Alexander A. Myburg***

Co-Supervised by ***Prof Fikret Isik and Prof Gary R. Hodge***

Submitted in partial fulfilment of the requirements for the degree ***Philosophiae Doctor***

Department of Biochemistry, Genetics and Microbiology

University of Pretoria

Eucalyptus plantation forestry is a renewable feedstock for woody biomass-based products in the traditional timber, pulp and paper industries, as well as emerging biomaterials, biochemicals and bioenergy industries. Due to human population growth, pressure on the limited land for housing, agricultural produce, the risk from biotic and abiotic stressors, and climate change poses significant risks to plantation forestry. Therefore, innovative breeding approaches are needed to explore and mitigate these pressures and ensure the future sustainability of eucalypt plantation forestry. This study explored the practical application of genome-assisted breeding technologies (genomic selection, GS and single-step genomic BLUP, ssGBLUP) as replacements or complementaries to traditional *Eucalyptus grandis* tree breeding (TB) approaches. First, the study focused on the feasibility of implementing genomic BLUP approaches that maximise genetic gains per unit time for improved growth and wood quality traits. Second, the study demonstrated the utility of genomic data in multivariate

approaches to simultaneously improve diameter growth and tolerance to *Leptocybe invasa* leaf gall wasp and *Botryosphaeria dothidea* / *Teratosphaeria zuluensis* fungal stem disease complex. Third, the study investigated the additional benefits of combining genomic selection (GS) technologies with accelerated floral induction approaches such as transgrafting with *Flowering locus T* (FT) transgenic scions (GS-FT) to maximise gains per unit time, as well as cost per unit genetic gains in precocious and non-precocious *Eucalyptus* tree species.

The study demonstrates that GS strategies will improve tree breeding efficiency, resulting in higher genetic gains per unit of time for growth and wood quality traits. We also showed that a multivariate GS breeding strategy to simultaneously improve diameter growth and tolerance to pests and pathogens is practically feasible in *E. grandis*. We showed no breeding cycle time differences in their GS-FT breeding strategies by accelerating floral induction in the precocious *E. grandis* and non-precocious *E. dunnii*. Transitioning to GS or GS-FT breeding strategies from the TB strategy will improve genetic gains per unit time but at a higher cost per unit of genetic gain. In contrast, transitioning from a GS to a GS-FT breeding strategy results in improved genetic gains per unit time and favourable cost per unit genetic gain in early and late flowering species. The GS-FT strategy benefits *E. dunnii* due to the considerable reduction of its naturally longer breeding cycle time by accelerating floral induction and eliminating progeny testing. Although an essential short-term factor, the influence of genotyping cost on GS in forest tree breeding is minimal and diminishes over time when new efficient genotyping technologies with better throughput come online. An essential outcome of the study is that the often-overlooked changes to operational breeding practices are required to accelerate GS strategies to realise genome-assisted breeding benefits.

PREFACE

Eucalyptus is a genus of widely adaptable hardwood tree species planted worldwide for their renewable biomass and supports numerous wood product industries, including biomaterial, biochemical, and bioenergy products. To sustainably meet the demand for this woody biomass, tree breeders need to deploy technologies to improve genetic gains while mitigating abiotic and biotic pressures. *E. grandis* is a subtropical species grown as a hybrid partner in clonal genotypes deployed in South Africa, mainly in *E. grandis* × *E. urophylla* and *E. grandis* × *E. nitens* F1 hybrid genotypes. Over the last two decades, *E. grandis* has been on the decline due to pressure from pests and pathogens that limit its breeding and commercial deployment.

As part of the tree breeding process, genetic testing is lengthy with many breeding cycle steps, including lengthy field trials. This genetic testing process seeks to determine genetic parameters to select biotic and abiotic tolerant trees with an improved yield of economic traits. Many of these biotic and abiotic risk factors are in an arms race with tolerant trees. Therefore long breeding cycles delay selection responses from these genetic tests to mitigate these risks and ensure improved yields. Such challenges are universal to tree breeding programmes worldwide, including in South Africa. In South Africa, particularly with *E. grandis*, due to pests and pathogens and being in its 5th generation, the species is not planted commercially in the sub-tropical region, and it is a vital hybrid partner to the commercially deployed clones in the sub-tropical region. Therefore, studies of GS technology in *E. grandis* are essential to resolve both growth and wood quality yield and ensure tolerance to biotic and abiotic stresses. New technology has emerged to accelerate floral induction using *Flower Locus T* transgene via transgrafting. Such technology will ensure that the breeding cycle time for tree species is at its shortest biological limit, thereby maximising the reduction in breeding cycle time and

accelerating risk mitigation and yield improvement. These technologies will require additional investment in genotyping cost and breeding cycle adjustment due to the new accelerated floral induction process. Adopting these new technologies means that benefits and costs assessments are needed to evaluate their viability in terms of genetic gains benefit per unit time and cost of these new breeding strategies compared to the traditional strategies.

This PhD study began in 2014 and concluded in 2021. The study aimed to develop marker-aided breeding strategies for *E. grandis* to improve genetic gains and mitigate pest and pathogen pressures. The study was a collaborative effort between Mondi South Africa, Tree Improvement Research, and the Forest Molecular Genetics (FMG) Programme at the University of Pretoria, Department of Biochemistry, Genetics, and Microbiology, and North Carolina State University (NCSU), Department of Forestry and Environmental Resources. Mondi Tree Improvement Research provided the *E. grandis* GS study training population comprised of a half-sib and full-sib progeny. The half-sib *E. grandis* progeny trial for the growth and pest and pathogen tolerance study was part of the main breeding population. The FMG Programme facilitated the SNP genotyping of the study populations, with NCSU providing data analysis support for quantitative genetics and genomic selection approaches.

Chapter 1 highlights the literature about genome-wide marker analysis in tree breeding. It also covers various genotyping technologies and genome-assisted analysis platforms used to predict genetic merit and facilitate artificial selection to improve genetic gain.

Chapter 2 investigated the utility of genomic selection (GS) via genomic BLUP of seven growth and wood quality traits in an operational *E. grandis* breeding population. The GS accuracy of the traits was estimated as the ratio between the genomic estimated breeding values

(GEBV) and the pedigree estimated breeding values (EBV) of the traits. GS efficiencies was estimated by comparing the relative accuracy of a 4-year GS breeding cycle with the 8-year traditional breeding cycle. Simulated over many years, we show improved genetic gains per unit time. However, significant operational breeding steps must be adjusted to accommodate GS as a viable breeding approach.

Chapter 3 investigated how tree breeders can improve the diameter growth of *E. grandis* under *Leptocybe invasa* (*Lepto*) and *Botryosphaeria dothidea* / *Teratosphaeria zuluensis* (*BotryoTera*) fungal stem disease complex. Using single-step genomic BLUP (ssGBLUP), we obtain genetic parameters for heritabilities, multivariate and multi-environment analysis of pedigree-linked *E. grandis* families over three generations. We observed a strong genetic correlation between diameter growth and *Lepto* tolerance, which resulted in indirect improvement in *Lepto* tolerance over the generations from the direct artificial selection of diameter growth. *BotryoTera* shared a negative genetic correlation with diameter growth. Resolving this problem, we proposed a GS breeding strategy to improve diameter growth and *Lepto* and *BotryoTera* tolerance by ensuring the three traits are accurate phenotyping to train a GS model.

Chapter 4 presents results in which traditional breeding strategies of the precocious *E. grandis* and non-precocious *E. dunnii* are compared to their proposed GS breeding strategy and a proposed GS strategy with an accelerated floral induction. We modelled a scenario under which accelerated floral induction is facilitated by the transgrafting of *Flowering locus T* (FT) transgenic scion (GS-FT) onto the selected rootstock. In both species, the operational and genotyping cost of the three breeding strategies was assessed. We compared the cost ratio, genetic gain benefit ratio, and the relative genetic gains benefit to cost ratio between the two

species and the three breeding strategies. We show that adopting GS and GS-FT will increase the cost per unit genetic gain and the genetic gains benefit per unit time, however, at an unfavourable relative benefit per unit cost ratio compared to the TB strategy in both species. Adopting the GS-FT strategy over the GS strategy improves the benefit per unit time, cost, and the relative benefit per unit cost. The accelerated GS-FT strategy removes floral induction as a barrier, resulting in similar breeding cycle times and breeding steps for both species benefiting *E. dunnii* more because of a more significant time reduction in *E. dunnii* vs *E. grandis*.

Chapter 5 presents the study's overarching findings and conclusions and synthesises the study summary, its limitations, recommendations, and future research contribution and implication.

Below are the research outputs throughout the PhD study, including peer-reviewed publications in ISI-rated journals and conference presentations that have emanated from the study:

ARTICLES

Vivas, M., Kemler, M., **Mphahlele, M. M.,** Wingfields, M. J., and Slippers, B. (2017). Maternal effects on phenotype, resistance and the structuring of fungal communities in *Eucalyptus grandis*. *Environmental and Experimental Botany* 140**,** 120-127.

Naidoo, S., Christie, N., Acosta, J. J., **Mphahlele, M. M.,** Payn, K. G., Myburg, A. A., and Kulheim, C. (2018). Terpenes associated with resistance against the gall wasp, *Leptocybe invasa*, in *Eucalyptus grandis*. *Plant Cell & Environment* 41**,** 1840-1851.

Mhoswa, L., O'Neill, M. M., **Mphahlele, M. M.,** Oates, C. N., Payn, K. G., Slippers, B., Myburg, A. A., and Naidoo, S. (2020). A Genome-wide association study for resistance

to the insect pest *Leptocybe invasa* in *Eucalyptus grandis* reveals genomic regions and positional candidate defence genes. *Plant and Cell Physiology* 61**,** 1285-1296.

Mphahlele, M. M., Isik, F., Mostert-O'neill, M. M., Raynolds, S. M., Hodge, G. R., and Myburg, A. A. (2020). Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. *Tree Genetics & Genome* 16**,** 49. doi: 10.1007/s11295-11020-01443-11291.

Mphahlele, M. M., F. Isik, G R. Hodge, A. A. Myburg. (2021). Genomic breeding for diameter growth and tolerance to *Leptocybe* gall wasp and *Botryosphaeria / Teratosphaeria* fungal disease complex in *Eucalyptus grandis*. *Frontiers in Plant Science*: 12:638969. doi: 10.3389/fpls.2021.638969

Mostert-O'Neill, M. M., Reynolds, S. M., **Mphahlele, M. M.,** van den Berg, G., Verryn, S. D., Acosta, J. J., Borevitz, J. O., and Myburg, A. A. (2022). Genomic signature of artificial selection during early domestication of a wood fibre drop. *New Phytologist* (doi: 10.1111/nph.18297)

CONFERENCES (ORAL PRESENTATIONS)

Mphahlele M. M., Hodge G. R., Isik F., Myburg A. A. "Phenotype, pedigree, and genomic BLUP analysis of growth and wood property traits using single-step approaches in semi-structured Eucalyptus grandis population". 11th South African Plant Breeders Association Symposium. 08-10 March 2016. Protea Hotel Technopark, Stellenbosch. South Africa.

Mphahlele M. M., Hodge G R., Isik F., Myburg A. A. "A single-step genomic breeding approach for diameter growth and tolerance to *Leptocybe* gall wasp and *Botryosphaeria / Teratosphaeria* fungal disease complex in *Eucalyptus grandis*". 13th South African Plant Breeders Association Symposium. 08-11 March 2020. Future Africa, University of Pretoria, South Africa.

Mostert-O'Neill M. M., Borevitz J. O., Acosta J. J., Reynolds S. M., **Mphahlele M. M.,** Van den Berg G., Verryn S. D., Myburg A. A. 2020. Domesticating *Eucalyptus grandis* for

changing climates using landscape genomics. Southern African Plant Breeding Symposium, 8 – 11 March, Pretoria, South Africa

Reynolds S. M., Tii-Kuzu Y. J., Ranade S. S., O'Neill M. M., Henning J. S., Swain T. L., Meyer B., Verryn S. D., **Mphahlele M. M.**, Van Deventer F., Van der Hoef A., Nel A., Le Roux T., De Waal L., Acosta J. J., Isik F., Hodge G. R., Payn K. G., Kanzler A, Naidoo S., Myburg A. A.. 2017. The development of a genomic diversity atlas for pine and eucalypt species of importance to the South African forestry industry. South African Forestry Research Seventh Forest Science Symposium, 18 – 20 July, Pietermaritzburg, South Africa

CONFERENCES (POSTER PRESENTATIONS)

Mphahlele M. M., Pienaar B., Du Plessis M., van Deventer F., Reynolds S.M., O'Neill M.M., Hodge G.R., Isik F. and Myburg A.A. "Opportunities and challenges of genomic selection in *Eucalyptus grandis* breeding. 6th ICFR Forest Science Symposium. 29-30 July 2014. Hilton College Theatre, Hilton College School, South Africa.

Mphahlele M. M., van Deventer F., Reynolds S. M., O'Neill M. M., Hodge G. R., Isik F. and Myburg A. A. "Genotyping and phenotyping of a *Eucalyptus grandis* training population towards the development of genomic selection models". IUFRO Tree Biotechnology Conference. 08-12 June 2015. Florence, Italy.

Mphahlele M.M., Pienaar B., Du Plessis M., van Deventer F., Reynolds S. M., O'Neill M. M., Hodge G. R., Isik F. and Myburg A. A. "Opportunities and challenges of genomic selection in *Eucalyptus grandis* breeding. 7th ICFR symposium 18-20 July 2017. One Life Church, Alexandra Road, Pietermaritzburg, South Africa.

Mhoswa L., O'Neill M. M., **Mphahlele M. M.,** Oates C., Payn K. G., Slippers B., Myburg A. A., Naidoo S. 2018. A genome-wide association study for tolerance to the insect pest Leptocybe invasa in *Eucalyptus grandis* reveals genomic regions and positional candidate defence genes. South African Society for Bioinformatics & South African Genetics Society (SASBi-SAGS) Joint Conference, 16 – 18 October, Golden Gate Highlands National Park, South Africa

O'Neill M. M., Reynolds S. M., **Mphahlele M. M.,** Louw A., Galloway G., Jones N., Kanzler A., Du Plessis M., Pienaar B., Myburg A. A. 2014. High-throughput SNP marker genotyping for genomic selection of growth and wood properties in *Eucalyptus*. Sixth Forest Science Symposium, 29 – 30 July, Pietermaritzburg, South Africa

ACKNOWLEDGEMENT

I want to express my deepest gratitude to the following people and organisations:

My supervisor, Professor Alexander A. Myburg, you have been part of my research career from undergrad. Your mentorship, guidance and supervision were exemplary, patient and motivating. Your influence in this PhD study and in my career is unmatched. I will forever be eternally grateful.

My co-supervisor, Professor Fikret Isik, you have been the source of excellent guidance in the data analysis of this PhD study. Your efforts to meticulously and clearly guide me in both R and ASReml platform is appreciated. Your co-authored book “Genetic Data Analysis for Plant and Animal Breeding, ISBN 978-3-319-55175-3” was invaluable to this PhD study. The skills you have taught me continue to make me a better tree breeder. Thank you.

My co-supervisor, Professor Gary R. Hodge, you have brought reason and connectedness to the study. Your guidance helped with interpreting the study and give the relationship of the study objectives to the practice of tree breeding. Your guidance was invaluable and gave meaning to the rationale of this study

To the Forest Molecular Genetics programme team, namely Mellissa Raynolds and Marja O’Neill-Mostert, your expertise on genotyping platforms were crucial to the study. All the genotyping used in the study found its way through your hands. Thank you. It is greatly

appreciated. To the rest of the team, your support, skills and the shared happy moments are a source of pride and joy. Thank you.

The Mondi South Africa, Research & Development team, namely Francois van Deventer, your institutional memory was crucial in establishing the breeding population and providing the pedigree information and their historical context. Thank you. Siyabonga Goodenough Dube, your hard work and determination is appreciated. Thank you for all the long hours and walks in the plantations doing the wood phenotyping and disease scoring. I would also like to thank Mondi South Africa Forest – Learning and Development for providing the necessary funding support to pursue this PhD study, namely Kerry Davies. Your belief in me is appreciated.

Thank you for the love and support to my parent, Thakamanyana Wilson Mphahlele and Rematsimela Windy Mphahlele. You are a source of encouragement and motivation throughout my life and PhD. I love you dearly. To my sibling Thato, Thuso and Mmatebogo Mphahlele, thank you for being a source of wonderful cheers, support and loving memories throughout my PhD.

To my lovely Wife, Reratilwe Ephenia Mmatshoene Mphahlele, with your love, support and encouragement, this PhD was as sure as the sunrise. You were the pillar of my motivations and efforts. You managed to support me during my PhD while you were busy with yours whilst also gave birth and raised our two beautiful kids Kgadi and Reagile Mphahlele. I drew from such strength you have demonstrated. What a wonder you are in my life. I love you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ABLUP	Pedigree relationship matrix BLUP
BLUP	Best linear unbiased perdition
BV	Breeding value
DGV	Direct genetic value
EBV	Empirical breeding value
FT	Flowering locus T
GBLUP	Genomic relationship matrix BLUP
GBS	Genotype-by-sequencing
GEBV	Genomic estimated breeding values
GS	Genomic selection
GS-FT	Genomic selection with FT floral induction
LD	Linkage disequilibrium
MAS	Marker assisted selection
NGS	Next generation sequencing
QTL	Quantitative trait loci
SNP	Single nucleotide polymorphism
ssGBLUP	Single-step genomic BLUP
TB	Traditional breeding

CHAPTER 1

LITERATURE REVIEW:

MOLECULAR MARKER TECHNOLOGIES AND TREE IMPROVEMENT

1.1 INTRODUCTION

Plants are the most significant contributors to permanent carbon biomass storage on Earth (~450 gigatons), mainly in the form of woody biomass (Bar-On et al., 2018). Over 4 billion hectares of terrestrial coverage globally have captured this woody biomass, with 0.9 billion hectares available for forest and woodlands restoration efforts (Bastin et al., 2019). Ecologically, forests play a role in preserving biodiversity, climate and water quality regulation. Forest plantations (over 277 million hectares globally) play a significant role in mitigating the demand pressures on native forests (Payn et al., 2015). Industrially, plantation forestry is vital for woody biomass biomaterials such as pulp and paper products, building materials, textile fibres, firewood, bio-energy and industrial chemicals (Perlack et al., 2005; Cetinkol et al., 2012; Devappa et al., 2015; Stafford et al., 2020). Sustainability in woody biomass production from plantation forests is vital, especially with increased pressure from the human population, demanding land and woody biomass-based industrial products, and increased biotic and abiotic pressures (Wingfield et al., 2015). Forest tree improvement efforts are crucial to mitigate such biotic and abiotic pressures and increase yields to meet these demands for woody biomass.

Tree breeding applies genetic principles of reproductive biology to economic value traits to maximise and sustain their genetic improvement. While some species in forest tree breeding have advanced to the second and third generations, some species are in their infancy in that the original geographic patterns (provenance) of genetic variation and adaptation and new infusions are still the subjects of study in provenance progeny trials. Traditional tree breeding approaches are subject to several challenges, such as long breeding cycles due to time to reproductive maturity and maturation of interest traits. Alleviating the reproductive challenges

may involve floral induction chemicals to improve precocity and placing selections in favourable environmental conditions (Griffin et al., 1993; Williams et al., 2003; Gardner and Bertling, 2005). Regardless of maturation, the selection of traits may be mitigated with molecular markers associated with the target traits (Grattapaglia, 2004; Grattapaglia and Kirst, 2008). Molecular markers identify DNA polymorphism that explains the genetic architecture of valuable traits and provides targets for forest molecular tree improvement strategies and opportunities to study the genetic diversity of traits and breeding populations (Grattapaglia et al., 2009, 2011, 2018).

Initial tree improvement strategies involving DNA markers linked phenotypic and genotypic data to mapped quantitative trait locus (QTL) of essential traits in trees. These traits included growth, wood quality, pests and pathogen resistance, abiotic and abiotic stress tolerance (Grattapaglia et al., 1995, 1996; Freeman et al., 2009) and assessing population diversity (Brondani et al., 1998). QTL regions are not synonymous with their causative polymorphisms, and consequently, their genetic basis is unclear, given the sophisticated genetic control of quantitative traits (Strauss et al., 1992). The inability of QTL studies to reveal the exact causative loci for a trait variation led researchers to an alternative approach in genome-wide genetic markers to tag each causative genetic variant. Genomic selection (GS), as introduced by Meuwissen et al. (2001), is an approach that predicts the genetic merit of individuals based on their aggregate genomic-wide genetic information associated with phenotypic information. In practice, GS uses whole-genome regression models of phenotypic information with genome-wide marker effect covariates to predict genetic merit in the form of genomic estimated breeding values (GEBV; Meuwissen et al., 2001).

Genomic selection can increase genetic gains per unit time by enabling the selection of difficult-to-measure and maturity-dependent traits earlier in the breeding cycle. These obstacles are particularly relevant in forest tree breeding since their expression of traits depends on planting trials over several years. GS is of particular benefit in forest trees due to their long breeding cycle times due to delayed reproductive maturity and the need for early selection of maturity-linked growth and wood quality traits (Grattapaglia et al., 2011). The availability of the GEBV for trees early in the breeding cycle has meant the adjustment of the traditional breeding (TB) cycle into the GS breeding cycle to achieve the requisite gains per unit time, demonstrated in crops (Bernardo and Yu, 2007; Bassi et al., 2016; Nyouma et al., 2019) and forest trees (Bartholome et al., 2016; Resende et al., 2017; Li and Dungey, 2018).

The genus *Eucalyptus*, mainly native to Australia, is host to crucial species to plantation forestry. These species possess environmental durability (Hughes et al., 1996) and have relatively fast growth with favourable wood qualities, which has led to their distribution in plantations on almost every continent. The "Big Nine" eucalypt species, *E. camaldulensis*, *E. grandis*, *E. tereticornis*, *E. globulus*, *E. nitens*, *E. urophylla*, *E. saligna*, *E. dunnii*, and *E. pellita*, and their hybrids constitute more than 90% of the *Eucalyptus* plantation forest area (Stanturf et al., 2013). *E. grandis*, the first *Eucalyptus* species to have its genome published, has demonstrated its research importance for global renewable energy and wood fibre resource (Myburg et al., 2014).

This review outlines some of the challenges and successes of applying genome-wide marker technologies in forestry tree breeding, focusing on how TB strategies may be adjusted and

complemented with GS breeding strategies and their plausible adoption and anticipated benefits.

1.2 MOLECULAR MARKER TECHNOLOGIES APPLIED IN TREES

Molecular marker technologies have complemented breeding strategies by providing a comprehensive understanding of molecular diversity and, consequently, molecular plant breeding, GS, molecular genetics and genome editing opportunities (Nadeem et al., 2018). Capturing all rare alleles with important additive and dominance effects on a population-wide scale requires genetic analysis of large mating designs or the development of high-density marker platforms to cover the genome across populations (Elbasyoni et al., 2018). The development of high-throughput genome-wide genotyping SNP marker technologies in structured *Eucalyptus* species (Silva-Junior et al., 2015b) and their associated application strategies have revolutionised tree breeding approaches with better estimations of genetic parameters, genotype-trait associations and accurate prediction of genetic merit (Resende et al., 2012).

In unstructured and complex populations, next-generation sequencing (NGS) technologies can simultaneously facilitate the discovery and perform genome-wide genotyping of SNP markers (Nadeem et al., 2018). NGS technologies allow for the resequencing of many related pooled plant genomes more cost-effectively with greater depth and reliability (Varshney et al., 2009; Metzker, 2010). Adopting genotyping-by-sequencing (GBS) technologies that rest on NGS capacity, identifying and tracking genetic variation, has become more efficient and precise within large populations and reduces nucleotide sequencing costs (He et al., 2014). GBS information has an additional advantage over SNP array data in allowing further analyses of

fine-tuned genome-wide association (GWAS) studies, genome diversity studies, genetic linkage analysis and molecular marker discoveries (He et al., 2014). GBS discovers SNPs in crop systems of varying genome sizes, breeding systems, and species with or without reference genomes (Kim et al., 2016). In a wheat population, GBS enabled the differentiation of genetic diversity between landraces and cultivars, partitioning them into subgroupings related to the year of selection and climatic zones (Alipour et al., 2017). GBS also demonstrated genetic variation and population genetic structure patterns under selection associated with ecological and diseased regions in flowering dogwood (Pais et al., 2020). Lodgepole pine (*Pinus contorta*) and white spruce (*Picea glauca*) are conifer species with large genomes, and GBS demonstrated itself as a suitable genotyping technology because of their unreferenced large, highly heterozygous genomes (Chen et al., 2013). Compared to SNP array data, GBS is the preferred marker platform for studying genetic diversity and accelerating breeding via GS in winter wheat (Elbasyoni et al., 2018). A genomic prediction study in a full-sib Scots pine (*Pinus sylvestris* L.) population using GBS presents similar prediction accuracy for both pedigree and genomic models, but the genomic models had higher selection efficiency (Calleja-Rodriguez et al., 2020). High throughput phenotypic technology combined with GBS in a GS study resulted in improved genetic gains for wheat (Crain et al., 2018). The robustness of an innovative genotyping platform such as GBS with high-throughput phenotypic can add value to GS study models, particularly in marginal species and unstructured populations without reference genomes.

Exome capture technology involves using biotinylated oligonucleotide probes to capture genomic DNA of protein-coding regions by hybridisation. The captured genomic DNA can be sequenced, targeting only the protein-coding portion, thereby significantly reducing resources (Choi et al., 2009). Exome capture sequencing platform was applied to discover and develop

SNPs resources for application in GWAS or GS studies in the absence of complete reference genomes in conifers like black spruce (*Picea mariana*; Pavy et al., 2016), radiata pine (*Pinus radiate*; Telfer et al., 2019) and Norway spruce (*Picea abies*; Chen et al., 2018). The exome capture sequencing platform discovers single nucleotide variants and indels in wood formation genes across the three eucalyptus species, *E. tereticornis*, *E. camaldulensis* and *E. grandis* (Dasgupta et al., 2015). It is valuable to capture genetic variation in the genome coding regions because it targets most of the causal gene variants. The approach may be advantageous to certain traits and limited to other traits depending on genetic architecture and their regulatory mechanism.

One of the first genome-wide marker platforms in *Eucalyptus* was the Diversity Array Technology (DArT), which generated moderate-density dominant polymorphic genome-wide markers (Sansaloni et al., 2010). Silva-Junior et al. (2015b) development of a high-density genome-wide genotyping SNP chip (EucHIP60K.Br) pooled DNA from multiple commercially important *Eucalyptus* species, generating approximately 64,000 informative SNP markers. The EucHIP60K.Br SNP chip is more informative and quantitative than microsatellites by allowing the calculation of a pairwise relationship matrix of the entire population (Telfer et al., 2015). The genotypic data generated from this genome-wide SNP chip expands our knowledge of the evolution and adaptation of *Eucalyptus* species such as *E. grandis* (Silva-Junior and Grattapaglia, 2015a) and basic biology of woody and tree health in perennial species such as eucalypts (Resende et al., 2016). The application includes ancestral mapping of *E. grandis*, identifying interspecific introgression associated with aridity and marker enriched regions for biotic and abiotic stress responses (Mostert-O'Neill et al., 2020) and consequences of artificial selection (Mostert-O'Neill et al., 2022).

1.3 DEVELOPMENT AND IMPLEMENTATION OF GENOMIC SELECTION STRATEGIES TO ACCELERATE TREE IMPROVEMENT

The outcrossing nature of *Eucalyptus* species, their large effective population sizes and broad geographic distribution predict plausible maintenance of low-frequency rare mutant alleles in the population over generations, mainly if they are selectively neutral (Grattapaglia and Kirst, 2008). Genetic and phenotypic diversity of *Eucalyptus* in natural or breeding populations is likely to contain an abundance of unique haplotypes and low-frequency alleles, some of which are what tree breeders seek when using marker-assisted selection (MAS) because some superior phenotypes in these populations may be attributable to such uncommon allelic variants (Ballesta et al., 2019; 2020). Genome-wide markers have led to methodologies of marker-trait association, which circumvent some of the limitations of QTL mapping. Genome-wide DNA marker-based selection approaches offer increased efficiency and accuracy in tree breeding strategies. These benefits are achieved by mapping, tagging, detecting, and quantifying the genetic linkages associated with the quantitative traits underlying causative allelic loci. Many small-effect QTLs rather than a few large-effect QTLs govern quantitative trait expression (Fisher, 1918). Hall et al., 2016, showed that the detection of large-effect QTLs was attributed to the studies of small population sizes. QTL studies in their design are biased toward highlighting large-effect QTLs against the small-effect QTL, a condition known as the "Beavis effect" (Beavis, 1998; Xu, 2003). In practice, the information from DNA markers (a) evaluate genetic diversity, (b) track genetic variants, and (c) facilitate the pyramid genome regions (Collard and Mackill, 2008). However, constraints exist in translating these uses into economic outputs due to the transferability of QTL across populations, genetic backgrounds and experimental designs (Strauss et al., 1992; Holland, 2004). In contrast, genome-wide genetic selection approaches use sophisticated mathematical and genetic models based on genome-

wide DNA markers to capture both big and small-effect QTL, linkage disequilibrium (LD) and genetic relatedness to predict genetic merits in GS modelling (Meuwissen et al., 2001). Other constraints include translating genotype-by-environment interactions, allele frequency differences and the application gaps between the researchers (molecular biologists) and the implementation strategists (tree breeders; Collard and Mackill, 2008).

The aggregation of SNP markers into haplotypes has additional benefits to the GS model by improving genomic prediction of low-heritability traits observed in *E. globulus* (Ballesta et al., 2019). Also, Bayesian models aggregate the heterogeneous variances and covariance structure of SNP markers across the genome using single-step genomic predictions to estimate variance components (Karama et al., 2020). The nature of field tests in tree breeding strategies means that tens of thousands of trees need to be planted, and if GS model approaches are adopted, it would be cost-prohibitive. Single-step genomic (ssG)BLUP analysis overcomes this limitation by blending the breeding population pedigree with the genomic relationship matrix of a subset of genotyped individuals (Legarra et al., 2009; Misztal et al., 2009; Aguilar et al., 2010; Christensen, 2012). Therefore, ssGBLUP analysis extends GS application to non-genotyped individuals (Legarra et al., 2014), allowing for multivariate and univariate analyses (Guo et al., 2014). ssGBLUP approach has improved genomic prediction efficiency in trees (Ratcliffe et al., 2017; Cappa et al., 2019; Imai et al., 2019). ssGBLUP approach requires a well-curated pedigree to ensure improved accuracy in its estimation of genetic parameters. The proportion of selfs in traditional pedigree records is one component never reported. Sib-ship reconstruction is an approach to identify the selfing proportion in genotyped siblings. Sib-ship reconstruction was demonstrated in *E. nitens* resulting in a selfing rate of 4% (Klapste et al., 2017). The rescaling of the input genomic relationship matrix in a ssGBLUP with this hidden selfing resulting in improved accuracy, particularly for inbreeding depression-sensitive traits such as

growth (Klapste et al., 2018). The ability to seamlessly integrate genomic information with traditional BLUP models has added to the value of adopting marker technologies into mainstream tree improvement efforts. The above technological platforms continue to shape much of molecular breeding advancements in forest tree improvement and GS model applications.

Genomic selection (GS) is a molecular breeding approach that estimates the aggregate effects of genome-wide polymorphic markers on measured phenotypes to estimate GEBV by exploiting genome-wide LD of these markers with QTL regions of the trait of interest (Meuwissen et al., 2001). Here QTL refers to any polymorphism that affects the trait of interest, regardless of the size of the effect and the ability to detect the QTL using genetic mapping or genome-wide association approaches. Genomic selection aligns with the infinitesimal model in that infinite small-effect loci rather than finite large-effect loci govern quantitative traits (Jannink et al., 2010b). Ridge regression models in genomic selection deal with the problem of a small number of observations and a large number of predictive variables (small p large n) to avoid multicollinearity in predictions to improve accuracies (de los Campos et al., 2013). They also resolved genomic selection accuracy when traits are controlled by many QTLs, leading to the genomic relationship estimated effects. In contrast, Bayesian models provide favourable accuracies for traits influenced by fewer QTL leading to marker-QTL LD estimated effects (Daetwyler et al., 2010). With strong marker-QTL LD, Bayesian model prediction accuracy would remain useful over generations but reduce with LD breakdown (Habier et al., 2007). Response to GS can fall short of phenotypic selection over generations because GS does not "discover" QTL, and selection response can drift if markers and QTL are not in strong LD. Therefore, fixing a marker effect will not necessarily fix the QTL. This may lead to a rapid decline in GS accuracies (Muir, 2007; Goddard, 2009; Rodriguez-Ramilo et al., 2015).

Reducing the number of markers results in reduced GS accuracies, particularly if the markers are not preselected based on their association with the trait of interest (Cleveland et al., 2010). Genomic selection models that assume equal marker effect variance produce better prediction accuracies than models that assume unequal marker effect variance when trait-linked markers are used (Zhang et al., 2010; Wang et al., 2012). Strong marker-QTL LD and genetic relationship strengthen the accuracy of GEBV predictions (Habier et al., 2007; Legarra et al., 2009; Forni et al., 2011). The challenge is to design a breeding strategy that integrates all of this information into an effective tree breeding program. It is plausible to expect that GS models that use stable trait-linked markers that affect traits in the same direction across populations and environments could produce better accuracies regardless of the GS model parameters. However, this approach is not always practical since environmental changes and genetic infusions are vital factors affecting tree breeding strategies and objectives, hence the need to adopt genome-wide DNA marker platforms.

Genomic selection strategies can maintain genetic diversity while increasing selection gains by managing inbreeding by reducing sibling co-selection (Daetwyler et al., 2007; Rodriguez-Ramilo et al., 2015). Response to GS shows that long-term gains are possible by managing genetic diversity of markers-QTL LD (Li et al., 2008), weighing favourably low-frequency marker alleles and adding more favourable marker alleles (Goddard, 2009; Jannink, 2010a) and managing inbreeding in the context of GEBV and not pedigree-based breeding value (Sonesson et al., 2012). The development and implementation of GS models in tree breeding must consider many factors related to the genetic parameters underlining the breeding objectives and the constraints within breeding operations and biological limits. GS challenges in tree breeding are the practical integration and adoption of GS models for a single trait or multiple trait selection under species-specific propagation and reproductive constraints, linked

operational-breeding steps and the commercially deployed or breeding product. There is a need for GS strategies to adapt and accommodate all these breeding and deployment strategies faced by tree breeders.

1.4 *EUCALYPTUS* FOREST PLANTATIONS AND BREEDING

The genus *Eucalyptus* comprises approximately 900 species, which are primarily endemic to the Australian continent (Brooker, 2000). Few of these species and their hybrids constitute commercial plantation forestry for wood fibre production. Wood fibre production has the potential to resource emerging industries of bioenergy and biomaterials, in addition to the traditional industries of pulp, paper, and timber production (Perlack et al., 2005), increasing and sustaining the demand for plantation forestry wood fibre. Phenotypic selections based on combined tree growth improvements, wood properties traits and tolerance to biotic and abiotic stresses have supported tree breeding programs for many years (Namkoong et al., 1980). The challenge is that growth has lower heritability than wood property traits, which are difficult and expensive to measure in trials (Raymond, 2002). However, both traits are essential determinants of forest plantation yields (Wimmer et al., 2002). The accuracy and reliability of such phenotypic selections and genetic improvements are dependent on optimally designed genetic field trials (Williams et al., 1999). Family-based selection approaches serve as the primary techniques for most forest genetic trials and continue to be practised to date (White and Hodge, 1988). The family-based selection approach used by tree breeders entails the accurate estimation of the actual genetic value or breeding value (BV) using their progeny for within and between families to effect forward and backward selections, respectively. Improved statistical equations such as Best Linear Unbiased Prediction (BLUP) have become the standard approach in tree breeding for estimating BV (Henderson, 1975; White and Hodge, 1988; Piepho et al., 2008). Multiple environment tests, accurate phenotypes, the known genetic

relationship amongst the trees mainly derived from half-sib or full-sib pedigree information, and established genetic parameters of traits of interest are information available to tree breeders to strengthen BLUP analysis to estimate BV.

A review of genomic selection as a tool for forest tree improvement was conducted by Grattapaglia et al. (2011), focusing on its influential factors, such as the number of markers, the size of the training population, and the genetic architecture of the target traits. Since then, the review has focused on GS's future research prospects, including its implementation strategies, validation and model updates, functional validation, and multi-environmental forecasting in response to climate change (Grattapaglia et al., 2018). Genomic selection studies have shown empirical results on their viable implementation in crop breeding systems (Zhong et al., 2009; Desta and Ortiz, 2014; Lin et al., 2014; Sallam et al., 2015; Crossa et al., 2017; Haile et al., 2018; Wang et al., 2018), conifer breeding system, (Resende Jr et al., 2012b; 2012a; Zapata-Valenzuela et al., 2013; Isik et al., 2015; Bartholome et al., 2016; EL-Dien et al., 2016; Duran et al., 2017), *Eucalyptus* breeding system (Zelener et al., 2005; Grattapaglia et al., 2011; Denis and Bouvet, 2013; Rambolarimanana et al., 2018; Suontama et al., 2019), as well as in hybrid breeding system in crops (Zhao et al., 2015) and forest trees (Tan et al., 2018).

GS studies in forest species have demonstrated acceptable prediction accuracies for adoption in pine species such as *Pinus taeda* (Resende Jr et al., 2012b; Zapata-Valenzuela et al., 2013), *Picea glauca* (Beaulieu et al., 2014) and *Pinus pinaster* (Isik et al., 2015; Bartholome et al., 2016). GS models resulted in prediction accuracy three-fold better than pedigree-based approaches and improved gains per unit time for growth and wood quality traits in black spruce (*Picea mariana*; Lenz et al., 2017). The genome sequencing of *E. grandis* (Myburg et al., 2014)

and the development of a robust *Eucalyptus* SNP (EUChip60K chip) platform for high-throughput SNP genotyping (Silva-Junior et al., 2015b) have created opportunities for studying and applying GS in *Eucalyptus* species. Genetic merit predictions using GS approaches has revised specific TB approaches in interspecific *Eucalyptus* hybrids (Resende et al., 2012; Tan et al., 2017), *Eucalyptus pellita* and *E. benthamii* breeding populations (Müller et al., 2017), *E. nitens* seed orchard population (Suontama et al., 2019) and elite clonal selections of *E. globulus* (Duran et al., 2017). Multi-trait genomic BLUP demonstrated improved growth and wood quality traits in an *E. robusta* breeding program (Rambolarimanana et al., 2018). In contrast, genomic BLUP in *E. nitens* improved solid wood gains compared to pedigree-based BLUP (Suontama et al., 2019). Genetic parameters are essential to maximising gains with directional selection, particularly with quantitative traits driven by additive genetic effects. Genetic markers are more precise than pedigree in dissecting these additive genetic effects from the non-additive genetic effects, therefore, maximising genetic yields in *Eucalyptus* hybrids (de Lima et al., 2019).

One of the challenges in deploying GS breeding strategies is desirable prediction accuracy, cost-efficiency, accurate phenotyping and robust genotyping platform targeted at a relatively small structured training population (Akdemir and Isidro-Sanchez, 2019). Genomic selection accuracies diminish with reducing relatedness in the breeding population from full-sibs, half-sibs and unrelated families, as observed in Norway spruce (*Picea abies*; Lenz et al., 2019) and black spruce (*Picea mariana*; Lenz et al., 2017). Increasing the number of markers improved the accuracy between full and half-sibs. Deep full-sib pedigree model has the potential to surpass the predictive ability of the DNA marker GS model, as observed in maritime pine over three generations (Bartholome et al., 2016). However, GS models will reduce the pedigree error rate and remove progeny testing by predicting superior genotypes from seedlings, thereby

reducing the breeding cycle. In Norway spruce (*Picea abies* (L.) Karst.), the pedigree BLUP predictive ability is 11-14% higher compared to the GBLUP predictive ability for tree height and wood quality traits (Chen et al., 2018). DNA markers that track pedigree are of little use in full-sib training populations. This was observed in Douglas-fir, in which cross-generational GS accuracy could not surpass a pedigree BLUP approach if enough markers captured the LD (Thistlethwaite et al., 2019). A genome-wide marker-based tree improvement approach complements the advancement of deep and structured full-sib pedigree.

The long-lived nature of forest trees requires long-term forest health management and protection. Forest health is crucial to the global ecosystem and the mitigation of climate change, increasing the risk of exotic and native pests and pathogens and other abiotic risk factors on plantation forests (Allen et al., 2010). The supply of wood fibre biomass to meet the increasing and sustained demand is under constant threat from the ever-increasing downward pressures of pests and pathogens (Wingfield et al., 2008) and changes in the traditional productive geographical region due to climate change (Irland et al., 2001; Booth, 2013). To overcome such downward pressures, advancements including tree breeding, cost-effective silviculture, harvesting and logistical practices (de Toit et al., 2010), and improved forest health (Wingfield et al., 2013; 2015) have thus far ensured a sustained supply of plantation wood fibre biomass to meet current demands.

Breeding for disease resistance is as essential as yield gains because the trees need to survive and thrive before any genetic potential in yield can be realised. Breeding for biotic resistance may involve two approaches; (a) qualitative resistance, in which susceptible individuals are weeded out early in the breeding cycle stages, and (b) quantitative resistance, which entails an

incremental improvement in resistance over generations. A marker-assisted selection approach has provided significant advances in breeding for disease resistance (Mamani et al., 2010). However, the approaches were biased toward identifying dominant resistance genes sensitive to failure under slight changes affecting host and pathogen. In contrast, quantitative resistance breeding is more stable but difficult and time-consuming to achieve (Silva et al., 2013). The latter approach is more suited for the GS model. GS can combine pest and pathogen tolerance breeding with increased yield objectives and is an appealing prospect for breeders (Poland, 2016). Many orthologues putative pathogenicity-related (PR) proteins are identified in *Eucalyptus* responsible for defence resistance (Naidoo et al., 2014). Their discovery is a start to achieving workable breeding strategies to improve resistance and tolerance in *Eucalyptus* forest plantations (Naidoo et al., 2019).

Phenotyping disease tolerance or resistance is challenging because it presents biological and temporal variation, thereby affecting the genetic signal reliability. A GS study for tolerance to *Dothistroma septosporum* in *Pinus radiata* resulted in favourable accuracy compared to a pedigree-based approach. (Klapste et al., 2020). The scoring of pests and pathogens can be notoriously categorical. However, Bayesian GS model approaches on categorical data of *Puccinia psidii* resulted in favourable prediction accuracy in *E. urophylla* tolerance (Silveira et al., 2019). Whereas a multivariate GS study in Norway spruce (*Picea abies* (L.) Karst) showed improved tolerance to its native white pine weevil (*Pissodes strobi* Peck) as well as growth and wood traits incorporated in the GS index model (Lenz et al., 2019). There is an opportunity to adopt GS breeding strategies, in which multivariate breeding objectives of improving growth and wood traits and pests and pathogen tolerance are viable.

There are many factors required to improve the accuracy of GS breeding strategies. They increase the training population size that is clonally tested in multiple environments, the accurate measurement of the target traits and the genetic architecture of such traits, and the study population genetic structure. The primary deliverable for tree breeders is to translate the GS model outputs into actual genetic gains and improved investment returns over time. Practising tree breeders are mandated to maintain genetic diversity and improve additive genetic gains while simultaneously developing commercial deployment products in the form of seeds or cloned varieties. Therefore, choosing a *Eucalyptus* species in which both the breeding objectives and commercial product development integrate within the GS breeding cycle strategy will ensure maximum investment returns. Focusing on *E. grandis* breeding, improving its genetic gains per unit time and improving its genetic gains per unit investment, tree breeders would need to improve their precision in the breeding cycle time management. They will need reliable control over biological processes such as reproductive and propagation mechanisms and investment in specialised infrastructure and precision equipment, such as SNP chips, phenotyping platforms, accurate environmental measurements and sophisticated mathematical models. *E. grandis* can be relatively easily propagated vegetatively. It is precocious and has a sequenced genome with a robust SNP chip. It can be deployed commercially as seeds or clonal variety and become a hybrid parent to many other species, making it a favourable candidate for adopting GS breeding strategies.

1.5 CONCLUSION REMARKS AND FUTURE PERSPECTIVE

There are many modelling approaches adopted in GS studies. These models primarily seek to understand better the genetic model of the target traits and the studied population. The target trait genetic models may be either dominant, additive, or epistatic, influenced by either cis or trans-acting regulatory elements. On the other hand, GS models also seek to understand the

genetic architecture of the study population, in the form of population structure, life history depicted in its origin and pedigrees, its genetic linkage patterns and genetic diversity (at the SNP or haplotype level) and it's reproductive and propagation characteristics. The other components that GS models seek to understand are the interaction of these genetic and population factors with the environment and their translation into the expressed phenotypes that can be accurately measured. Many statistical GS models aggregate some of the above factors at varying degrees of complexity. The less complicated and more reproducible the models and more closely aligned with the infield decision while minimising time and investment will result in effective and efficient adoption by tree breeders.

A study on white spruce (*Picea glauca*) involving multiple traditional and GS deployment strategies showed that the GS strategies had higher expected land valuation based on site index (Chamberland et al., 2020). There is an obvious attraction to understanding the impact of GS model improvements in the forest value chain. There are as many questions to answer with GS strategies on their investment value to the forest value chain expanding from the predicted GEBV. Other outstanding questions relate to the outcomes of GS models in *Eucalyptus* breeding cycles, given their many biological characteristics and associated operational breeding steps. Maximising the actual gains per unit time while minimising the investment per unit genetic gains is essential for adopting GS as an alternative to current TB systems.

The expressed phenotypes are crucial to the decisions making process by tree breeders, including their genetic regulatory systems. There is a gap in studying GS model aggregation with genetic regulatory mechanisms of target traits. Genetic regulatory mechanisms represent a coalescing of genetic control, functions and responses to expressed phenotypes (Bastiaanse

et al., 2019; Amaral et al., 2020). Therefore, multiple trait selection strategies, within and between-family information, multi-environmental testing and the time efficiency of the selection, and factors influencing the expressed phenotypes are essential considerations to achieve yield gains using tree-breeding principles. Also, how to update the GS models under these complex factors.

Conventional tree breeding in forestry has contributed significantly to the recent domestication and yield gains in plantation *Eucalypts*. Fundamental to this continual improvement is the vast amount of genetic variation in *Eucalyptus* species (Silva-Junior et al., 2015b). Adequate progeny trials and trial designs with genetic and statistical theory and pedigree information greatly influence tree breeders' selection decisions. This PhD study interrogates GS in practice in *E. grandis* breeding populations. (a) The study integrates traditional breeding approaches with GS breeding approaches to maximise gains per unit of time for growth and wood quality traits. (b) The study further performed a multivariate genomic BLUP and proposed a GS breeding strategy to improve growth and pest and pathogen tolerance in *E. grandis*. (c) Finally, the study models GS breeding strategies combined with accelerated floral induction technology to improve genetic gains per unit time and the relative cost-efficiencies per unit genetic gain in precocious and non-precocious *Eucalyptus* species.

1.6 REFERENCES

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: A unified approach to utilise phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 92**,** 743-752.

Akdemir, D., and Isidro-Sanchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. *Scientific Report* 9**,** 1446.

Alipour, H., Bihamta, M. R., Mohammadi, V., Peyghambari, S. A., Bai, G., and Zhang, G. (2017). Genotyping-by-Sequencing (GBS) revealed molecular genetic diversity of Iranian wheat landraces and cultivars. *Frontiers in Plant Science* 8**,** 1293.

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., Mcdowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. H., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J-H., Allard, G., Running, S. W., Semerci, A., and Cobb, N. (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* 259**,** 660-684.

Amaral, J., Ribeyre, Z., Vigneaud, J., Sow, M. D., Fichot, R., Messier, C., Pinto, G., Nolet, P., and Maury, S. (2020). Advances and promises of epigenetics for forest trees. *Forests* 11**,** 976.

Ballesta, P., Bush, D., Silva, F. F., and Mora, F. (2020) Genomic predictions using low-density SNP markers, pedigree and GWAS information: A case study with the non-model species *Eucalyptus cladocalyx*. *Plants* 9, 99

Ballesta, P., Maldonado, C., Perez-Rodriguez, P., and Mora, F. (2019). SNP and haplotype-based genomic selection of quantitative traits in *Eucalyptus globulus*. *Plants* 8, 331.

Bar-On, Y. M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. *PNAS* 115**,** 6506-6511.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science* 242**,** 23-36.

Bastiaanse, H., Zinkgraf, M., Canning, C., Tsai, H., Lieberman, M., Camai, L., Henry, I., and Groover, A. (2019). A comprehensive genomic scan reveals gene dosage balance impacts on quantitative traits in *Populus* trees. *PNAS* 116**,** 13690-13699.

Bastin, J-F., Finegold, Y., Garcia, C., Mollicone, D., Rezende, M., Routh, D., Zohner, C. M., and Crowther, T. W. (2019). The global tree restoration potential. *Science* 365**,** 76-79.

Bartholome, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., and Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. *BCM Genomics* 17**,** 604.

Beaulieu, J., Doerksen, T., Clement, S., Mackay, J., and Bousquet, J. (2014). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* 113**,** 343-352.

Beavis, W. D. (1998). *QTL analyses: power, precision, and accuracy in molecular dissection of complex traits.* Cleveland. Boca Raton, FL: CRC Press.

Bernardo, R., and Yu, J. (2007). Prospects for genome-wide selection for quantitative traits in maise. *Crop Science* 47**,** 1082-1090.

Booth, T. H. (2013). Eucalypt plantations and climate change. *Forest Ecology and Management* 301**,** 28-34.

Brondani, R. P. V., Brondani, C., Tarchini, R., and Grattapaglia, D. (1998). Development, characterisation and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics* 97**,** 816-827.

Brooker, M. I. H. (2000). A new classification of genus *Eucalyptus* L"Her. (*Myrtaceae*). *Australian Systematic Botany* 13**,** 79-148.

Calleja-Rodriguez, A., Pan, J., Funda, T., Chen, Z-Q., Baison, J., Isik, F., Abrahamsson, S., and Wu, H. X. (2020). Genomic prediction accuracies and abilities for growth and wood quality traits of Scots pine, using genotyping-by-sequencing (GBS) data. *bioRxiv 607648* In press.

Cappa, E. P., De Lima, B. M., Da Silva-Junior, O. B., Carcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Science* 284**,** 9-15.

Cetinkol, O. P., Smith-Moritz, A. M., Cheng, G., Lao, J., George, A., Hong, K., Henry, R., Simmons, B. A., Healzwood, J. L., and Holmes, B. M. (2012). Structural and chemical characterisation of hardwood from tree species with applications as bioenergy feedstocks. *PLOS One* 7**,** e52820.

Chamberland, V., Robichaud, F., Perron, M., Gelinas, N., Bousquet, J., and Beaulieu, J. (2020). Conventional versus genomic selection for white spruce improvement: a comparison of costs and benefits of plantations on Quebec public lands. *Tree Genetics & Genome* 16**,** 17.

Chen, C., Mitchell, S. E., Elshire, R. J., Buckler, E. S., and El-Kassaby, Y. A. (2013). Mining conifers' mega-genome using rapid and efficient multiplexed high-throughput genotyping-by-sequencing (GBS) SNP discovery platform. *Tree Genetics & Genome* 9**,** 1537-1544.

Chen, Z-Q., Baison, J., Pan, J., Karlsson, B., Anderson, B., Westin, Rosario, M., Gil, G., and Wu, H. X. (2018). Accuracy of genomic selection for growth and wood quality traits in two control pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BCM Genomics* 19**,** 946.

Choi, M., Schill, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloglu, A., Ozen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., and Lifton, R. P. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *PNAS* 106, 45**,** 19096-19101.

Christensen, O. F. (2012). Compatability of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genetics Selection Evolution* 44.

Cleveland, M. A., Forni, S., Deeb, N., and Maltecca, C. (2010). Genomic breeding value prediction using three bayesian methods and application to reduce density marker panels. *BCM Proceedings* 4**,** 1.

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transaction of the Royal Society* 363**,** 557-572.

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., and Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *The Plant Genome* 11**,** 1-14.

Crossa, J., Perez-Rodriguez, Cuevas, J., Montesinos-Lopez, O., Jarquin, D., de Los Campos, G., Burgueno, J., Gonzalez-Camacho, J., Perez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., and Varshney, R. K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science* 22, 11**,** 961-975.

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J.A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185**,** 1021-1031.

Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2007). Inbreeding and genome-wide selection. *Journal Animal Breeding and Genetics* 124**,** 369-376.

Dasgupta, M. G., Dharanishanthi, V., Agawal, I., and Krutovsky, K. V. (2015). Development of genetic markers in *Eucalyptus* species by target enrichment and exome sequencing. *PLOS One* 10**,** e0116528.

de Los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H, D., and Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327-345.

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

de Toit, B., Smith, C.W., Little, K.M., Boreham, G., and Pallet, R.M (2010). Intensive, site-specific silviculture: Manipulating resource availability at establishment for improved stand productivity. A review of South African research. *Forest Ecology and Management* 259**,** 1836-1845.

Denis, M., and Bouvet, J-M. (2013). Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. *Tree Genetics and Genomes* 9**,** 37-51.

Desta, Z. A., and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. *Trends in Plant Science* 19**,** 592-601.

Devappa, R. K., Rakshit, S. K., and Dekker, R. F. H. (2015). Forest biorefinery: Potential of poplar phytochemicals as value-added co-products. *Biotechnology Advances* 33**,** 681-716.

Duran, R., Isik, F., Zapara-Valenzuela, J., Balocchi, C., and Valanzuela, A. (2017). Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genetics & Genome* 13, 74.

El-Dien, O. G., Ratcliffe, B., Klapste, J., Porth, I., Chen, C., and El-Kassaby, Y. A. (2016). Implementation of the realised genomic relationship matrix to open-pollinated white spruce family testing for disentangling additive from non-additive genetic effects. *Genes Genomics Genetics***,** 743-753.

Elbasyoni, I., Lorenzo, A. J., Guttieri, M., Frels, K., Baenziger, P. S., Poland, J., and Akhunov, E. (2018). A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Science* 270**,** 123-130.

Fisher, R. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 58**,** 399-433.

Forni, S., Aguilar, I., and Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* 43, 1.

Freeman, J. S., Whittock, S.P., Potts, B.M., and Vaillancourt, R.E. (2009). QTL influencing growth wood properties in *Eucalyptus globulus*. *Tree Genetics and Genomes* 5**,** 713-722.

Gardner, R. A. W., and Bertling, I. (2005). Effect of winter chilling and paclobutrazol on floral bud production in *Eucalyptus nitens*. *South African Journal of Botany* 71**,** 238-249.

Goddard, M. E. (2009). Genomic selection: Prediction of accuracy and maximisation of long term responses. *Genetica* 136**,** 245-257.

Grattapaglia, D. (2004). Integrating genomics into *Eucalyptus* breeding. *Genetic and Molecular Research* 3**,** 369-379.

Grattapaglia, D., and Kirst, M. (2008). *Eucalyptus* applied genomics: from gene sequence to breeding tools. *New Phytologist* 179**,** 911-929.

Grattapaglia, D., Bertolucci, F. L., and Sederoff, R. R. (1995). Genetic mapping of QTLs controlling vegetative propagation in *Eucalyptus grandis* and *E. urophylla* using pseudo-testcross strategy and RAPD markers. *Theoretical and Applied Genetics* 90**,** 933-947.

Grattapaglia, D., Bertolucci, F.L.G., Penchel, R., and Sederoff, R.R. (1996). Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* 144**,** 1205-1214.

Grattapaglia, D., Macos, D., and Resende, D.V. (2011). Genomic selection in forest tree breeding. *Tree Genetics and Genomes* 7**,** 241-255.

Grattapaglia, D., Plomion, C., Kirst, M., and Sederoff, R. R. (2009). Genomics of growth traits in forest trees. *Current Opinion in Plant Biology* 12**,** 148-156.

Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Muller, B. S. F., Tan, B., Isik, F., Ratcliffe, D., and El-Kassaby, Y. A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science* 9**,** 1693.

Griffin, A. R., Whiteman, P., Rudge, T., Burgess, I. P., and Moncur, M. (1993). Effect of paclobutrazol on flower-bud production and vegetative growth in two species of *Eucalyptus*. *Canadian Journal of Forest Research* 23**,** 640-647.

Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., and Su, S. (2014). Comparison of single-trait and multiple-trait genomic prediction models. *BCM Genetics* 15**,** doi:10.1186/1471-2156-1115-1130.

Hall., D. Hallingback, H. R., and Wu, H. X. (2016) Estimation of number and size of QTL effects in forest tree traits. *Tree Genetics & Genomics* 2, 110

Habier, D., Fernando, R. L., and Dekker, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177**,** 2389-2397.

Haile, J. K., N'diaye, A., Clarke, F., Clarke, J., Knox, R., Rutkoski, J., Bassi, F. M., and Pozniak, C. J. (2018). Genomic selection for grain yield and quality traits in durum wheat. *Molecular Breeding* 38**,** 78.

He, J., Zhao, X., Laroche, A., Lu, Z-X., Liu, H. K., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science* 5, 484.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31**,** 423-447.

Holland, J. B. (2004). "Implementation of molecular markers for quantitative traits in breeding programs - challenges and opportunities", in *Proceedings of the 4th International Crop Science Congress.* (Australia, Brisbane: www.cropscience.org.au).

Hughes, L., Cawsey, E. M., and Westoby, M. (1996). Climatic range sizes of *Eucalyptus* species in relation to future climate change. *Global Ecology and Biogeography Letters* 2**,** 23-39.

Imai, A., Kuniga, T., Yoshioka, T., Nonaka, K., Mitani, N., Fukamachi, H., Heihata, N., Yamamoto, M., and Hayashi, T. (2019). Single-step genomic prediction of fruit-quality traits using phenotypic records of non genotyped relatives in citrus. *PLOS One* 14**,** e0221880.

Irland, L. C., Adams, D., Alig, R., Betz, C.J., Chen, C-C., Hutchins, M., Mccarl, B.A., Skog, K., and Sohngen, B.L. (2001). Assessing socioeconomic impacts of climate change on US forests, wood-product markets, and forest recreation: The effects of climate change on forests will trigger market adaptations in forest management and in wood-products industries and may well have significant effects on forest-based outdoor recreation. *BioScience* 51**,** 753-764.

Isik, F., Bartholome, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C., and Bouffier, L. (2015). Genomic selection in maratime pine. *Plant Science* 242**,** 108-119.

Jannink, J-L. (2010a). Dynamics of long-term genomic selection. *Genetics Selection Evolution* 43**,** 35.

Jannink, J-L., Lorenz, A. J., and Iwata, H. (2010b). Genomic selection in plant breeding: from theory to practice. *Briefing in Functional Genomics* 9**,** 166-177.

Karama, E., Lund, M. S., and Su, G. (2020). Multi-trait single-step genomic prediction accounting for heterogeneous (co)variances over the genome. *Heredity* 124**,** 274-287.

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L-S., and Peterson, A. H. (2016). Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science* 242**,** 14-22.

Klapste, J., Dungey, H. S., Graham, N. J., and Telfer, E. J. (2020). Effect of trait's expression level on single-step genomic evaluation of resistance to *Dothistroma* needle blight. *BCM Plant Biology* 20**,** 205.

Klapste, J., Suontama, M., Telfer, E., Graham, N., Low, C., Stovold, T., Mckinley, R., and Dungey, H. (2017). Exploration of genetic architecture through sib-ship reconstruction in advanced breeding population of *Eucalyptus nitens*. *PloS One* 12**,** e0185137.

Klapste, J., Suontanna, M., Dungey, H. S., Telfer, E. J., Graham, N. J., Low, C. B., and Stovold, G. T. (2018). Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. *Journal of Heredity* 109**,** 802-810.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92**,** 4656-4663.

Legarra, A. C., O. F., Aguilar, I., and Misztal, I. (2014). Single step, a general approach for genomic selection. *Livestock Science* 166**,** 54-65.

Lenz, P. R. N., Beaulieu, J., Mansfield, S. D., Clement, S., Desponts, M., and Bousquet, J. (2017). Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BCM Genomics* 18**,** 335.

Lenz, P. R. N., Nadeau, S., Mottet, M-J., Perron, M., Isabel, N., Beaulieu, J., and Bousquet, J. (2019). Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. *Evolutionary Applications* 13, 1**,** 76-94.

Li, Y., and Dungey, H. S. (2018). Expected benefit of genomic selection over forward selection in conifer breeding and deployment. *PLOS One* 13**,** e0208232.

Li, Y., Kadarmideen, H. N., and Dekkers, J.C.M. (2008). Selection on multiple QTL with control of gene diversity and inbreeding for long-term benefit. *Journal Animal Breeding and Genetics* 125**,** 320-329.

Lin, Z., Hayes, B. J., and Daetwyler, H. D. (2014). Genomic selection in crops, trees and forages: a review. *Crop and Pasture Science* 65**,** 1177-1191.

Mamani, E. M. C., Bueno, N. W., Faria, D. A., Guimaracs, D. M. S., Lau, D., Alfenas, A. C., and Grattapaglia, D. (2010) Positioning of the major locus for *Puccinia psidii* rust resistance (*Ppr*1) on the *Eucalyptus* reference map and its validation across unrelated pedigrees. *Tree Genetics & Genome* 6, 953-962

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Review Genetics* 11**,** 31-46.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157**,** 1819-1829.

Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full and genomic information. *Journal of Dairy Science* 92**,** 4648-4655.

Mostert-O'neill, M. M., Reynolds, S. M., Acosta, J. J., Lee, D. J., Borevitz, J. O., and Myburg, A. A. (2020). Genomic evidence of introgression and adaptation in a model subtropical tree species, *Eucalyptus grandis*. *Molecular Ecology* 1**,** 14.

Mostert-O'neill, M. M., Reynolds, S. M., Acosta, J. J., Lee, D. J., Borevitz, J. O., and Myburg, A. A. (2022). Genomic consequences of artificial selection during early domestication of a wood fibre crop. *New* Phytologist

Muir, W. M. (2007). Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal Animal Breeding and Genetics* 124**,** 342-355.

Müller, B. S. F., Neves, L. G., De Almeida Filho, J. E., Resende, M. F. R., Muñoz, P. R., Dos Santos, P. E. T., Filho, E. P., Kirst, M., and Grattapaglia, D. (2017). Genomic prediction

in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genomics* 18**,** 524.

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D.,Goodstein, D.M., Dubchak, I., Poliakov, A., Mizrachi, E., Kullan, A.R.K., Hussey, S.G., Pinard, D., Van Der Merwe, K., Signh, P., Van Jaarsveld, I., Silva-Junior, O.B., Togawa, R.C., Pappas, M.R., Faria, D.A., Sansaloni, C.P., Petroli, C.D., Yang, X., Ranjan, P., Tschaplinski, T.J., Ye, C-Y., Li, T., Sterck, L., Vanneste, K., Murat, F., Soler, M., San Clemente, H., Saidi, N., Cassan-Wanh, H., Dunand, C., Hefer, C.A., Bornberg-Bauer, E., Kersting, A.R., Vining, K., Amarasinghe, V., Ranik, M., Naithani, S., Elser, J., Boyd, A.E., Liston, A., Spatafora, J.W., Dharmwardhana, P., Rija, R., Sullivan, C., Romanel, E., Alves-Ferreira, M., Kulheim, C., Foley, W., Carocha, V., Paiva, J., Kudrna, D., Brommonschenkel, S.H., Pasquali, G., Byrne, M., Rigault, P., Tibbits, J., Spokeviciuos, A., Jones, R.C., Steane, D.A., Vaillancourt, R.E., Potts, B.M., Joubert, F., Barry, K., Pappas Jr, G.J., Strauss, S.H., Jaiswal, P., Grima-Pettenati, J., Salse, J., Van De Peer, Y., Rokhsar, D.S., and Schmutz, J. (2014). The genome of *Eucalyptus grandis*. *Nature* 510**,** 356-375.

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G., and Baloch, F. S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment* 32**,** 261-285.

Naidoo, S., Kulheim, C., Zwart, L., Mangwanda, R., Oates, C. N., Visser, E. A., Wilhen, F. E., Mamni, T. B., and Myburg, A. A. (2014). Uncovering the defence responses of *Eucalyptus* to pests and pathogens in the genomics age. *Tree Physiology* 34**,** 931-943.

Naidoo, S., Slippers, B., Plett, J. M., Coles, D., and Oates, C. N. (2019). The road to resistance in forest trees. *Frontiers in Plant Science* 10**,** 273.

Namkoong, G., Barnes, R.D., and Burley, J. (1980). Screening for yield in forest tree breeding. *Commonwealth Forestry Review* 59**,** 61-68.

Nyouma, A., Bell, J. M., Jacob, F., and Cros, D. (2019). From mass selection to genomic selection: one century of breeding for quantitative yield components of oil palm (*Elaeis guineensis* Jacq.). *Tree Genetics & Genome* 15**,** 69.

Pais, A. L., Whetten, R. W., and Xiang, Q-Y (2020). Population structure, landscape genomics, and genetic signatures of adaptation to exotic disease pressure in *Cornus florida* L.; insights from GWAS and GBS data. *Journal of Systematics and Evolution* 0**,** 1-25.

Pavy, N., Gagnon, F., Deschenes, A., Boyle, B., Beaulieu, J., Abd Bousquet, J. (2016). Development of highly reliable in silico SNP resource and genotyping assay from exome capture and sequencing: an example from black spruce (*Picea mariana*). *Molecular Ecology Resources* 16, 2**,** 588-598.

Payn, T., Carnus, J-M., Freer-Smith, P., Kimberley, M., Kollert, W., Liu, S., Orazio, C., Rodriguez, L., Silva, L. N., and Wingfield, M. J. (2015). Changes in planted forests and future global implications. *Forest Ecology and Management* 352**,** 57-67.

Perlack, R. D., Turhollow, W. L .L., Af, G. R. L., Stokes, B. J., Erbach, D. C. (2005). "Biomass as a feedstock for a bioenergy and bioproduct industry: the technical feasibility of a billion-ton annual supply", (ed.) *O.R.N. Laboratory*. (USA: Department of Energy).

Piepho, H. P., Mohring, J., Melchinger, and Buchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161**,** 209-228.

Poland, J. A. R., J. (2016). Advances and challenges in genomic selection for disease resistance. *Annual Review of Phytopathology* 54**,** 79-98.

Rambolarimanana, T., Ramamonjisoa, L., Verhaegen, D., Tsy, J-M. L. P., Jacquin, L., Cao-Hakouanzi, G., and Bouvet, J-M. (2018). Performance of multi-trait genomic selection for *Eucalyptus robusta* breeding program. *Tree Genetics & Genome* 14**,** 71.

Ratcliffe, B., El-Dien, O. G., Cappa, E . P., Porth, I., Klápste, J., Chen, C., and El-Kassaby, Y. A. (2017). Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. *Genes Genomics Genetics* 7**,** 935-942.

Raymond, C.A. (2002). Genetics of *Eucalyptus* wood properties *Annual Forestry Science* 59**,** 525-531.

Resende Jr, M. F. R., Munoz, P., Acosta, J.J., Peter, G.F., Davis, J.M., Grattapaglia, D., Resende, M.D.V., and Kirst, M. (2012a). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193**,** 617-624.

Resende Jr, M. F. R., Munoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F., and Kirst, M. (2012b). Accuracy of genomic selection methods in standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190**,** 1503-1510.

Resende, M. D. V., Resende Jr, M. F. R., Sansaloni, C. P., Petroli, C. D., Missiaggia, A.A., Aguiar, A. M., Abad, J. M., Takahashi, E. K., Rosado, A. M., Faria, D. A., Pappas Jr, G.J., Kilian, A., and Grattapaglia, D. (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* 193**,** 617-624.

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevero, C. F., Takahashi, E. K., Silva-Junior, O. B., and Grattapaglia, D. (2016). Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in *Eucalyptus*. *New Phytologist* 213**,** 1287-1300.

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevero, C. F., Takahashi, E. K., Silva-Junior, O. B., and Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity* 119**,** 245-255.

Rodriguez-Ramilo, S. T., Garcia-Cortes, L. A., and Rodriguez De Cara, M. A. (2015). Artificial selection with traditional or genomic relationship: consequences in coancestry and genetic diversity. *Frontiers in Genetics* 6**,** 127.

Sallam, A. H., Endelman, J. B., Jannink, J-L., and Smith, K. P. (2015). Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *The Plant Genome* 8**,** 1.

Sansaloni, C., Petroli, C., Carling, J., Hudson, C. J., Steane, D. A., Myburg, A. A., Grattapaglia, D., Vaillancourt, R. E., and Kilian, A. (2010). A high-density diversity arrays technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* 6, 16.

Silva-Junior, O. B., and Grattapaglia, D. (2015a). Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytologist* 208**,** 830-845.

Silva-Junior, O. B., Faria, D.A., and Grattapaglia, D. (2015b). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206**,** 1527-1540.

Silva, P. H. M., Miranda, A. C., Moraes, M. L. T., Furtado, E. L., Stape, J. L., Alvares, C. A., Sentelhas, P. C., Mori, E. S., Ans Sebbenn, A. M. (2013). Selecting for rust (*Puccinia psidii*) resistance in *Eucalyptus grandis* in Sao Paulo State, Brazil. *Forest Ecology and Management* 303**,** 91-97.

Silveira, L. S., Martins Filho, A., Azevendo, C. F., Barbosa, E. C., Resende, M. D. V., and Takahashi, E. K. (2019). Bayesian models applied to genomic selection for categorical traits. *Genetics and Molecular Research* 18**,** 4.

Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. (2012). Genomic selection required genetic control of inbreeding *Genetic Selection Evolution* 44.

Stafford, W., De Lange, W., Nahman, A., Chunilall, V., Lekha, P., Andrew, J., Johakimu, J., Sithole, B., Trotter, D. (2020). Forestry biorefineries. *Renewable Energy* 154**,** 461-475.

Stanturf, J. A., Vance, E. D., Fox, T. R., and Kirst, M. (2013). *Eucalyptus* beyond its native range: Environmental issues in exotic bioenergy plantations. *International Journal of Forestry Research* 2013, ID 463030.

Strauss, S. H., Lande, R., and Namkoong, G. (1992). Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian Journal of Forest Research* 22**,** 1050-1061.

Suontama, M., Klápště, J., Telfer, E., Graham, N., Stovold, T., Low, C., Mckinley, R., and Dungey, H. (2019). Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* 122**,** 370-379.

Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., and Ingvarsson, P. K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BCM Plant Biology* 17**,** 110.

Tan, B., Grattapaglia, D., Wu, H. X., and Ingvarsson, P. K. (2018). Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Science* 267**,** 84-93.

Telfer, E., Graham, N., Macdonald, L., Li, Y., Klapste, J., Resende Jr, M., Neves, L. G., Dungey, H., and Milcox, P. (2019). A high-density exome capture genotype-by-sequencing panel for forestry breeding in *Pinus radiata*. *PLOS One* 14**,** e0222640.

Telfer, E. J., Stovold, G. T., Li, Y., Silva-Junior, O. B., Grattapaglia, D. G., and Dungey, H. S. (2015). Parentage reconstruction in *Eucalyptus nitens* using SNPs and microsatellite markers: A comparative analysis of marker data power and robustness. *PLOS ONE* 10**,** e0130601.

Thistlethwaite, F. R., Ratcliffe, B., Klapste, J., Porth, I., Chen, C., Stoehr, M. U., and El-Kassaby, Y. A. (2019). Genomic selection of juvenile height across a single-generational gap in Douglas-fir. *Heredity* 122**,** 848-863.

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27**,** 522-530.

Wang, C-L., Ma, P-P., Zhang, Z., Ding, X-D., Liu, J-F., Weng, Z-Q., and Zhang, Q. (2012). Comparison of five methods for genomic breeding value estimation for the common dataset of the 15th QTL-MAS workshop. *BMC Proceedings* 6**,** S13.

Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018). Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal* 6**,** 330-340.

White, T. L., And Hodge, G.R. (1988). Best linear prediction of breeding values in forest tree improvement programs. *Theoretical and Applied Genetics* 76**,** 719-727.

Williams, D. R., Potts, B. M., and Smethurst, P. J. (2003). Promotion of flowering in *Eucalyptus nitens* by paclobutrazol was enhanced by nitrogen fertilizer. *Canadian Journal of Forest Research* 33**,** 74-81.

Williams, E. R., John, J.A., and Whitaker, D. (1999). Examples of block designs for plant and tree breeding trials. *Australian and New Zealand Journal of Statistics* 41**,** 277-284.

Wimmer, R., Downes, G. M., Evans, R., Rasmussen, G., and French, J. (2002). Direct effects of wood characteristics on pulp and handsheet properties of *Eucalyptus globulus*. *Holzforschung* 56**,** 244-252.

Wingfield, M. J., Brocherhoff, E.G., Wingfield, B.D., and Slippers, B. (2015). Planted forest health: The need for a global strategy. *Science* 349**,** 832-836.

Wingfield, M. J., Roux, J., Slippers, B., Hurley, B.P., Garnas, J., Myburg, A.A., and Wingfield, B.D. (2013). Established and new technologies reduce increasing pest and pathogen threats to Eucalypt plantations. *Forest Ecology and Management* 301**,** 35-42

Wingfield, M. J., Slippers, B., Hurley, B. P., Coutinho, T. A., Wingfield, B. D., and Roux, J. (2008). Eucalypt pest and diseases: Growing threats to plantation productivity. *Southern Forests* 70**,** 139-144.

Xu, S. (2003). Theoretical basis of the beavis effect. *Genetics* 165**,** 2259-2268.

Zapata-Valenzuela, J., Whetten, R. W., Naele, D., Mckeand, S., and Isik, F. (2013). Genomic estimated breeding values using genomic relationship matrices in a cloned population of Loblolly Pine. *Genes Genomics Genetics* 3**,** 909-916.

Zelener, N., Poltri, S.N.M., Bartoloni, N., Lopez, C.R., and Hopp, H.E. (2005). Selection strategy for seedling seed orchard design based on trait selection index and genomic analysis by molecular markers: a case study for *Eucalyptus dunnii*. *Tree Physiology* 25**,** 1457-1467.

Zhang, Z., Liu, J., Ding, X., Bijma, P., De Koning, D-J., and Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding value using a trait-specific marker derived relationship matrix. *PLOS One* 5**,** e12648.

Zhao, Y., Mette, M. F., and Reif, J. C. (2015). Genomic selection in hybrid breeding. *Plant Breeding* 134**,** 1-10.

Zhong, S., Dekker, J. C. M., Fernando, R. L., and Jannink, J-L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A Barley case study. *Genetics* 182**,** 355-364.

CHAPTER 2

EXPECTED BENEFITS OF GENOMIC SELECTION FOR GROWTH AND WOOD QUALITY TRAITS IN *EUCALYPTUS GRANDIS*

Makobatjatji M. Mphahlele[1,2], Fikret Isik[3], Marja M. Mostert-O'Neill[2], S. Melissa Reynolds[2], Gary R. Hodge[3,4], and Alexander A. Myburg[2*]

[1]Mondi Forests, Research and Development Department, Trahar Technology Centre - TTC, P.O. Box 12, Hilton, 3245, South Africa
[2]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa
[3]Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA
[4]Camcore, North Carolina State University, P.O. Box 7626, Raleigh, NC 27695, USA

Author contributions:

This data chapter has been published in the international ISI journal, Tree Genetics & Genomics as Mphahlele, M. M., Isik, F., Mostert-O'Neill, M. M., Reynolds, S. M., Hodge, G. R., and Myburg, A. A. (2020). Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. *Tree Genetics & Genome* 16**,** doi.org/10.1007/s11295-11020-01443-11291. I carried out the experimental design, data collection, data analysis, and drafting of this paper as part of this PhD thesis. F.I. assisted with the modelling and scripts of the data analysis tools. F. I., G. H., and A. A. M. supervised the study and helped with data interpretation and the drafting of the manuscript. M.O. assisted with SNP data analysis, and M.R. assisted with DNA extraction and pedigree reconstruction. All authors read, edited and approved the final manuscript.

2.1 ABSTRACT

Genomic selection (GS) can substantially reduce breeding cycle times in forest trees compared to traditional breeding cycles. Practical implementation of GS in tree breeding requires assessing significant genetic gains over time, which may differ among species and breeding objectives. We present a GS study of growth and wood quality traits in an operational *Eucalyptus grandis* breeding program in South Africa. The training population consisted of 1,575 full and half-sib individuals, genotyped with the *Eucalyptus* (EUChip60K) SNP chip resulting in 15,040 informative SNP markers. The accuracy of the GS models ranged from 0.47 (diameter) to 0.67 (fibre width). We compared a four-year GS breeding cycle equivalent to half of a traditional eight-year *E. grandis* breeding cycle and obtained GS efficiencies ranging from 1.20 (wood density) to 1.62 (fibre length). Simulated over 17 years, the ratio of the accumulated genetic gains between three GS cycles and two traditional breeding cycles ranged from 1.53 (diameter) to 3.35 (wood density). To realise these genetic gains per unit time in *E. grandis* breeding, we show that significant adjustments have to be made to integrate GS into operational breeding steps.

2.2 INTRODUCTION

Eucalyptus species constitute about 14 million of 261 million hectares of plantation forestry (Carle and Holmgren, 2008). The genus *Eucalyptus* has more than 800 different species, of which the "Big Nine" species *E. urophylla, E. tereticornis, E. camaldulensis, E. saligna, E. dunnii, E. grandis, E. pellita, E. nitens, E. globulus,* and their hybrids support the bioenergy, biochemical and biomaterials industries (Shepherd et al., 2011; Stanturf et al., 2013). These industries' sustainability relies on the high adaptability, fast growth, and superior wood quality of *Eucalyptus* species in forest plantations. Climate change affects traditionally productive forestry plantation areas (Irland et al., 2001; Booth, 2013), as are pests and pathogens aggravated by global trade (Wingfield et al., 2008; Wingfield et al., 2015). Faster and more agile breeding approaches are therefore needed to ensure the future sustainability of eucalypt forestry plantations.

Traditional forestry tree breeding systems still face challenges, mainly due to the long rotation cycles associated with reproductive maturity and the time-to-maturity of commercially important traits such as wood quality. The selection of superior families and individuals in these tree breeding systems primarily relies on the empirical breeding values of valuable traits such as tree growth, wood quality and tolerance to biotic and abiotic stresses (Namkoong et al., 1980; White and Hodge, 1988). Genetic relatedness is a critical consideration in determining genetic merit. Pedigree information in BLUP analysis uses additive relationships among individuals to derive the variance-covariance relationships among all observations when making genetic value predictions (Piepho et al., 2008). However, pedigree relationships represent the average proportion of shared alleles (at infinite loci) identical by descent (IBD), often ignoring Mendelian sampling effects among segregating individuals in families. This average proportion leads to an overestimation of

genetic parameters, affecting variables and their correlated responses (Veerkamp et al., 2011).

The inclusion of single nucleotide polymorphism (SNP) markers, the most abundant form of DNA polymorphism in plant genomes (Agarwal et al., 2008; Mammadov et al., 2012), makes tracking of the Mendelian sampling effects of individuals in families possible (Hill and Weir, 2010). Genomic relationship matrix derived from SNP markers accurately estimates the genomic proportions that are IBD to capture Mendelian segregation (within families) and allow detection of more precise relationships (between families) and the correction of erroneous pedigree records (Hayes et al., 2009). The genomic relationship matrix is blended with the pedigree matrix of non-genotyped individuals to modify covariances of the ancestors and descendants of the genotypes individuals to predict genetic merit better, an approach called single-step genomic BLUP (Legarra et al., 2009; Christensen, 2012; Misztal et al., 2013; Isik et al., 2017), with applications in tree breeding (Ratcliffe et al., 2017; Klapste et al., 2018; Cappa et al., 2019). The blending of pedigree and genome markers matrices is a cost-effective approach to maximise breeding value prediction accuracy in non-genotyped tree breeding populations with shallow open-pollinated pedigree structures.

Genomic selection (GS), as a breeding tool, predicts genotyped individual genetic merit without phenotypes (defined as *genomic estimated breeding values*, GEBVs). The prediction is based on the aggregate modelling of the training population genomic and phenotypic information (Meuwissen et al., 2001). In practice, GS approaches maximise genetic gain per unit time and cost by predicting breeding values early in the breeding cycle, eliminating field testing. GS studies have demonstrated encouraging results in its application in many genetic

improvement systems, such as livestock (Schaeffer et al., 2006; Hayes et al., 2008; Luan et al., 2009; Bouquet et al., 2013; Garcia-Ruiz et al., 2016; Wolc et al., 2016; Wiggans et al., 2017) and crops (Bassi et al., 2016; Crain et al., 2018; Haile et al., 2018; Cros et al., 2019; Voss-Fels et al., 2019). GS continues to revolutionise breeding approaches by not just enabling accurate prediction of related individuals but by allowing complex interrogation of genetic and environmental interactions and identifying genomic regions that are stable or responsive to specific environments (Crossa et al., 2017).

GS is of particular benefit in forest trees due to the extended breeding cycles because of delayed reproductive maturity and the need for early selection of late expressing (mature) growth and wood quality traits. Grattapaglia et al. (2011) performed the first simulation study to demonstrate the potential of GS in tree breeding and highlighted crucial factors to consider, such as marker density, size of training population, trait heritability and QTL number. Since then, numerous studies have demonstrated GS as a tool for accelerated tree improvement (Grattapaglia et al., 2018). GS studies in forest species have shown acceptable prediction accuracies for adoption in conifer species such as *Pinus taeda* (Resende Jr et al., 2012; Zapata-Valenzuela et al., 2013), *Picea glauca* (Beaulieu et al., 2014) and *Pinus pinaster* (Isik et al., 2015; Bartholome et al., 2016). The reference genome sequence of *E. grandis* (Myburg et al., 2014) and the development of a robust genome-wide SNP genotyping (EUChip60K) chip platform (Silva-Junior et al., 2015b) have created opportunities to study and apply GS in *Eucalyptus* species and their hybrids. Prediction of genetic merit using GS has performed well in interspecific *Eucalyptus* hybrids (Resende et al., 2012; Tan et al., 2017), where linkage disequilibrium is high, as well as in open-pollinated pure-species breeding populations of *E. pellita* and *E. benthamii* (Müller et al., 2017). Suontama et al. (2019) was able to demonstrate improved breeding value accuracy as

well as increased genetic gains in an *E. nitens* solid-wood breeding population. An elite clonal population of *E. globulus* demonstrated encouraging results of GEBV estimates for wood density and stem volume (Duran et al., 2017). Predicting the genetic merit of individuals earlier in the breeding cycle can increase gains per unit time in trees (Bartholome et al., 2016; Resende et al., 2017; Li and Dungey, 2018). Adopting GS as an alternative or a complementary tree breeding strategy for growth and wood quality traits requires a practical demonstration of its benefits concerning traditional tree breeding (TB) approaches.

In this study, we interrogate the benefits of implementing GS for growth and wood quality traits in the context of an established *E. grandis* breeding programme. First, we evaluated the predictive ability of GS based on genome-wide SNP markers. We then analysed the efficiency of an accelerated GS breeding strategy compared to the conventional approach. Finally, we investigated the genetic gains achieved by GS breeding for growth and wood quality traits in *E. grandis*.

2.3 MATERIAL AND METHODS

2.3.1 Training population and phenotype assessment

The training population was derived from a series of *E. grandis* (Hill ex Maiden) breeding trials provided by the Mondi South Africa tree breeding programme. The training population is composed of a total of 1,548 trees, genotyped and phenotyped from four trials. The genotyped and phenotyped individuals per family were selected from Tygerskloof (eight per family), Montigny (seven per family) and Port Durnfort (nine per family) traits to span the phenotypic range relative to the diameter plot average, whereas, for the Ncalu trial, all available standing trees were sampled. The Ncalu trial is a full-sib progeny trial, whereas the Montigny, Tygerskloof and Port Durnfort trials are half-sib progeny age-age correlation

trials. An additional 27 parental selections were also genotyped. The number of related parents between the trials ranged from 5 to 40, and all progeny are from the same breeding generation. **Table 2.1** presents the trial design information, environmental conditions, family relatedness, and the summary statistics of the measured growth and wood quality traits. Growth and wood quality traits, such as diameter at breast height and tree height, were measured at age seven using a diameter tape and a Vertex Hypsometer (Haglof, Sweden), respectively. Non-destructive wood sampling was performed at breast height from bark-to-bark by extracting 22 mm increment cores and wood shavings. Extractives were removed from the increment cores overnight by soaking in acetone, after which basic wood density was determined using the water displacement method (Tappi methods 258 om-02). Weighted fibre length and fibre width were measured using the MorFi Compact Fibre and Shive Analyzer (TachPap, France) from wood shavings macerated with a 1:1 solution of acetic acid and 50% hydrogen peroxide for four hours at 90°C. Percentage α-cellulose, as well as the syringyl to guaiacyl (S/G) lignin monomer ratio, were estimated using near-infrared (NIR) spectroscopy models with the OPUS/QUANT Spectroscopy Software Version 6 (Bruker, Massachusetts) calibrated from *E. grandis* breeding material partially related to the training population **(Table 2.2)**. The raw phenotypic data was used directly in the downstream analyses. Pedigree records showed that nine of the seed parents are shared between the full-sib and half-sib families as both seed and pollen contributors **(Table 2.3)**. The full-sib progeny in the training population was confirmed with microsatellite DNA fingerprinting (Brondani et al. 1998), as presented in **Table 2.3**.

TABLE 2.1 | Environmental and plant materials of the *E. grandis* study population. Trial design, environmental conditions, overlap in families and summary statistics of measured phenotypes are presented.

	Mistley (CSO)	**Ncalu (NCA)**	**Montigny (MNT)**	**Tygerskloof (TYG)**	**Port Durnfort (PDF)**
Environment					
Latitude (South)	29^0 36'	30^0 20'	32^0 10'	31^0 16'	27^0 47'
Longitude (East)	30^0 22'	30^0 08'	28^0 35'	27^0 47'	27^0 47'
Altitude (m)	1023	1134	47	1118	32
MAP (ml)	832	809	1086	785	1454
MAT (^{0}C)	17	16	21	17	21
MAT min. (^{0}C)	5	4	11	4	11
MAT max. (^{0}C)	25	25	29	26	28
Trial design					
Plot design	Single	Single	Block	Block	Block
Progeny type	Parents	Full-sib	Half-sib	Half-sib	Half-sib
Tree per plot design	1	1x1	10x10	10x10	10x10
Families (Controls)	30	88 (5)	49	40	20
Replications	10	15	1	1	1
Number of plots	300	1395	49	40	20
Genotyped trees	27	709	343	320	176
Phenotyped trees		709	343	320	176
Family overlap					
Mistley		27	9	8	6
Ncalu			8	7	5
Montigny				40	20
Tygerskloof					18
Traits	**n**	**Min.**	**Max.**	**Mean**	**Std.dev**
Fibre length (mm)	1528	0.46	0.93	0.73	0.06
Fibre width (µm)	1528	15.86	23.52	19.70	1.07
α-Cellulose (%)	1538	36.00	54.10	46.06	2.08
S/G ratio	1538	1.26	2.77	2.14	0.22
Basic wood density ($kg.m^{-3}$)	1537	278.41	574.68	391.03	40.07
Diameter (cm)	1539	4.80	28.30	16.49	4.60
Height (m)	1539	7.90	30.90	20.74	4.76

MAP – mean annual precipitation and **MAT** – mean annual temperature. **Footnote:** The effective population size is 47.5, ranging from 21.9 for the full-sib trial to 46.3 for the half-sib trials. We used the formula described by Kimura, M., and Crow, J. F. (1963) for monoecious diploids**:** $\boldsymbol{N_e = \frac{N\underline{k}-2}{\underline{k}-1+\frac{V}{\underline{k}}}}$ where $\boldsymbol{\underline{k}}$ is the mean number of progeny per parent, $\boldsymbol{V}$ is the variance of the number of progeny per parent, and $\boldsymbol{N}$ is the number of parents. **Reference:** Kimura, M., and Crow, J. F. (1963). The measurement of effective population number. Evolution **17**(3): 279-288.

TABLE 2.2 | Near-infrared (NIR) spectroscopy models used to predict cellulose content and S/G lignin monomer ratio in the training population. The NIR model was calibrated using 200 samples obtained from two *E. grandis* breeding trials (128 samples) and a subset from the Tygerskloof (TYG) trial (72 samples). NIR calibration set (fitted vs true) and the validation set (predicted vs true) parameters are presented. Data processing of the NIR scans for model development involved a combination of the first (1st) derivative and vector normalization (SNV) computation using OPUS/QUANT Spectroscopy Software Version 6 (Bruker, Massachusetts).

Calibration	**Method**	**Samples**	**Rank**	**Slope[a]**	**R^2 [b]**	**RPD[c]**	**RMSEE[d]**	
α-Cellulose	SNV	124	9	0.919	92	3.52	0.661	
S/G ratio	1st derivative + SNV	96	10	0.918	92	3.5	0.0591	
Validation	**Method**	**Samples**	**Rank**	**Slope[a]**	**R^2 [b]**	**RPD[c]**	**RMSECV[e]**	**Bias[f]**
α-Cellulose	SNV	67	9	0.907	90	3.08	0.745	0.00353
S/G ratio	1st derivative + SNV	67	10	0.885	87	2.75	0.0733	-0.00168

[a]Linear regression slope of the calibration and validation data
[b]Coefficient of determination
[c]Residual prediction deviation
[d]Root mean square error of estimation
[e]Root mean square error of cross-validation
[f]Deviation from the linear regression slope

TABLE 2.3 | Full and half-sib pedigree information of the training population. The full-sib progeny (706) in the Ncalu trial were confirmed with microsatellite DNA fingerprinting. The half-sib progeny (203) belong to nine of the seed parents using breeders' pedigree records with unknown pollen parents. A large number of trees (639) from half-sib families are not shown in the table because their seed and pollen parents are unknown. The total SNP genotyped population is 1,575 individuals, including the full-sib and half-sib progeny and the 27 parental selections.

	Pollen parents							
Seed parents	Unknown	Parent 03	Parent 07	Parent 08	Parent 15	Parent 16	Parent 18	Total
Parent 01			9					9
Parent 02		10	7	7		5	2	31
Parent 03	56		7		5	11	4	83
Parent 04		8	8	11				27
Parent 05	24	8	9					41
Parent 06		10	8		9	5	7	39
Parent 07	15	12		8		8		43
Parent 08		10			9	11	9	39
Parent 09		10			11			21
Parent 10	1	9	10		7	7	7	41
Parent 11		6		6				12
Parent 12			7		10	9	6	32
Parent 13		8	9		8		8	33
Parent 14		8	9	13	8			38
Parent 15	28	7	5	8		4	4	56
Parent 16		9	9		8		9	35
Parent 17		11	6	11	3			31
Parent 19		8		10	11	7	9	45
Parent 20		2		5	7	4	4	22
Parent 21		9			13	10		32
Parent 22	23	8	5	9	10	8		63
Parent 23	15	6	10	8	12	11	10	72
Parent 24			9		6	7		22
Parent 25		1						1
Parent 26	23							23
Parent 27	18							18
Total	203	160	127	96	137	107	79	**909**

2.3.2 Genotyping

DNA was extracted from immature xylem and cambium tissue scrapings of mature trees using the NucleoSpin DNA extraction kit (Machery-Nagel, Germany). The *Eucalyptus* (EUChip60K) SNP chip, as described by (Silva-Junior et al. 2015) available from GeneSeek (Neogen, Lansing, MI, USA), was used for genotyping the 1,575 *E. grandis* trees. Informative SNP markers were retained with a call rate of over 90% and a minor allele frequency above 0.05 using the SVS software v8.4.3 (Golden Helix, Inc. Bozeman, MT). The genotype information was coded based on the additive gene content model to zero, one, and two, representing major homozygous (0.48), heterozygous (0.33) and homozygous minor (0.19) alleles, respectively (frequencies given in bracket). Missing allele data were imputed based on allelic distribution, assuming Hardy–Weinberg equilibrium using the *synbreed* 0.10-2 R package (Wimmer et al., 2012).

2.3.3 Linkage disequilibrium

Linkage disequilibrium as a coefficient of determination (r^2) was estimated with pairs of informative marker loci with a pairwise distance of up to 50 kb. The above analysis was performed using the SVS software v8.4.3 (Golden Helix, Inc. Bozeman, MT) for the entire study population, its half-sib and full-sib subpopulations. The coefficient of determination or allele frequency was computed as $\boldsymbol{r^2 = D^2/[P_A P_B(1-P_A)(1-P_B)]}$, and $\boldsymbol{r}$ is the correlation of the allele frequency (Hill and Robertson, 1968), where the frequency of allele $\boldsymbol{A}$ is $\boldsymbol{P_A}$, and for $\boldsymbol{B}$ is $\boldsymbol{P_B}$, and the allele pair (haplotype) is $\boldsymbol{P_{AB}}$ then the coefficient of LD as $\boldsymbol{D = P_{AB} - P_A P_B}$.

2.3.4 Linear mixed models analysis

Due to the differences in progeny types and the trial designs of single tree plots in Ncalu with 15 replication versus multiple plots at single replication in Montigny, Tygerskloof and Port

Durnfort, we could only consider the site of the trials as our fixed model term in the analysis. The G×E random model term was not included because of the low connectedness of the trials. Linear mixed models were fitted to estimate variance components and solve for fixed and random effects. The matrix notation for the linear mixed models used is as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{e} \qquad (1)$$

where $\boldsymbol{y}$ is a vector of measured phenotypes, $\mathbf{X}$, and $\mathbf{Z}$ are the incidence matrix for the fixed and random effects, respectively. The $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are the vectors of fixed and random effect coefficients, respectively; $\boldsymbol{e}$ is the residual effects vector. The site effect was considered as a fixed factor, while tree effect was treated as random. The expectations of $\boldsymbol{y}$, $\mathbf{u}$ and $\boldsymbol{e}$ are $\boldsymbol{E}(\boldsymbol{y}) = \mathbf{X\beta}$, $\boldsymbol{E}(\mathbf{u}) = \mathbf{0}$ and $\boldsymbol{E}(\mathbf{e}) = \mathbf{0}$ and the variances are $\boldsymbol{Var}(\boldsymbol{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$, $\mathbf{Var}(\mathbf{e}) = \mathbf{R} = \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$, and $\mathbf{Var}(\mathbf{u}) = \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}$, respectively, where $\mathbf{I}$ is the identity matrix, $\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$ is the variance associated with the residuals, and $\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}$ is the variance associated with the random effect. The $\boldsymbol{Var}(\mathbf{u})$ was scaled by the numerator relationship matrix $\mathbf{A}$ derived from the pedigree or by the $\mathbf{G}$ derived from the SNP markers. A restricted maximum likelihood approach was used to estimate variance components with the ASReml-R 3.0 R package (Butler et al., 2009) in the R environment (R_Core 2016).

The Henderson, (1975) mixed model equations were solved based on pedigree (ABLUP) to predict the empirical breeding values (EBV) of individuals:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\boldsymbol{\lambda} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\boldsymbol{y} \\ \mathbf{Z}'\boldsymbol{y} \end{bmatrix} \qquad (2)$$

where $\mathbf{A}^{-1}$ is the inverted additive genetic relationship matrix derived from the pedigree, $\boldsymbol{\lambda} = \boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}/\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}$ is the shrinkage factor. The direct genetic value (DGV) of individuals were predicted by solving the mixed model equations by substituting the $\mathbf{A}^{-1}$ matrix with the inverted realised

genomic relationship matrix $\mathbf{G^{-1}}$. The genomic relationship was computed as described in (VanRaden, 2008):

$$\boldsymbol{G} = \frac{(\mathbf{Z}-\mathbf{P})(\mathbf{Z}-\mathbf{P})'}{\mathbf{2}\sum \boldsymbol{p_i}(\mathbf{1}-\boldsymbol{p_i})} \quad \textbf{(3)}$$

where $\boldsymbol{Z}$ and $\boldsymbol{P}$ are two matrices of dimension $\boldsymbol{n}$ (individuals) × $\boldsymbol{p}$ (markers). The gene content values in matrix $\boldsymbol{Z}$ are −1 (homozygote major allele), 0 (heterozygote), and 1 (homozygote minor allele). The allele frequencies in matrix $\boldsymbol{P}$ are presented as $\mathbf{2}(\boldsymbol{p_i} - \mathbf{0.5})$, where $\boldsymbol{p_i}$ is the allele frequency observed at the marker $\boldsymbol{i}$ for all individuals. The variance of alleles summed across all loci is $\mathbf{2}\sum \boldsymbol{p_i}\,(\mathbf{1} - \boldsymbol{p_i})$.

The prediction accuracy of empirical breeding values (EBV) from the ABLUP models and direct genetic values (DGV) from the GBLUP models were estimated as:

$$\boldsymbol{r} = \sqrt{\mathbf{1} - (\boldsymbol{SE^2}/(\mathbf{1} + \boldsymbol{F})\boldsymbol{\sigma_u^2})} \quad \textbf{(4)}$$

where $\boldsymbol{SE^2}$ is squared standard error (variance) of the predictions, $\boldsymbol{\sigma_u^2}$ is the genetic variance component, and $\boldsymbol{F}$ is the inbreeding coefficient of the individuals (Gilmour et al., 2015), which is assumed to be zero for non-inbred individuals.

2.3.5 Genomic selection validation

A cross-validation approach was implemented to investigate the predictive ability of GBLUP models based on the above linear mixed model approaches. The cross-validation sampling strategy involved 50 replications of random re-sampling of the whole population split into 90% training set ($\boldsymbol{TS}, n = 1425$) and 10% validation set ($\boldsymbol{VS}, n = 150$). This sampling strategy mimics the situation were 90% of the population is phenotyped and genotyped, and 10% is genotyped but without phenotypes. Genomic estimated breeding values (GEBV) of individuals in the validation set (genotyped individuals without phenotypes) were predicted by solving

Henderson's linear mixed model equations in GBLUP. The variance components were fixed in the cross-validation analysis. The accuracy of the genomic selection models in the validation set was determined as the correlation ($r_{GEBV:EBV}$) between their GEBV and their EBV, whereas the predictive accuracy of the training set was determined as the correlation ($r_{GEBV:DGV}$) of their GEBV and their DGV.

2.3.6 Genomic selection efficiency

Genomic selection efficiency determines the selection response effectiveness by adopting a GS breeding strategy versus the TB strategy. The selection response for GS is presented as the ratio between the selection accuracy ($r_{GEBV:EBV}$) of the GS strategy and its breeding cycle time (t_{GS}), whereas the selection response for TB is obtained as the ratio between the breeding value prediction accuracy and its breeding cycle time (t_{TB}). The relative GS efficiency, in turn, was calculated as the ratio between GS and TB selection responses (Grattapaglia et al., 2011).

$$GSE = \left(\frac{r_{GEBV:EBV}/t_{GS}}{\sqrt{1-\left(SE^2/(1+F)\sigma_u^2\right)}/t_{TB}} \right) \qquad (5)$$

A sliding scale of GS breeding cycle times from three to nine years was used to simulate and compare the relative GS efficiency to that of the TB cycle, which takes eight years to complete for *E. grandis*.

2.3.7 Genetic gains

Variance components were fixed when solving the linear mixed models to estimate EBV, DGV and GEBV for genetic gain calculations on the same scale. The expected genetic gain (%) of the traits was estimated by selecting the top 10% of individuals from the population expected to contribute to the next generation.

$$\%\boldsymbol{G} = \left(\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{BV}_{\boldsymbol{ij}} / \boldsymbol{\mu_i}\right) * \mathbf{100} \qquad \textbf{(6)}$$

where $\boldsymbol{\mu_i}$ is the population mean for trait $\boldsymbol{i}$ and $\boldsymbol{BV_{ij}}$ is the breeding value for individual $\boldsymbol{j}$ for trait $\boldsymbol{i}$. GS and TB genetic gains were simulated over 17 years, accounting for all breeding operations, including nursery activities **(Figure 2.7)**. We assumed a fixed genetic gain of 10% and that the performance of GS is the same in every breeding cycle. The proposed 10% increase in genetic gains every breeding cycle corresponds with empirical generational gains of *E. pellita* (Leksono et al., 2008) and *E. grandis* (Verryn et al., 2009). The ratio of the expected genetic gains accumulated over the 17 years for GS versus TB should indicate the genetic gains benefit per unit time of GS over TB.

2.4 RESULTS

2.4.1 Genome-wide SNP markers

The EUChip60K SNP chip generated 64,639 SNP markers (60,904 unique and 3735 replicated). Of these, about 15,040 informative SNP markers (MAF > 0.05 and present 90% of individuals in the *E. grandis* study population) were retained with 0.7% alleles missing and imputed, assuming Hardy–Weinberg equilibrium. The distribution of the 15,040 SNP markers per chromosome ranged from 1076 (chromosome 6) to 1962 (chromosome 10) **(Figure 2.1).**

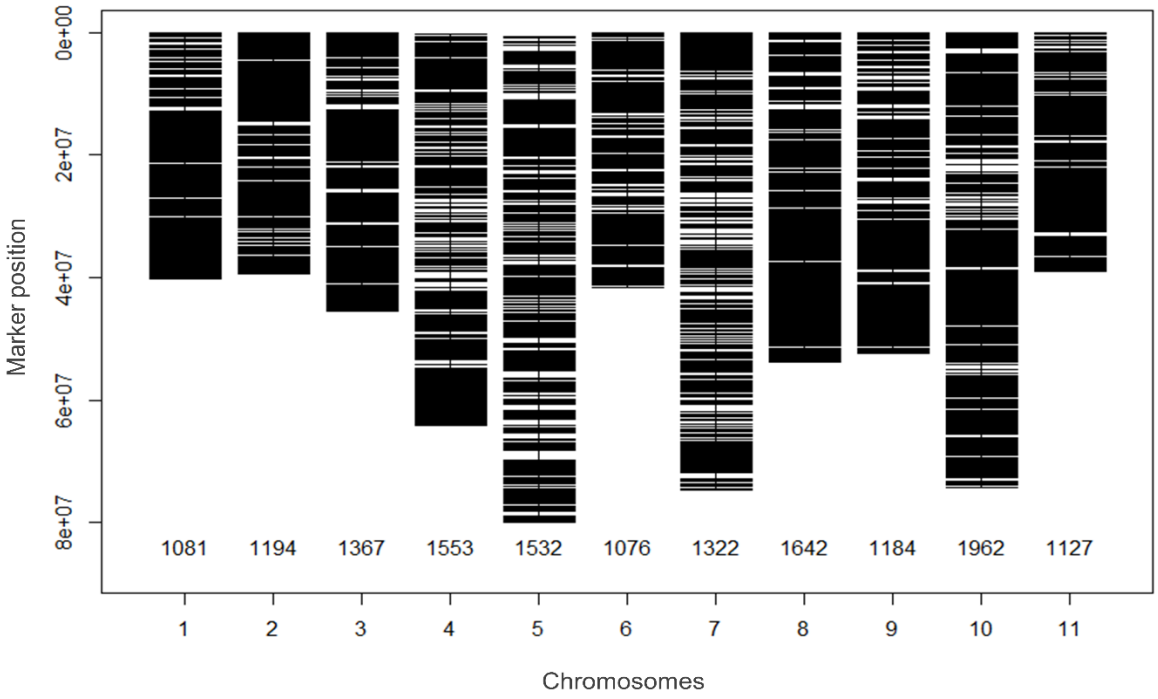


FIGURE 2.1 | SNP marker distribution across chromosomes and gene content frequency. The total number of markers is 15040, and the gene content frequency of the genotypes are AG (0.267), AA (0.303), GG (0.288), CC (0.07), AC (0.065), and NA's (0.716).

The version 1.0 genome assembly map of *E. grandis* was used to determine the EUChip60K SNP chip marker physical positions. These initial SNP marker positions included inaccurate assignment of some markers due to scaffolding errors later improved in version 2.0 (Bartholome et al., 2015). This is evident from the coverage plot of the 64,639 SNP markers on chromosomes 1, 3, 6 and 7 **(Figure 2.2C)** and the 15,040 retained SNP markers **(Figure 2.2D)** compared to gene positions on the version 2.0 assembly **(Figure 2.2B)**. The 15,040 polymorphic SNP markers have an average density of ~25 markers per Mb with average marker intervals of 41 kb ranging from 30 bp to 1.71 Mb, translating to an average of 1 SNP marker every ~2.76 genes. On average, the genetic linkage map of *E. grandis* has a resolution of 1 SNP marker for every 0.48 cM (~ 2 markers per cM) (Bartholome et al., 2015).

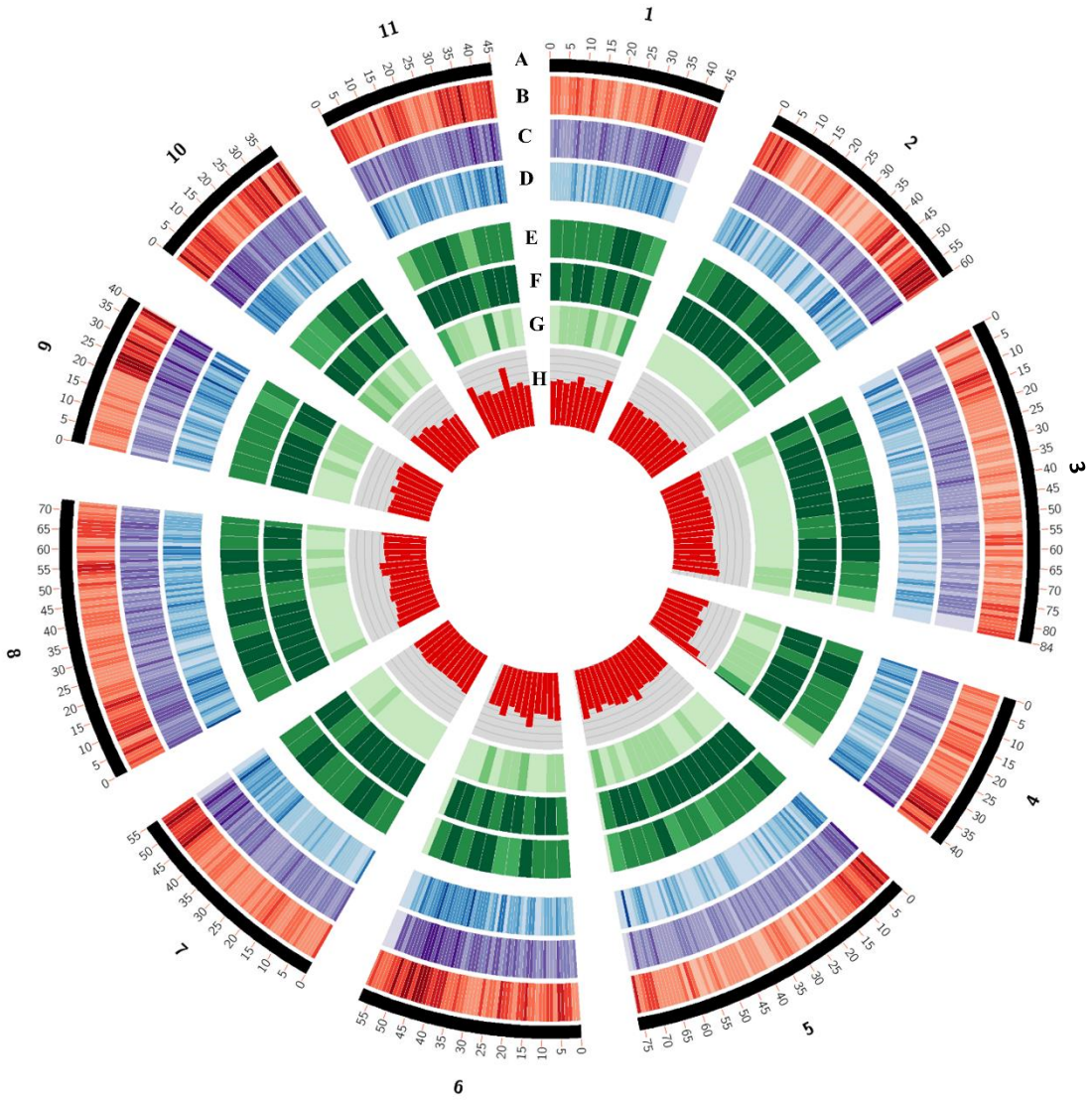
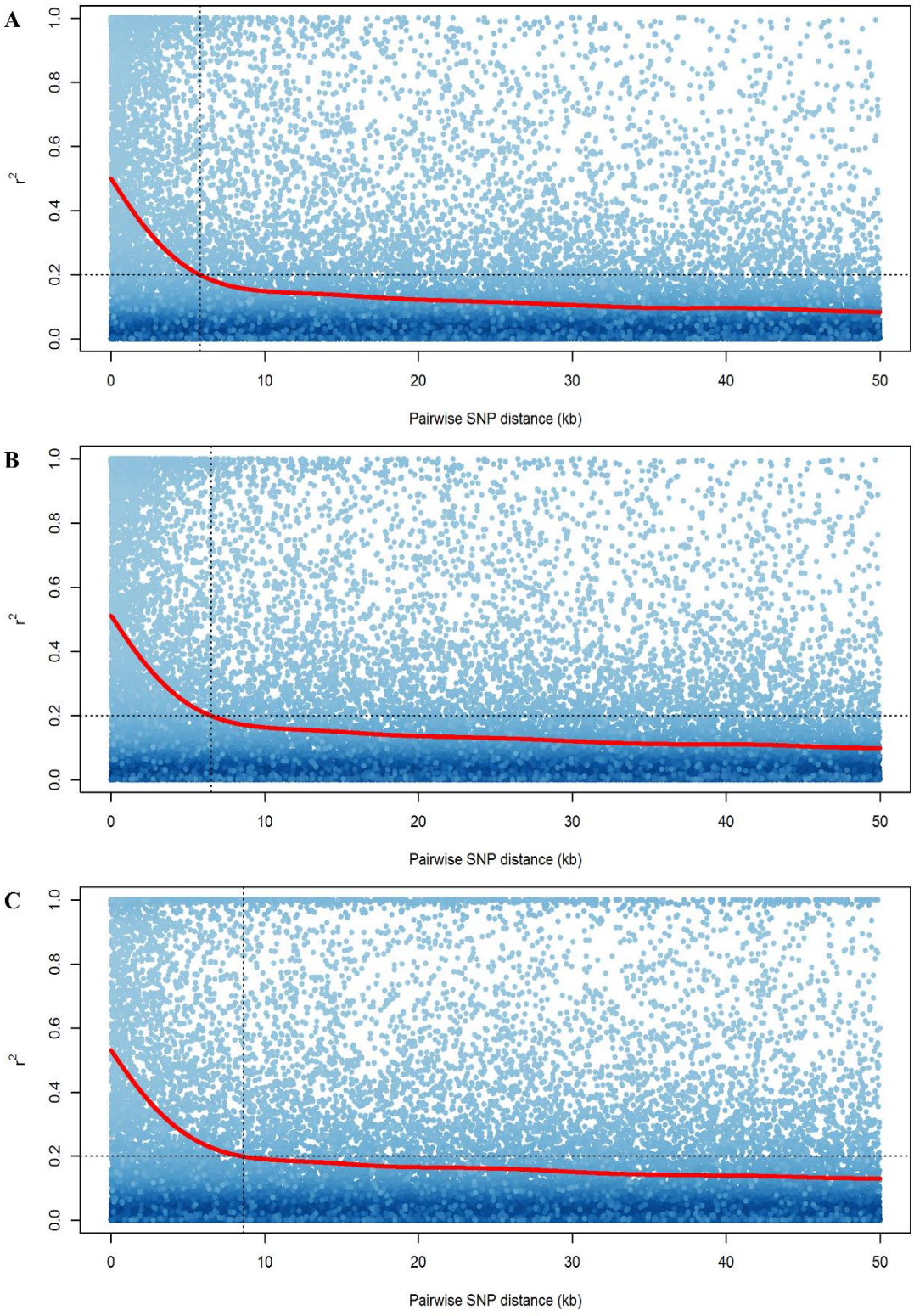


FIGURE 2.2 | Chromosome-scale overview of genomic features in *E. grandis*. **(A)** Physical genome assembly map of *E. grandis* version 2.0 in 5 Mb intervals across the 11 chromosomes. **(B)** Gene density heatmap (red) in 1 Mb intervals. **(C)** SNP density heatmap (purple) for the 64,639 SNPs derived from the *Eucalyptus* (EUChip60K) SNP chip in 1 Mb intervals. **(D)** Density heatmap (blue) of the filtered (for call rate > 0.90 and MAF > 0.05) 15,040 SNPs used in the study in 1 Mb intervals. Heatmaps (green) showing the frequency of marker pairs within 5 Mbp intervals with **(E)** LD of $0.2 < r^2 < 0.25$, **(F)** LD of $0.25 < r^2 < 0.6$, **(G)** LD of $r^2 > 0.6$, **(H)** is the mean LD per 5 Mbp interval.

2.4.2 Genome-wide linkage disequilibrium

The pairwise LD as the coefficient of determination (r^2) was estimated in the study population for all pairs of markers, resulting in 10,664,866 pairwise combinations with a mean $r^2 = 0.015$. The average genome-wide LD of the study population of marker pairs of distance within 50 kb is $r^2 = 0.18$, decaying to $r^2 < 0.20$ after 6.5 kb **(Figure 2.3B)**. The average genome-wide LD of the half-sib subpopulation of markers pairs of markers within 50 kb is $r^2 = 0.16$, decaying to $r^2 < 0.20$ within 5.8 kb **(Figure 2.3A)**, and for the full-sib subpopulation is $r^2 = 0.20$, decaying to $r^2 < 0.20$ within 8.6 kb **(Figure 2.3C)**. Variable LD is observed across the genome when considering 5 Mb intervals. A high proportion of the genome-wide LD ranges in the study population is considered medium-range LD ranging from 0.25 to 0.6, with a mean distance of 20.4 Mb **(Figure 2.2F)**. The mean distance of the low LD of 0.2 to 0.25 ranges from 13.1 Mb (chromosome 9) to 28.2 Mb (chromosome 5) **(Figure 2.2E)**, whereas the high LD (>0.6) ranges from 0.4 Mb (chromosome 2) to 2.3 Mb (chromosome 5) **(Figure 2.2G)**. The mean LD per 5 Mb bin aligned with the proportion of the marker pair in high LD are observed in chromosomes 4 and 11 **(Figure 2.2H).**



FGIRUE 2.3 | Linkage disequilibrium decay in the *E. grandis* study population. Genome-wide scatter plot of LD coefficient of determination (r^2) of SNP marker pairs within 50 kb distance of each other in the *E. grandis* genome. The decay curve (red) is the smoothed line of LD (r^2) generated with the *smooth.spline* R function. **(A)** Shows LD is decaying at $r^2 < 0.2$ after 5.8 kb for half-sib progeny, **(B)** LD is decaying at $r^2 < 0.2$ after 6.5 kb for all progeny, and **(C)** LD is decaying at $r^2 < 0.2$ after 8.6 kb for full-sib progeny.

2.4.3 Genetic parameters and relationship

The genomic relationship matrix was able to identify 63 selfed individuals from half-sib families with a coefficient of relationship greater than 0.75. The correction of these individuals in the pedigree resulted in a mean inbreeding coefficient increase from 1.0000 to 1.0041 **(Table 2.4)**. We observed lower than expected heritability estimates for growth and wood quality traits such as diameter (0.06), height (0.05) and cellulose (0.05) with the ABLUP models, with higher heritability estimates for wood quality traits such as S/G ratio (0.44) and fibre width (0.67) (**Table 2.5)**. The GBLUP models resulted in increased heritability estimates for both growth and wood quality traits, with the highest increase for wood density (0.18 to 0.33) and the lowest for S/G ratio (0.44 to 0.45) (**Table 2.5)**. However, the heritability for fibre width decreased from 0.67 to 0.58. Pairwise Pearson correlations of EBV for growth and wood quality, and their distributions, are presented in **Figure 2.4**. Diameter and height correlate at 0.87 and correlates with cellulose at 0.25 and 0.24, respectively, with basic density correlating with height at 0.25 and diameter growth at 0.37.

TABLE 2.4 | Summary statistics of the pedigree and realised genetic relationship matrices. The corrected pedigree includes the revision of 63 (7%) individuals identified as selfs in the half-sib families (coefficient of relationship > 0.75) based on the genomic relationship matrix.

	Minimum	Maximum	Mean
Uncorrected Pedigree			
Coefficient of relationship	0.0000	0.5000	0.0202
Inbreeding coefficient	1.0000	1.0000	1.0000
Corrected Pedigree			
Coefficient of relationship	0.000	1.0000	0.0199
Inbreeding coefficient	1.000	1.5000	1.0041
Genomic Relationship			
Coefficient of relationship	-0.2631	1.2085	-0.0008
Inbreeding coefficient	1.0884	1.6258	1.2324

TABLE 2.5 | Genetic variance components of the *E. grandis* study population. The additive genetic (σ_u^2), and residual (σ_e^2) variance components and narrow-sense heritability (h^2) estimates and their standard errors (se) for the growth and wood quality traits from the ABLUP and GBLUP models.

	ABLUP			GBLUP		
	σ_u^2	σ_e^2	$h^2(se)$	σ_u^2	σ_e^2	$h^2(se)$
Fibre length	0.001	0.003	0.17 (0.055)	0.001	0.003	0.22 (0.035)
Fibre width	0.774	0.382	0.67 (0.093)	0.652	0.465	0.58 (0.036)
Cellulose	0.190	3.362	0.05 (0.026)	0.481	3.084	0.13 (0.031)
S/G ratio	0.020	0.020	0.44 (0.085)	0.020	0.024	0.45 (0.039)
Density	249.283	1101.067	0.18 (0.075)	440.855	903.703	0.33 (0.040)
Diameter	1.048	16.520	0.06 (0.032)	1.597	16.037	0.09 (0.030)
Height	0.690	12.049	0.05 (0.026)	1.099	11.646	0.09 (0.027)

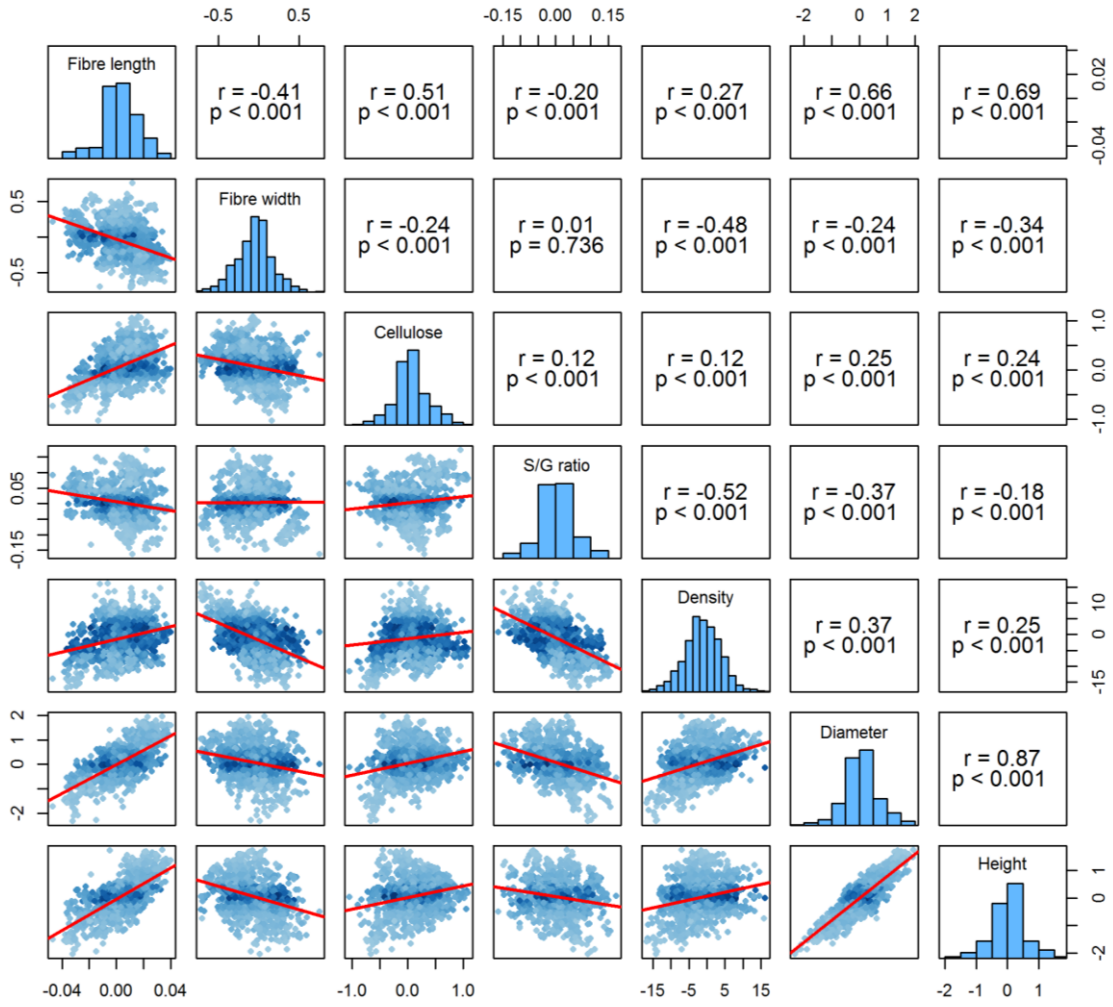


FIGURE 2.4 | Estimated breeding values (EBV) pairwise correlations of the growth and wood quality traits. The lower triangle shows the scatter plots between the traits with a fitted linear regression (*red*), the distribution of the traits are shown on the diagonal, and the upper triangle provides the Pearson correlations (*r)* between the traits (H_0: $r = 0$).

2.4.4 Genomic selection accuracy and expected genetic gains

The use of the genomic relationship matrix resulted in an average increase of 7% in the breeding value prediction accuracy of DGV (ranging from 0.81 to 0.94) compared to the EBV (ranging from 0.76 to 0.87) across all traits **(Table 2.6)**. This resulted in an average increase of 24% in the expected genetic gain of the growth and wood quality traits for DGV compared to EBV **(Figure 2.5)**. This suggested that using a genomic realised relationship is efficient to capture true genetic relationships. The genetic gains of GEBV are on average 7% lower compared to DGV; however, still higher than the EBV across all traits except for cellulose and diameter, which were marginally lower **(Figure 2.5)**. The accuracy of the GS model ($r_{GEBV:EBV}$) was calculated as an average of the 50 cross-validation folds ranging from 0.54 (density) to 0.67 (fibre width), whereas the prediction accuracy of the training set of the GS model ($r_{GEBV:DGV}$) ranged from 0.86 (diameter) to 0.98 (S/G ratio) **(Figure 2.6)**.

TABLE 2.6 | The average prediction accuracy (r) of the estimated breeding value (EBV) and direct genetic value (DGV) for the growth and wood quality traits. The range of the prediction accuracy and the standard deviations are presented.

	r_{EBV}			r_{DGV}		
	Accuracy	**Min-Max**	**Std.dev**	**Accuracy**	**Min-Max**	**Std.dev**
Fibre length	0.81	0.71-0.96	0.039	0.87	0.77-0.97	0.030
Fibre width	0.92	0.71-0.99	0.015	0.94	0.84-0.99	0.016
Cellulose	0.76	0.71-0.92	0.034	0.84	0.71-0.96	0.034
S/G ratio	0.87	0.71-0.99	0.024	0.92	0.81-0.98	0.021
Density	0.81	0.71-0.97	0.038	0.89	0.79-0.98	0.026
Diameter	0.77	0.71-0.92	0.035	0.82	0.66-0.94	0.037
Height	0.76	0.71-0.92	0.035	0.81	0.65-0.94	0.037

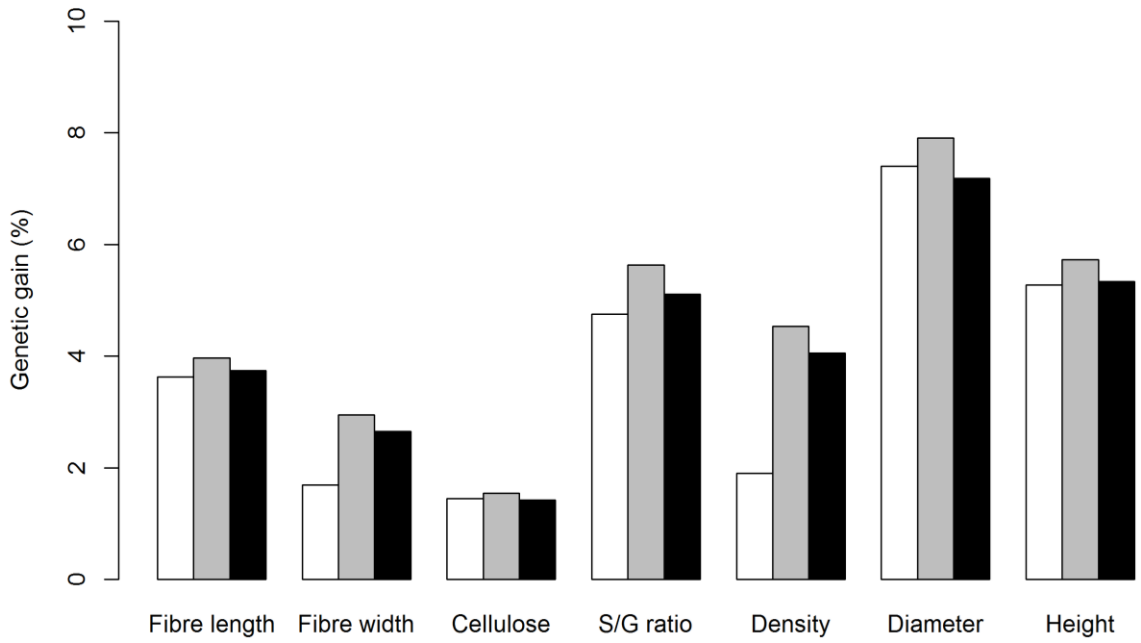


FIGURE 2.5 | Percentage genetic gains for growth and wood quality traits. The gains were estimated from selecting the top 10% of individuals from the EBV (*white*), DGV (*grey*) and GEBV (*black*) predictions for the individual traits. The variance components from the models were fixed in solving the mixed models.

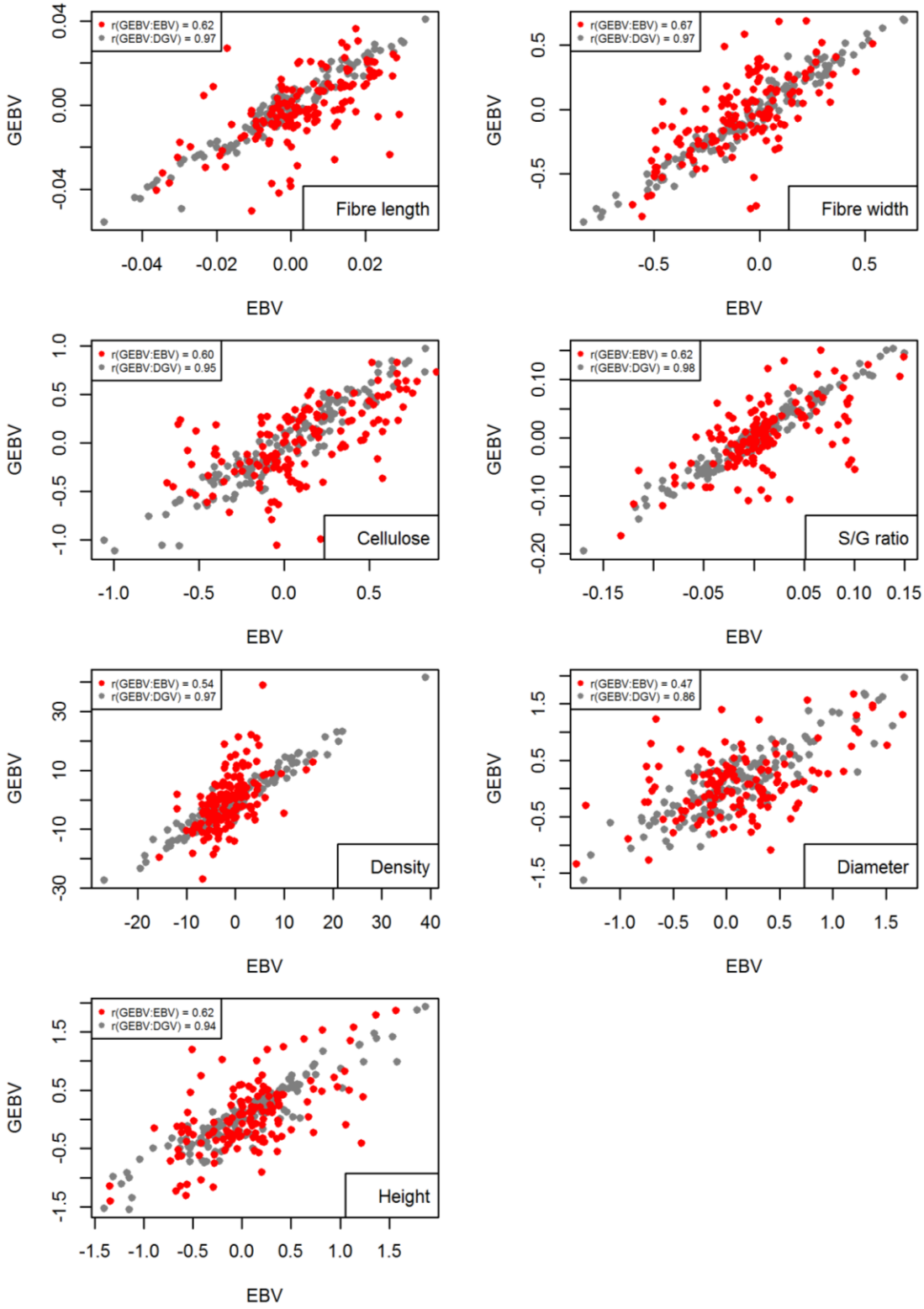


FIGURE 2.6 | Genomic selection accuracy (r) scatter plots for growth and wood quality traits for *E. grandis.* Depicted by the *grey dots* is the relationship between DGV (*x-axis*) and GEBV (*y-axis*), demonstrating the accuracy of the training set, while *red dots* show the relationship between EBV (*x*-axis) and the GEBV (*y-axis*), demonstrating the accuracy of the

validation set.

2.4.5 Genomic selection efficiency and genetic gains

Compared to the TB cycles, the relative efficiency of the GS breeding cycles over 17 years, considering all operational breeding steps in the breeding cycles, showed that for *E. grandis*, we could complete two eight-year conventional breeding cycles and three four-year GS breeding cycles in the same period, including a full-sib clonal trial to validate and update the GS model **(Figure 2.7)**. For *E. grandis*, a four-year GS breeding cycle necessitates flower induction treatments in a non-trials environment to enable controlled pollination between GEBV-selected individuals. With a four-year GS breeding cycle (seed-to-seed), the relative efficiency of the GS strategy was higher than an eight-year TB breeding cycle, ranging from 1.20 (wood density) to 1.62 (fibre length) **(Figure 2.8)**. Note that GS predictions do not consider the actual genetic and phenotypic correlations of the traits at half-rotation (4 years) versus full-rotation (8 years) because the training and updating of the GS model will use full-rotation phenotype data for implementation **(Figure 2.7)**.

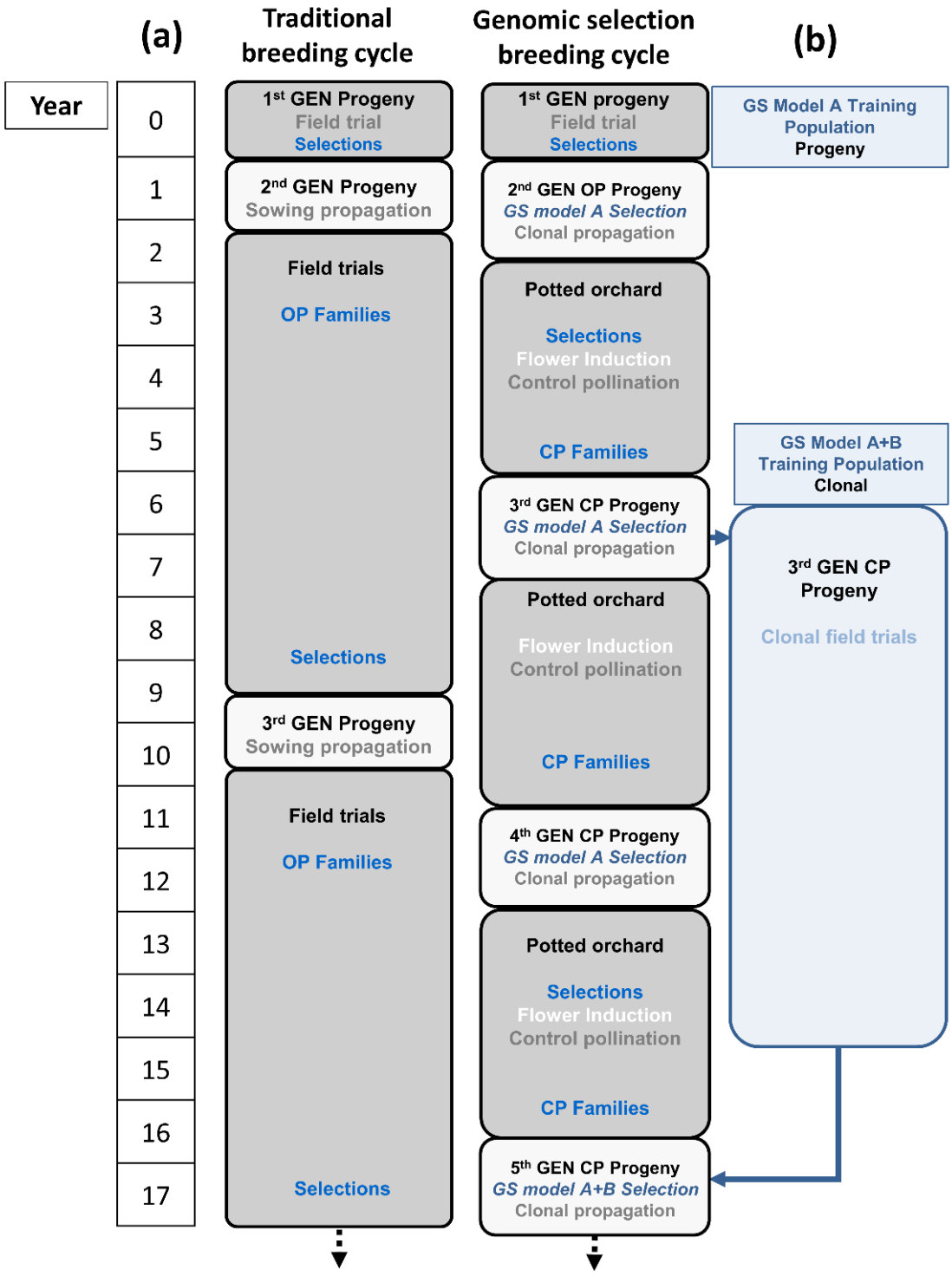


FIGURE 2.7 | Comparison of traditional and genomic selection breeding cycles over 17 years for *E. grandis*. (a) Two complete traditional eight-year TB cycles and (b) three complete four-year GS breeding cycles with a full rotation full-sib clonal trial to validate and update the current GS model. GS model A predictions are performed in years 1, 6, and 12, with the updated model (GS model A+B) from the control pollinated clonal progeny field trials used for prediction in year 17.

However, should the GS strategy take five years to complete its breeding cycle, then the relative efficiency for traits was lower, ranging from 1.17 (fibre width) to 1.29 (fibre length), with wood density less efficient at 0.96 **(Figure 2.8)**. The relative efficiency of GS breeding compared to TB diminishes with increasing GS breeding cycle time, underscoring the significant effect of reproductive biology (seed-to-seed) on the feasibility of implementing GS in *E. grandis*.

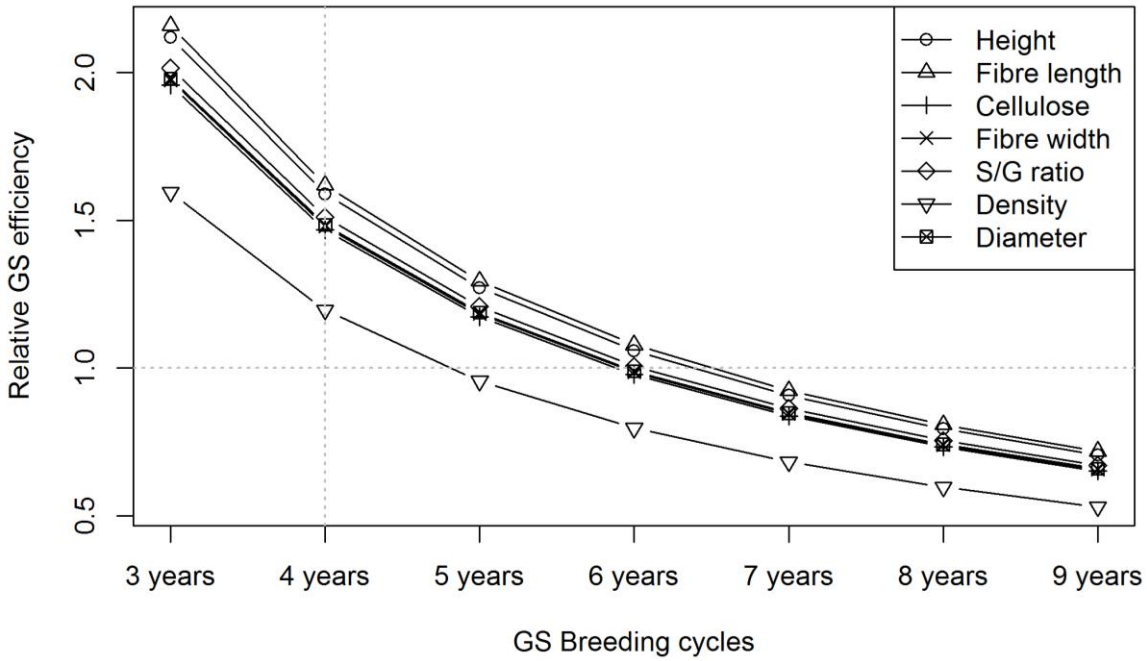


FIGURE 2.8 | The relative efficiency of GS over time for the growth and wood quality traits compared to the eight-year TB cycle. GS efficiency decreases with the increase of its breeding cycle times. The *grey dotted line* indicates the intercept between the earliest time point (*x-axis*) to achieve reproductive maturity with flowering induction treatments under non-trial conditions and when the 8-year TB cycle is as efficient as the GS breeding (*y-axis*). The relative efficiency of GS is better than the TB, then the *y-axis* is >1.0, and when GS is less efficient than the TB, then the *y-axis* is <1.0.

The accumulated genetic gains as a benefit of the TB strategy over the 17 years ranged from 3.0% (cellulose) to 15.5% (diameter), whereas for the GS strategy, it ranged from 4.7%

(cellulose) to 23.8% (diameter) **(Table 2.7)**. The genetic gains ratio, which is the benefit of GS compared to the TB over the 17 years, ranged from 1.53 (diameter) to 3.35 (wood density) **(Table 2.7)**, suggesting that there is an improved benefit in genetic gains over the 17 years with GS breeding compared to a TB approach in *E. grandis.*

TABLE 2.7 | The ratio of the genetic gains of genomic selection (GS) compared to traditional breeding (TB) accumulated over 17 years. The simulated conditions over the 17 years are a constant 10% increase in the percentage of genetic gain every breeding cycle and similar performance of GS Model A every breeding cycle.

Traits	Traditional breeding (TB)			Genomic selection (GS)				Genetic gain ratio GS/TB
	Cycle1	Cycle2	Total	Cycle1	Cycle2	Cycle3	Total	
Fibre length	3.6	4.0	**7.6**	3.7	4.1	4.5	**12.4**	1.62
Fibre width	1.7	1.9	**3.6**	2.7	2.9	3.2	**8.8**	2.47
Cellulose	1.5	1.6	**3.0**	1.4	1.6	1.7	**4.7**	1.55
S/G ratio	4.8	5.2	**10.0**	5.1	5.6	6.2	**16.9**	1.69
Density	1.9	2.1	**4.0**	4.1	4.5	4.9	**13.4**	3.35
Diameter	7.4	8.1	**15.5**	7.2	7.9	8.7	**23.8**	1.53
Height	5.3	5.8	**11.1**	5.3	5.9	6.5	**17.7**	1.60

2.5 DISCUSSION

2.5.1 Deviations from Hardy-Weinberg equilibrium

The mean pedigree inbreeding coefficient adjusted slightly from 1.0 to 1.0041 after correcting 63 (7%) selfed individuals based on the genomic relationship matrix, suggesting a slight disruption of this study population from the Hardy-Weinberg equilibrium principles. Other slight disruptions include two generations of half-sib recurrent artificial selection and non-random mating (control pollinated) of 46% of the study population. The selfing rate is lower than the expected range of 10 – 38% (van Wyk, 1981). The above suggests that the current TB strategy, at least over two generations of recurrent artificial selection, has not disrupted the equilibrium of allelic and genotypic frequencies in this population.

2.5.2 Linkage disequilibrium pattern

At long genetic distances, LD reflects population genetic factors such as breeding systems, geographic population substructures, and effective population size. In contrast, it reflects recent population genetic factors that cause gene-frequency evolution at short genetic distances, such as natural selection, gene conversion, mutation, and genetic drift (Slatkin, 2008). We observed a slightly higher level of LD (decaying after 6.5 kb) compared to that observed in a wild *E. grandis* population (decaying within 5.7 kb) using the same dense SNP marker genotypes (Silva-Junior and Grattapaglia, 2015a). The higher LD decay distance levels are expected since our study population has experienced artificial selection and comprised 46% non-random mating individuals. However, the half-sib subpopulation shows a similar LD decay distance of 5.8 kb as the wild *E. grandis*, even after two generations of half-sib recurrent selection. We suggest that the overall amount of LD in this *E. grandis* half-sib subpopulation has not shifted from the wild population. However, the introduction of non-random mating has managed to increase the LD decay distance (8.6 kb) of the full-sib subpopulation. Based on the average marker distance of 41 kb and LD decay at 6.5 kb, a substantial proportion of the genome is not in LD, even though LD extends over much larger distances in some regions of the genome.

2.5.3 LD influences of GS model performance

High LD facilitates the tagging of causative QTLs by increasing the likelihood of marker-QTL association, ensuring reliable and unbiased GEBV predictions (Villanueva et al., 2005; Grattapaglia et al., 2011; Habier et al., 2013). In barley, GBLUP predictions were demonstrated to rely more on the genomic relationship than LD, particularly over few generations (Zhong et al., 2009), whereas Bayesian models were more suited to capture both LD and genomic relationship over many generations when higher LD was observed in the population (Habier et al., 2007; et al., 2010). In forest populations with low LD, GBLUP derived from low-density

markers achieves similar predictive ability as Bayesian models (Zapata-Valenzuela et al., 2013; Isik et al., 2015) as well as with high-density markers (Duran et al., 2017; Tan et al., 2017).

2.5.4 Genomic selection in practice

Genomic selection benefits are well documented, foremost by accelerating breeding cycles and improving breeding value accuracy. Genome-wide DNA markers have also enabled better estimation of genetic parameters than pedigree records (Klapste et al., 2014). The sustained decrease in genotyping cost means that more innovative and complex strategies can be explored with GS approaches pursuing different breeding objectives (Grattapaglia et al., 2018). Such strategies may include regular updating of GS models to ensure their continued efficacy across generations, improve marker-trait association accuracy, and develop GS models that perform across multiple and changing environments. The advantage of GBLUP is that it is computationally less demanding than other whole-genome regression models. It allows the inclusion of the experimental design factors and the G×E interactions in the model and the modelling of variance-covariance structures to account for the heterogeneity (Isik et al., 2015). This study showed a healthy relationship (upwards of 0.98) between the GEBV and DGV, albeit with deflated predictions resulting in the lower percentage genetic gains **(Figure 2.6)**. We also investigated the practical implications of implementing GS strategies in an operational *E. grandis* breeding programme under realistic conditions, including the challenges of using individual trees from unbalanced trials as opposed to a clonally replicated training population. The study is unique because it demonstrates in practice the actual operational breeding cycles required for implementing GS strategies compared to TB strategies. For this to happen, vital adjustments to operational breeding steps are required to accommodate and realise the benefits of adopting a GS breeding strategy for *E. grandis*.

2.5.5 Genetic parameters

In the training population, the heritability estimates of diameter (0.05) and height (0.06) were lower than expected for *E. grandis*. Harrand et al. (2009) previously reported heritability estimates of 0.16 and 0.14 for diameter and height, respectively, in *E. grandis*. The low heritability estimates observed in this study could be attributed to experimental design inefficiency due to the unbalanced trials, resulting in increased residual errors. Nevertheless, the use of the genomic relationship matrix resulted in higher and more precise heritability estimates **(Table 2.5)**, higher breeding value prediction accuracies **(Table 2.6)** and higher genetic gains **(Figure 2.7)** for the growth and wood quality traits compared to the pedigree relationship matrix.

2.5.6 Predictive ability

The GS model prediction accuracy for the growth and wood quality traits ranged from 0.47 to 0.67 **(Figure 2.6)**. The accuracies shown are similar to those obtained in GS studies of other forest trees, for example, growth and stem-form in maritime pine (*Pinus pinaster*) ranging from 0.54 to 0.65 (Bartholome et al., 2016) and from 0.39 to 0.49 (Isik et al., 2015). Similar accuracies were also observed in the selection of *E. globulus* clones density (0.60) and volume (0.73) (Duran et al., 2017). The prediction accuracy of disease, growth and wood quality traits of *P. taeda* breeding range from 0.20 to 0.46 (Resende Jr et al., 2012). The genetic gains from the GEBV were lower than the DGV but higher than that of the EBV **(Figure 2.5)**. The genetic gains from GEBV are lower than those from the DGV because of the lower GS model prediction accuracy but higher than that of the EBV because of the improved average breeding value prediction accuracy (ranging from 0.81 to 0.87) and heritability estimates associated with the genomic relationship matrix. Together, these results suggested that there is sufficient GS

prediction accuracy as well as improved genetic gains over pedigree to justify the implementation of GS strategies in *E. grandis* breeding.

2.5.7 Genomic selection efficiency and the benefit to genetic gains

The relative efficiency of adopting the GS breeding strategy in *E. grandis* was investigated using simulation. We considered the operational breeding steps and the respective breeding cycle times required for a current TB strategy versus the proposed GS breeding strategy **(Figure 2.8)**. One TB cycle takes eight years (full-rotation) for *E. grandis.* There is a strong correlation between growth and wood quality trait measurements at half-rotation vs full-rotation (Osorio et al., 2003; Wu et al., 2007; Luo et al., 2010; Rweyongeza, 2016). Age-age genetic correlations higher than 0.90 have been reported in *E. nitens* for growth traits such as heights, diameter and volume, and basic density (Greaves et al., 1997). Although *E. grandis* can reach reproductive maturity at four years (half-rotation) under experimental (field trial) conditions, the proportion of these individuals with flowers is small. Therefore, there is a need to wait until year eight to increase the percentage of flowering individuals so that most of the chosen selections have seeds to turn over the generation.

The GS breeding cycle should take four years incorporating all the necessary adjustments to advance GEBV selected seedlings into the next generation **(Figure 2.7)** because, under non-trial conditions, flowering precocity in *Eucalyptus* species can be enhanced with growth regulators such as paclobutrazol (Griffin et al., 1993; Hasan and Reid,1995; Williams et al., 2003; Gardner and Bertling, 2005). In the case of *E. grandis*, flowering can be induced well within four years. This means that a GS strategy can be implemented with a plausible four-year seed-to-seed breeding cycle. Therefore, together with accelerated flowering, the GS strategy can produce predicted GEBV seedlings in the nursery, unlike the TB approach, where seedlings from early flowering may not have EBV. A four-year GS breeding cycle would represent a

50% reduction compared to the TB cycle as simulated by (Grattapaglia et al., 2011). The operational adjustments required for GS approaches will differ when considering different species, breeding objectives, reproductive biology, propagation requirements and performances of seedlings vs clones to realise the expected accelerated gains per unit time (Isik, 2014; Resende et al., 2017; Li and Dungey, 2018). In our projection, the four-year GS breeding cycle will result in a relative efficiency ranging from 1.20 to 1.62 for the growth and wood quality traits compared to the eight-year TB breeding cycle of *E. grandis* **(Figure 2.7)**. Thus far, the GS breeding study has demonstrated encouraging GS prediction accuracy for growth and wood quality traits, higher genetic gains, and higher efficiency than the TB strategy. Over the 17 years, the predicted ratio of the accumulated genetic gains for the GS and TB cycles range from 1.53 to 3.35, demonstrating the benefit of GS in increased gains per unit time (**Table 2.7**). There is a realistic expectation that, despite the need for repeated genotyping over the generations to update the GS model, the total cost of genotyping will reduce as high-throughput technology evolves over the 17 years, approaching zero compared to phenotyping cost, DNA isolation and operational expenses including infrastructure maintenance that will increase with inflation over the period. Our results suggest that a GS breeding strategy can be implemented for *E. grandis* and is likely to realise the expected genetic gains per unit time. However, it will be essential to manage the significant adjustments and additional cost of operational breeding steps required to accommodate the GS strategy, as these may increase the cost per unit genetic gain in the short term but will be offset in the longer term by increased profits at the mill.

2.5.8 Implementation of a GS strategy

The increased cost per unit genetic gain does come with additional benefits in that the TB and GS breeding strategies can advance two separate but complementary crucial breeding objectives, which are to maintain genetic diversity and accelerate genetic gains, respectively. Current breeding strategies combine these two breeding objectives, mainly to maximise

resource allocation. However, when the two objectives are separated, proper attention can be given to maximize their respective outputs. Compared to the two open-pollinated TB cycles, the proposed three control-pollinated GS breeding cycles will more rapidly accumulate favourable alleles. The control pollination strategy embedded in the GS breeding approach should allow mating designs that maximise gains while limiting inbreeding, increase population LD, and allow controlled infusion of genotypes from the TB breeding cycle into the GS breeding cycle. Controlled pollination is essential for developing desirable training populations for GS in that it confines allelic diversity within the training population (Habier et al., 2007; Rutkoski et al., 2015). The rapid development of a deep full-sib pedigree within a GS strategy has the potential to surpass the benefit provided by genome-wide markers in predicting genetic merit (Bartholome et al., 2016). The replicated clonal full-sib progeny field trials running concurrently with the GS approach simulated in this study will update and validate the GS model for G×E interactions. We have only highlighted the breeding implications of GS in *E. grandis* in this study. However, it is essential to note that the GS strategy will realize commercial products at least four years earlier than the TB approach, offering an opportunity to recoup some of the investments required earlier for GS practical implementation.

2.6 CONCLUSION

The use of genome-wide DNA markers presents multiple advantages over shallow pedigree. These benefits include more accurate estimates of genetic relationships among individuals, higher and more precise heritability estimates, improved breeding value prediction accuracy and increased genetic gains. The study population is not far removed from the wild in its LD pattern, even when subjected to half-sib recurrent selection over three generations or one generation of non-random mating. Significant practical adjustments to the TB cycle are

required to realise the GS breeding strategy efficiency in *E. grandis*. Enhanced benefits in the form of gains per unit time are achieved through the shortened operational breeding cycle of GS, mainly by overcoming the reproductive limits such as the time from unimproved seed to improved seed.

Looking forward, GS will become a common practice that will provide breeders with much more critical information to achieve breeding goals and produce elite clones for deployment. More sophisticated analysis models will be needed to help breeders accurately compute all of the information gained from interactions between the genome, phenome and the environment. Practical studies like this are necessary for tree breeding programs to inform breeding objectives and strategic decisions for tree breeders and managers.

2.7 REFERENCES

Agarwal, M., Shrivastava, N., and Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27**,** 617-631.

Bartholome, J., Mandrou, E., Mabiala, A., Jenkins, J., Nabihoudine, I., Klipp, C., Schmutz, J., Plomion, C., and Gion, F-M. (2015). High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytologist* 206**,** 1238-1296.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., and Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science* 242**,** 23-36.

Bartholome, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., and Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. *BCM Genomics* 17**,** 604.

Beaulieu, J., Doerksen, T., Clement, S., Mackay, J., and Bousquet, J. (2014). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* 113**,** 343-352.

Booth, T. H. (2013). Eucalypt plantations and climate change. *Forest Ecology and Management* 301**,** 28-34.

Bouquet, A., and Juga, J. (2013). Integrating genomic selection into dairy cattle breeding programmes: a review. *Animal* 7**,** 705-713.

Butler, D. G., Cullis, B. R., Gilmour, A. R., and Goqel, B. J. (2009). "ASReml-R reference manual").

Cappa, E. P., De Lima, B. M., Da Silva-Junior, O. B., Carcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in

Eucalyptus using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Science* 284**,** 9-15.

Carle, J., and Holmgren, P. (2008). Wood from planted forests a global outlook 2005 - 2030. *Forest Product Journal* 58**,** 6-18.

Christensen, O. F. (2012). Compatability of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genetics Selection Evolution* 44, 37

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., and Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *The Plant Genome* 11**,** 1-14.

Cros, D., Mbo-Nkoulou, L., Bell, J. M., Oum, J., Masson, A., Soumahoro, M., Tran, D. M., Achour, Z., Guen, V. L., and Clement-Demanage, A. (2019). Within-family genomic selection in rubber tree (*Hevea brasiliensis*) increases genetic gain for rubber production. *Industrial Crops and Products* 138**,** 1-13.

Crossa, J., de Los Campos, G., Maccaferri, M., Tuberosa, R., Burgueno, J., and Perez-Todriguez, P. (2017). Extending the marker x environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. *Crop Science* 56**,** 2193-2209.

Duran, R., Isik, F., Zapara-Valenzuela, J., Balocchi, C., and Valanzuela, A. (2017). Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genetics & Genome* 13, 74.

Garcia-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-Lopez, F. J., and Van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *PNAS* 113**,** E4928.

Gardner, R. A. W., and Bertling, I. (2005). Effect of winter chilling and paclobutrazol on floral bud production in *Eucalyptus nitens*. *South African Journal of Botany* 71**,** 238-249.

Gilmour, A. R., Gogel, B.J., Cullis, B.R., Welham, S.J., and Thompson, R. (2015). "ASReml guide release 4.1 functional specifications". 4.1 ed. (Hemel Hempstead, HP1 1ES, UK: VSN International Ltd (www.vsni.co.uk)).

Grattapaglia, D., Macos, D., and Resende, D.V. (2011). Genomic selection in forest tree breeding. *Tree Genetics and Genomes* 7**,** 241-255.

Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Muller, B. S. F., Tan, B., Isik, F., Ratcliffe, D., and El-Kassaby, Y. A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science* 9**,** 1693.

Greaves, B. L., Borralho, N. M. G., Raymond, C. A., Evans, R., and Whiteman, Ph. (1997). Age-age correlations in, and relationships between basic density and growth in *Eucalyptus nitens*. *Silvae Genetica* 45**,** 264-270.

Griffin, A. R., Whiteman, P., Rudge, T., Burgess, I. P., and Moncur, M. (1993). Effect of paclobutrazol on flower-bud production and vegetative growth in two species of *Eucalyptus*. *Canadian Journal of Forest Research* 23**,** 640-647.

Habier, D., Fernando, R. L., and Dekker, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177**,** 2389-2397.

Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP decoded: A look into the black box of genomic prediction. *Genetics* 194**,** 597-607.

Habier, D., Tetens, J., Seefried, F-R., Lichtner, P., and Thaller, G. (2010). The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution* 42, 5.

Haile, J. K., N'diaye, A., Clarke, F., Clarke, J., Knox, R., Rutkoski, J., Bassi, F. M., and Pozniak, C. J. (2018). Genomic selection for grain yield and quality traits in durum wheat. *Molecular Breeding* 38**,** 78.

Harrand, L., Hernandez, J. J. V., Upton, J. L., and Valverde, G. R. (2009). Genetic parameters of growth traits and wood density in *Eucalyptus grandis* progenies planted in Argentina. *Silvae Genetica* 58**,** 1-2.

Hasan, O., and Reid, J .B. (1995). Reduction of generation time in *Eucalyptus globulus*. *Plant Growth Regulation* 17**,** 53-60.

Hayes, B. J., Bowman, P. J., Camberlain, A. J., and Goddard, M. E. (2008). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92**,** 433-443.

Hayes, B. J., Visscher, P.M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research* 91**,** 47-60.

Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38**,** 226-231.

Hill, W.G., and Weir, B. S. (2010). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetic Resources* 93**,** 47-64.

Irland, L. C., Adams, D., Alig, R., Betz, C.J., Chen, C-C., Hutchins, M., Mccarl, B.A., Skog, K., and Sohngen, B. L. (2001). Assessing socioeconomic impacts of climate change on US forests, wood-product markets, and forest recreation: The effects of climate change on forests will trigger market adaptations in forest management and in wood-products industries and may well have significant effects on forest-based outdoor recreation. *BioScience* 51**,** 753-764.

Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests* 45**,** 379-401.

Isik, F., Bartholome, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C., and Bouffier, L. (2015). Genomic selection in maratime pine. *Plant Science* 242**,** 108-119.

Isik, F., Holland, J., and Maltecca, C. (2017). *Genetic data analysis for plant and animal breeding.* Springer International Publishing. ISBN 978-3-319-55177-7.

Klapste, J., Lstiburek, M., and El-Kassaby, Y. A. (2014). Estimates of genetic parameters and breeding values from western larch open-pollinated families using marker-based relationship. *Tree Genetics & Genomes* 10**,** 241-249.

Klapste, J., Suontanna, M., Dungey, H. S., Telfer, E. J., Graham, N. J., Low, C. B., and Stovold, G. T. (2018). Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. *Journal of Heredity* 109**,** 802-810.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92**,** 4656-4663.

Leksono, B., Kurinobu, S., and Ide, Y. (2008). Realized genetic gains observed in second generation seedling seed orchards of *Eucalyptus pellita* in Indonesia. *Journal of Forest Research* 13**,** 110-116.

Li, Y., and Dungey, H. S. (2018). Expected benefit of genomic selection over forward selection in conifer breeding and deployment. *PLOS One* 13**,** e0208232.

Luan, T., Woolliams, J. A., Lien, S., Kent, M., Svendsen, M., and Meuwissen, T. H. E. (2009). The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. *Genetics* 183**,** 1119-1126.

Luo, J., Zhou, G., Wu, B., Chen, D., Cao, J., Lu, W., Pegg, R. E., and Arnold, R. J. (2010). Genetic variation and age-age correlations of *Eucalyptus grandis* at Dongmen Forest Farm in southern China. *Australian Forestry* 73**,** 67-80.

Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. (2012). SNP markers and their impact on plant breeding. *International Journal of Plant Genomics* 2012**,** 11.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157**,** 1819-1829.

Misztal, I., Aggrey, S. E., and Muir, W. M. (2013). Experiences with a single-step genome evaluation. *Poultry Science* 92**,** 2530-2534.

Müller, B.S.F., Neves, L. G., De Almeida Filho, J. E., Resende, M. F. R., Muñoz, P. R., Dos Santos, P. E. T., Filho, E. P., Kirst, M., and Grattapaglia, D. (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of *Eucalyptus*. *BMC Genomics* 18**,** 524.

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D.,Goodstein, D.M., Dubchak, I., Poliakov, A., Mizrachi, E., Kullan, A. R. K., Hussey, S.G., Pinard, D., Van Der Merwe, K., Signh, P., Van Jaarsveld, I., Silva-Junior, O.B., Togawa, R.C., Pappas, M.R., Faria, D.A., Sansaloni, C.P., Petroli, C.D., Yang, X., Ranjan, P., Tschaplinski, T.J., Ye, C-Y., Li, T., Sterck, L., Vanneste, K., Murat, F., Soler, M., San Clemente, H., Saidi, N., Cassan-Wanh, H., Dunand, C., Hefer, C.A., Bornberg-Bauer, E., Kersting, A.R., Vining, K., Amarasinghe, V., Ranik, M., Naithani, S., Elser, J., Boyd, A.E., Liston, A., Spatafora, J.W., Dharmwardhana, P., Rija, R., Sullivan, C., Romanel, E., Alves-Ferreira, M., Kulheim, C., Foley, W., Carocha, V., Paiva, J., Kudrna, D., Brommonschenkel, S.H., Pasquali, G., Byrne, M., Rigault, P., Tibbits, J., Spokeviciuos, A., Jones, R.C., Steane, D.A., Vaillancourt, R.E., Potts, B.M., Joubert, F., Barry, K., Pappas Jr, G.J., Strauss, S.H., Jaiswal, P., Grima-Pettenati, J., Salse, J., Van De Peer, Y., Rokhsar, D.S., and Schmutz, J. (2014). The genome of *Eucalyptus grandis*. *Nature* 510**,** 356-375.

Namkoong, G., Barnes, R.D., and Burley, J. (1980). Screening for yield in forest tree breeding. *Commonwealth Forestry Review* 59**,** 61-68.

Osorio, L. F., White, T. L., and Huber, D. A. (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 (Article). *Theoretical and Applied Genetics* 106**,** 735-743.

Piepho, H.P., Mohring, J., Melchinger, and Buchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161**,** 209-228.

Ratcliffe, B., El-Dien, O. G., Cappa, E . P., Porth, I., Klápste, J., Chen, C., and El-Kassaby, Y. A. (2017). Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. *Genes Genomics Genetics* 7**,** 935-942.

Resende Jr, M. F. R., Munoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F., and Kirst, M. (2012). Accuracy of genomic selection methods in standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190**,** 1503-1510.

Resende, M. D. V., Resende Jr, M.F.R., Sansaloni, C.P., Petroli, C.D., Missiaggia, A.A., Aguiar, A.M., Abad, J.M., Takahashi, E.K., Rosado, A.M., Faria, D.A., Pappas Jr, G.J., Kilian, A., and Grattapaglia, D. (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* 193**,** 617-624.

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevero, C. F., Takahashi, E. K., Silva-Junior, O. B., and Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity* 119**,** 245-255.

Rutkoski, J., Singh, R. P., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J. L., and Sorrells, M. E. (2015). Efficient use of historical data for genomic selection: A case study of stem rust resistance in wheat. *The Plant Genome* 8, 1.

Rweyongeza, D. W. (2016). A new approach to prediction of the age-age correlation for use in tree breeding. *Annals of Forestry Science* 73**,** 1099-1111.

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal Animal Breeding and Genetics* 123**,** 218-223.

Shepherd, M., Bartle, J., Lee, D, J., Brawner, J., Bush, D., Turnbull, P., Macdonel, P., Simmons, B., and Henry, R. (2011). Eucalypts as a biofuel feedstock. *Biofuels* 2**,** 639-657.

Silva-Junior, O. B., and Grattapaglia, D. (2015a). Genome-wide patterns of recombination, linkage disequilibrium and nucleotide diversity from pooled resequencing and single nucleotide polymorphism genotyping unlock the evolutionary history of *Eucalyptus grandis*. *New Phytologist* 208**,** 830-845.

Silva-Junior, O. B., Faria, D.A., and Grattapaglia, D. (2015b). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206**,** 1527-1540.

Slatkin, W. (2008). Linkage disequilibrium - understanding the evolutionary past and mapping the future. *Nature Reviews Genetics* 9**,** 477-485.

Stanturf, J. A., Vance, E. D., Fox, T. R., and Kirst, M. (2013). *Eucalyptus* beyond its native range: Environmental issues in exotic bioenergy plantations. *International Journal of Forestry Research* 2013, ID 463030.

Suontama, M., Klápště, J., Telfer, E., Graham, N., Stovold, T., Low, C., Mckinley, R., and Dungey, H. (2019). Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* 122**,** 370-379.

Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., and Ingvarsson, P. K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BCM Plant Biology* 17**,** 110.

Van Wyk, G. (1981). Inbreeding effects in *Eucalyptus grandis* in relation to degree of relatedness *South African Forestry Journal* 116**,** 60-63.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91**,** 4414-4423.

Veerkamp, R. F., Mulder, H. A., Thompson, R., and M. P. L. Calus. (2011). Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *Journal of Dairy Science* 94**,** 4189-4197.

Verryn, S. D., Snedden, C. L., and Eatwell, K. A. (2009). A comparison of deterministically predicted genetic gains with those realised in a South African *Eucalyptus grandis* breeding program. *Southern Forests: a Journal of Forest Science* 71**,** 141-146.

Villanueva, B., Pong-Wong, R., Fernandez, J., and Toro, M. A. (2005). Benefits from marker-assisted selection under an additive polygenic genetic model. *Journal of Animal Science* 83**,** 1747-1752.

Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics* 132**,** 669-686.

White, T. L., and Hodge, G.R. (1988). Best linear prediction of breeding values in forest tree improvement programs. *Theoretical and Applied Genetics* 76**,** 719-727.

Wiggans, G. R., Cole, J. B., Hubbard, S. M., and Sontegard, T. S. (2017). Genomic selection in dairy cattle: THe USDA experience. *Annual Review of Animal Bioscieces* 5**,** 309-327.

Williams, D. R., Potts, B. M., and Smethurst, P. J. (2003). Promotion of flowering in *Eucalyptus nitens* by paclobutrazol was enhanced by nitrogen fertilizer. *Canadian Journal of Forest Research* 33**,** 74-81.

Wimmer, V., Albrecht, T., Auinger, H-J., and Schon, C-C. (2012). synbreed: a framework for the analysis of genomic predictions data using R. *Bioinformatics* 28**,** 2086-2087.

Wingfield, M. J., Brocherhoff, E.G., Wingfield, B.D., and Slippers, B. (2015). Planted forest health: The need for a global strategy. *Science* 349**,** 832-836.

Wingfield, M. J., Slippers, B., Hurley, B. P., Coutinho, T. A., Wingfield, B. D., and Roux, J. (2008). Eucalypt pest and diseases: Growing threats to plantation productivity. *Southern Forests* 70**,** 139-144.

Wolc, A., Kranis, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Avendano, A., Watson, K. A., de Los Campos, G., Fernando, R. L., Garrick, D. J., and Dekkers, J. C. M. (2016). Implementation of genomic selection in the poultry industry. *Animal Frontiers* 6**,** 23-31.

Wu, H. X., Powell, M. B., Yang, J. L., Ivkovic, M., and Mcrae, T. A. (2007). Efficiency of early selection for rotation-aged wood quality traits in radiata pine. *Annual Forestry Science* 64**,** 1-9.

Zapata-Valenzuela, J., Whetten, R. W., Naele, D., Mckeand, S., and Isik, F. (2013). Genomic estimated breeding values using genomic relationship matrices in a cloned population of Loblolly Pine. *Genes Genomics Genetics* 3**,** 909-916.

Zhong, S., Dekker, J. C. M., Fernando, R. L., and Jannink, J-L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: A Barley case study. *Genetics* 182**,** 355-364.

CHAPTER 3

GENOMIC BREEDING FOR DIAMETER GROWTH AND TOLERANCE TO *LEPTOCYBE* GALL WASP AND *BOTRYOSPHAERIA* / *TERATOSPHAERIA* FUNGAL DISEASE COMPLEX IN *EUCALYPTUS GRANDIS*

Makobatjatji M. Mphahlele[1,2], Fikret Isik[3], Gary R. Hodge[3,4], and Alexander A. Myburg[2*]

[1]Mondi Forests, Research and Development Department, Trahar Technology Centre - TTC, P.O. Box 12, Hilton, 3245, South Africa
[2]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa
[3]Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA
[4]Camcore, North Carolina State University, P.O. Box 7626, Raleigh, NC 27695, USA

Author contributions:

This data chapter has been published in the international ISI journal, Frontier Plant Science as Mphahlele MM, Isik F, Hodge GR and Myburg AA (2021) Genomic Breeding for diameter growth and tolerance to *Leptocybe* gall wasp and *Botryosphaeria/Teratosphaeria* fungal disease complex in *Eucalyptus grandis*. Frontier Plant Science. 12:638969. doi:10.3389/fpls.2021.638969. I carried out the experimental design, data collection, data analysis, and drafting of this paper as part of my PhD thesis. F.I. assisted with the modelling and scripts of the data analysis tools. F. I., G. H., and A. A. M. supervised the study and helped with data interpretation and the drafting of the manuscript. All authors read, edited and approved the final manuscript.

3.1 ABSTRACT

Eucalyptus grandis is one of the most important species for hardwood plantation forestry around the world. At present, its commercial deployment is in decline because of pests and pathogens such as *Leptocybe invasa* gall wasp (*Lepto)* and often co-occurring fungal stem diseases such as *Botryosphaeria dothidea* and *Teratosphaeria zuluensis* (*BotryoTera*). This study analysed *Lepto*, *BotryoTera*, and stem diameter growth in an *E. grandis* multi-environmental, genetic trial. The study was established in three subtropical environments. Diameter growth and *BotryoTera* incidence scores were assessed on 3334 trees, and *Lepto* incidence was assessed on 4463 trees from 95 half-sib families. Using the *Eucalyptus* EUChip60K SNP chip, a subset of 964 trees from 93 half-sib families were genotyped with 14,347 informative SNP markers. We employed single-step genomic BLUP (ssGBLUP) to estimate genetic parameters in the genetic trial. Diameter and *Lepto* tolerance showed a positive genetic correlation (0.78), while *BotryoTera* tolerance had a negative genetic correlation with diameter growth (-0.38). The expected genetic gains for diameter growth and *Lepto* and *BotryoTera* tolerance were 12.4%, 10%, and negative 3.4%, respectively. We propose a genomic selection breeding strategy for *E. grandis* that addresses some of the present population structure problems.

3.2 INTRODUCTION

Fast-growing plantation forests are essential to the pulp, paper, and timber industries and the emerging biorefinery and biomaterials industries (Perlack et al., 2005; Cetinkol et al., 2012; Devappa et al., 2015; Stafford et al., 2020). The sustainability of many of these industries is dependent on woody biomass from plantation-growing *Eucalyptus* trees. *Eucalyptus* species are adaptable, fast-growing, generally resilient to pests and pathogens, and have the desired wood qualities for diverse wood products (Malan, 1993; Stafford et al., 2020). Volume growth and wood density are essential for forest plantation productivity (Raymond, 2002). However, pest and pathogen challenges have increased in severity in the past decades, posing a significant risk to *Eucalyptus* plantation forestry productivity and sustainability in subtropical regions (Wingfield et al., 2015). How to ensure continued genetic gains for volume growth in the presence of severe pest and pathogen challenges has become an essential question for plantation species such as *Eucalyptus grandis*.

Leptocybe invasa Fisher & La Salle is one of the most damaging insect pests of *Eucalyptus* species that affects growth by forming galls on leaves and leaf petioles. The insect is native to Queensland, Australia, known as the Blue Gum Chalcid wasp (Hymenoptera: Eupholidea). It has spread across the globe, infesting a wide range of commercially grown *Eucalyptus* species and their hybrids, resulting in severe losses in young plantations and nursery seedlings (Mendel et al., 2004; Nyeko et al., 2010; Chang et al., 2012; da Silva et al., 2020). First reported in the Mediterranean Basin and the Middle East in 2000 (Viggiani et al., 2000; Mendel et al., 2004), *L. invasa* subsequently spread throughout countries in Africa, America and Asia (Nyeko, 2005; Wiley et al., 2008; Prabhu, 2010; Zhu et al., 2012). Two parasitoid species of *L. invasa* from Australia, *Quadrastichus mendeli* and *Selitrichodes kryceri*, were deployed as biological controls to manage severe infestation levels in

Eucalyptus plantations in Israel (Kim et al., 2008). Tracking the introduction of *L. invasa* in South Africa, *Q. mendeli* was recently discovered, and the biological control potential of *L. invasa* in South African *Eucalyptus* plantations was investigated (Bush et al., 2018). Another recently discovered parasitoid species of *L. invasa* from Australia, *S. neseri*, was described and investigated for its parasitism rates in South Africa, ranging from 9.7% to 71.8% (Dittrich-Schroder et al., 2014).

Resistance-linked DNA markers for molecular breeding are an alternative strategy to manage pest challenges. Towards this, simple sequence repeat (SSR) markers have been identified that jointly explained 3% to 37% of the variation of resistance in *E. grandis* and were validated in *E. tereticornis,* explaining 24% to 48% of the variation of resistance (Zhang et al., 2018). Due to the significant variation that exists within and between *Eucalyptus* species, there is an opportunity to breed for *L. invasa* tolerance (Mendel et al., 2004; Thu et al., 2009; Durand et al., 2011; Sangtongpraow et al., 2011; Dittrich-Schroder et al., 2012; Nugnes et al., 2015; Zheng et al., 2016). A recent genome-wide association study in an *E. grandis* breeding population identified candidate genomic regions on chromosomes 3, 7 and 8 that contained putative candidate genes for tolerance. These candidate genomic regions explained ~17.6% of the total phenotypic variation of *L. invasa* tolerance (Mhoswa et al., 2020).

Teratosphaeria zuluensis, a fungal pathogen that causes stem canker, previously known as *Coniothyrium* canker, is a devastating stem disease of *Eucalyptus* species and is one of the most severe pathogens of plantation-grown *Eucalyptus* spp. (Wingfield et al., 1996; Crous et al., 2009; Aylward et al., 2019). It was first recognised in South Africa in 1989 and described in 1996 (Wingfield et al., 1996). *T. zuluensis* has been reported on *Eucalyptus* spp.

in Malawi, Mozambique and Zambia (Jimu et al., 2015), Hawaii (Cortinas et al., 2004), Ethiopia (Gezahgne and Wingfield, 2003), Argentina and Vietnam (Gezahgne et al., 2004b). Infections from *T. zuluensis* results in necrotic spots on green branches and the main stem, giving a "cat-eye" appearance that develops into large cankers on susceptible trees. *T. zuluensis* infection reduces wood quality by penetrating the cambium to form black kino-filled pockets and may lead to tree death (Wingfield et al., 1996; Gezahgne and Wingfield, 2003).

Botryosphaeria dothidea is also a devastating fungal pathogen of eucalypt species affecting the stem. *B. dothidea* is known to have endophytic characteristics with instances of opportunistic latent infections (Smith et al., 1996; Slippers et al., 2009). Species of the *Botryosphaeriaceae* family infect plants via natural apertures (Bihon et al., 2011) and wounding (Epstein et al., 2008). *B. dothidea* infection results in longitudinal cracks that penetrate the bark into the xylem, forming kino-filled pockets in the wood, and stem cankers and tip die-back (Smith et al., 1994). It infects eucalypts in many countries, including the Congo (Roux et al., 2000), Australia (Burgess et al., 2019), South Africa (Smith et al., 1994), Ethiopia (Gezahgne et al., 2004a), Venezuela (Mohali et al., 2007), Colombia (Rodas et al., 2009), Uruguay (Perez et al., 2008), and China (Chen et al., 2011). Field assessment of the two fungal stem pathogens has revealed that the symptoms of *B. dothidea* and *T. zuluensis* can be present separately or concurrently at varying levels on trees in the population in the form of a fungal stem disease complex.

In general, tree breeding strategies use pedigree information to estimate genetic merit, often in trials with large numbers of individuals in open-pollinated families. The availability of a reference genome sequence of *E. grandis* (Myburg et al., 2014) and the development of a

robust single nucleotide polymorphism (SNP, EUChip60K) chip for high-throughput genotyping in multiple eucalypt species (Silva-Junior et al., 2015) have created opportunities for implementing new breeding strategies based on the genomic prediction of breeding values. While conventional pedigree relationships represent the average proportion of shared alleles, SNP markers can track Mendelian segregation patterns enabling the detection of unknown (cryptic) relationships and more precise estimation of known relationships (Habier et al., 2007; Hayes et al., 2009; Hill and Weir., 2010). However, the genotyping of all individuals in large open-pollinated tree breeding populations would be prohibitively expensive. Single-step genomic (ssG)BLUP analysis is an attractive alternative that blends the known pedigree of the entire population with the genomic relationship matrix of a subset of genotyped individuals (Legarra et al., 2009; Misztal et al., 2009; Christensen and Lund, 2010). Thereby, ssGBLUP analysis extends the benefits of applying genomic BLUP to non-genotyped individuals (Legarra et al., 2014), therefore allowing for multivariate and univariate analysis (Guo et al., 2014) in livestock (Lourenco et al., 2015; Ma et al., 2015) and forest trees (Ratcliffe et al., 2017; Klapste et al., 2018; Cappa et al., 2019; Klapste et al., 2020).

Improving forest plantation productivity requires recurrent selection of multiple traits, such as growth, wood quality and tolerance to pests and pathogens. A multivariate analysis involves estimating genetic correlations between traits to understand their correlated responses (Burdon, 1977). The correlated phenotypes of growth and pest and disease traits are attributable to shared genetic factors (pleiotropy) and/or linked genetic factors (linkage disequilibrium) and their interrelationships with environment factors (Falconer and Mackay, 1996). Being able to partition these components will help improve breeding strategies for correlated traits (Chen and Lubberstedt, 2010).

In this study, we measured breeding trials of *E. grandis* comprised of trees from three half-sib pedigree linked generations and some unrelated families for diameter growth at breast height, tolerance to stem disease caused by the co-occurrence of *B. dothidea* and *T. zuluensis* (*BotryoTera*) and tolerance to leaf gall caused by *L. invasa* (*Lepto*). The study aimed to obtain genetic parameters and genetic gains for growth, pest and pathogen tolerance in this multi-generation breeding trial comparing ABLUP (pedigree-based BLUP analysis) and ssGBLUP models. We further investigated the additive genetic correlations and genotype-by-environmental (G×E) interactions of diameter growth and tolerance to *Lepto* and *BotryoTera*. Based on the results, we discuss the utility of genomic selection in *E. grandis* for simultaneous improvement of growth and tolerance to the gall wasp and fungal stem disease.

3.3 MATERIALS AND METHODS

3.3.1 Breeding history and phenotyping of the study population

Eucalyptus grandis W. Hill ex Maiden was introduced to South Africa in the early 1900s and included various government breeding populations as a timber resource for the mining industry. Private breeding programs only started in the early 1970s, initiated from government landrace breeding populations. Breeding objectives for these landrace breeding populations gradually shifted to target traits for pulp and paper products rather than timber production in successive generations and trial series **(Figure 3.1)**. We had access to seed from two first-generation selections from the 2nd trial series in this study population, 32 third-generation selections from the 3rd trial series, and 28 fourth-generation selections from the 4th trial series **(Table 3.1)**. Also included in the study were 33 unrelated (no pedigree link) families as controls, from seed sourced in the early 1990s from selections in Swaziland. The

62 half-sib pedigree linked families and the 33 unrelated control families were planted across three sites Mtunzini, Kwambonambi and Nyalazi, in KwaZulu Natal, a sub-tropical region in South Africa (**Figure 3.2 and Table 3.1)**. Families from the different generations were planted together in the three trial sites. The experimental design was a randomised complete block planted at single tree plots at 15 replicates per family.

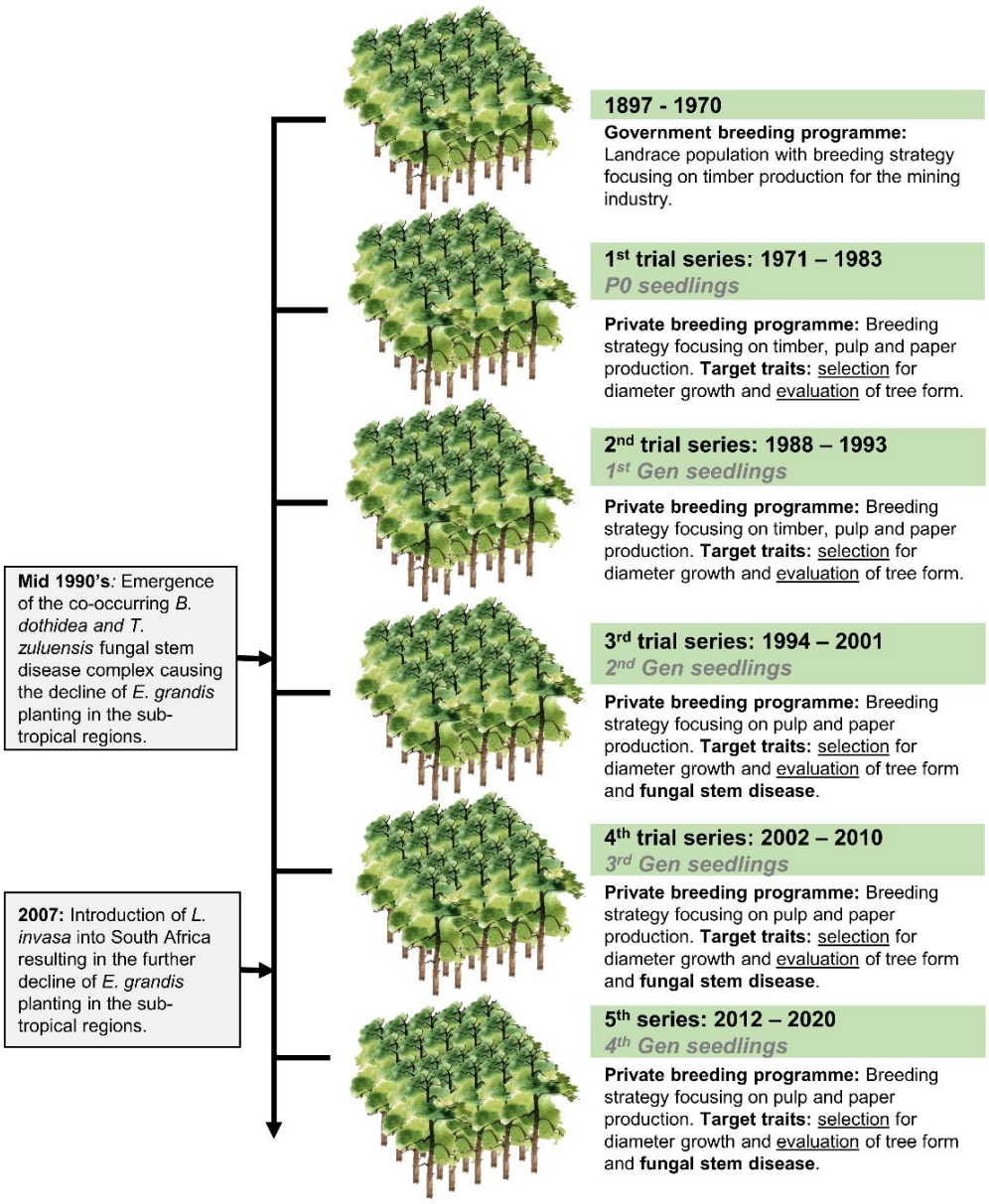


FIGURE 3.1 | Historical overview of E. grandis breeding in South Africa, from the government to private breeding and the introduction of major pests and pathogens. The trial series timeline, as well as the generational timeline, are shown. Selection strategies are noted for each trial series, shifting from timber to pulp and paper related traits and pest and disease tolerance. Selection refers to the selection of phenotyped individuals based on their breeding values, whereas evaluation refers to the selection of individuals based on

visual screening without phenotypes and breeding values.

TABLE 3.1 | Environmental and trial design information of the study population. The number of families and their pedigree generation are indicated with the phenotyped and genotyped individuals.

		Nyalazi	Kwambonambi	Mtunzini
Site environment	Latitude (South)	28° 12'32.01" S	28° 38' 56.43" S	29° 1' 52.11" S
	Longitude (East)	32° 20' 42.79" E	32° 9' 13.81" E	31° 39' 23.73" E
	Altitude (m)	47	63	69
	[a]MAP (mm)	999	1196	1220
	[b]MAT (C)	21	21	21
	MAT min. (C)	12	11	11
	MAT max. (C)	30	29	28
Distance (km)	Nyalazi		50	112
	Kwambonambi			66
Trial design	Progeny type	Half-sib	Half-sib	Half-sib
	Trial design	RCB	RCB	RCB
	Replications	15	15	15
	Plot design	Single-tree plot	Single-tree plot	Single-tree plot
Trial pedigree	*Unrelated families*	33	33	31
	2nd Gen families	2	2	
	3rd Gen families	36	38	33
	4th Gen families	51	53	48
	Total families	122	126	112
	Number of trees	1830	1890	1680
Survival (%)	Diameter (4yrs)	68	59	58
	BotryoTera (3yrs)	83	66	68
	Lepto(1.5yrs)	89	80	78
Phenotyped (*n*)	Diameter	1290	1074	970
	BotryoTera	1573	1216	1144
	Lepto	1687	1465	1311
Genotyped pedigree	*Unrelated families*	33	33	31
	2nd Gen families	2	2	
	3rd Gen families	32	32	30
	4th Gen families	28	28	28
	Total families	95	95	89
	Number of trees	321	325	318
Genotyped (*n*)	Diameter	246	291	273
	BotryoTera	321	325	318
	Lepto	321	325	318

[a]MAP – mean annual precipitation
[b]MAT – mean annual temperature

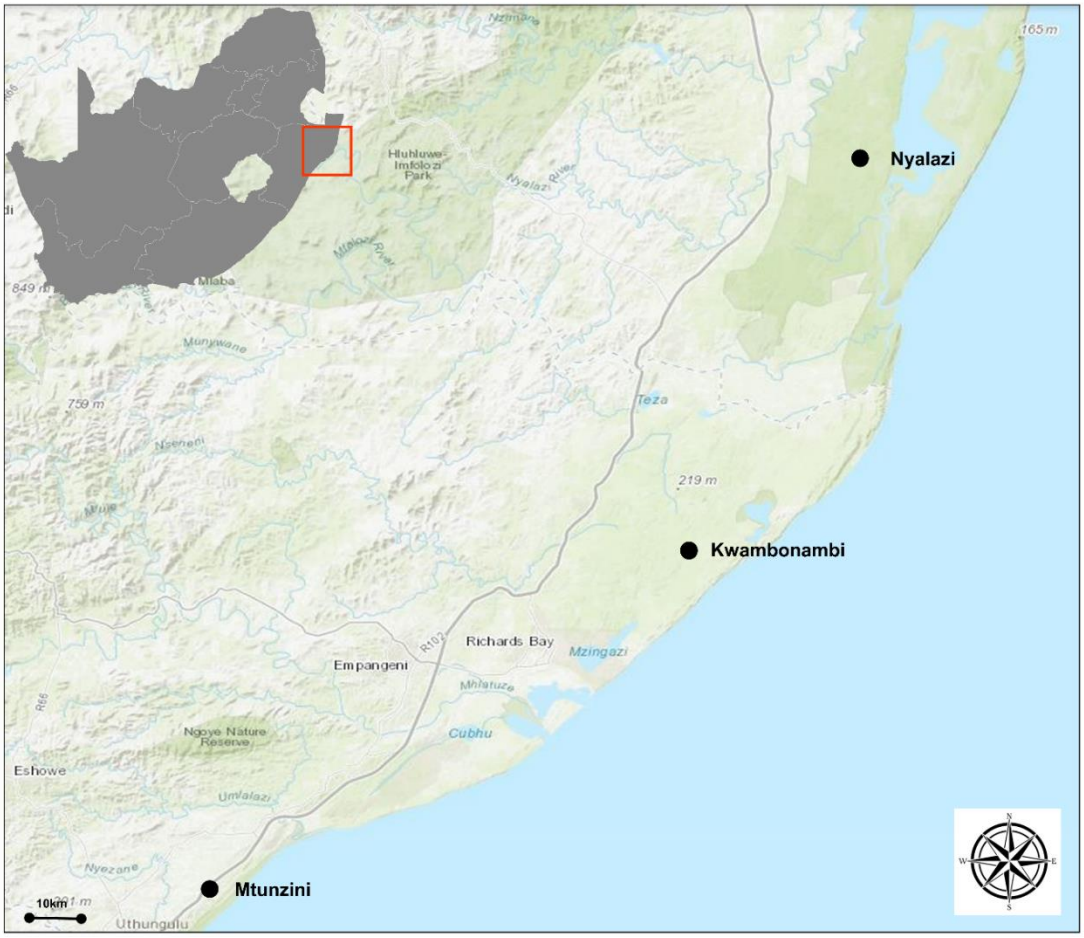


FIGURE 3.2 | Geographical representation of the trial sites in the KwaZulu Natal province, South Africa. The region has a sub-tropical climate. The distance (straight line) between Mtunzini and Nyalazi is 112 km. The details of the environmental conditions are in **Table 3.1**. Darker shades of green indicate nature reserves.

Field tolerance to *Lepto* was assessed at age one using a four-scale incidence score in which trees with Score 4 shows no evidence of an attack on the leaf midrib or petiole, Score 3 shows evidence of an attack on the leaf midrib or petiole without galls, and Score 2 indicates trees with an attack on the leaf midrib or petiole with galls. Trees with Score 1 present a lethal outcome from an attack on the leaf midrib or petiole with galls (**Figure 3.3**). Field tolerance to *BotryoTera* was assessed at age three using an incidence score in which Score 6 represents trees with no spots/cracks or redness and trees with Score 5 show symptoms of *T. zuluense* spots with redness, whereas trees with Score 4 have with *B. dothidea* cracks with

redness. Trees with Score 3 shows symptoms with *T. zuluense* spots and *B. dothidea* cracks with redness, and Score 2 represents trees with heavy *T. zuluense* spots, and *B. dothidea* cracks with redness, and Score 1 represents trees with heavy *T. zuluense* spots and *B. dothidea* cracks with redness and cankers **(Figure 3.4)**. Diameter growth at breast height (1.3 m over-bark) was measured at age four.

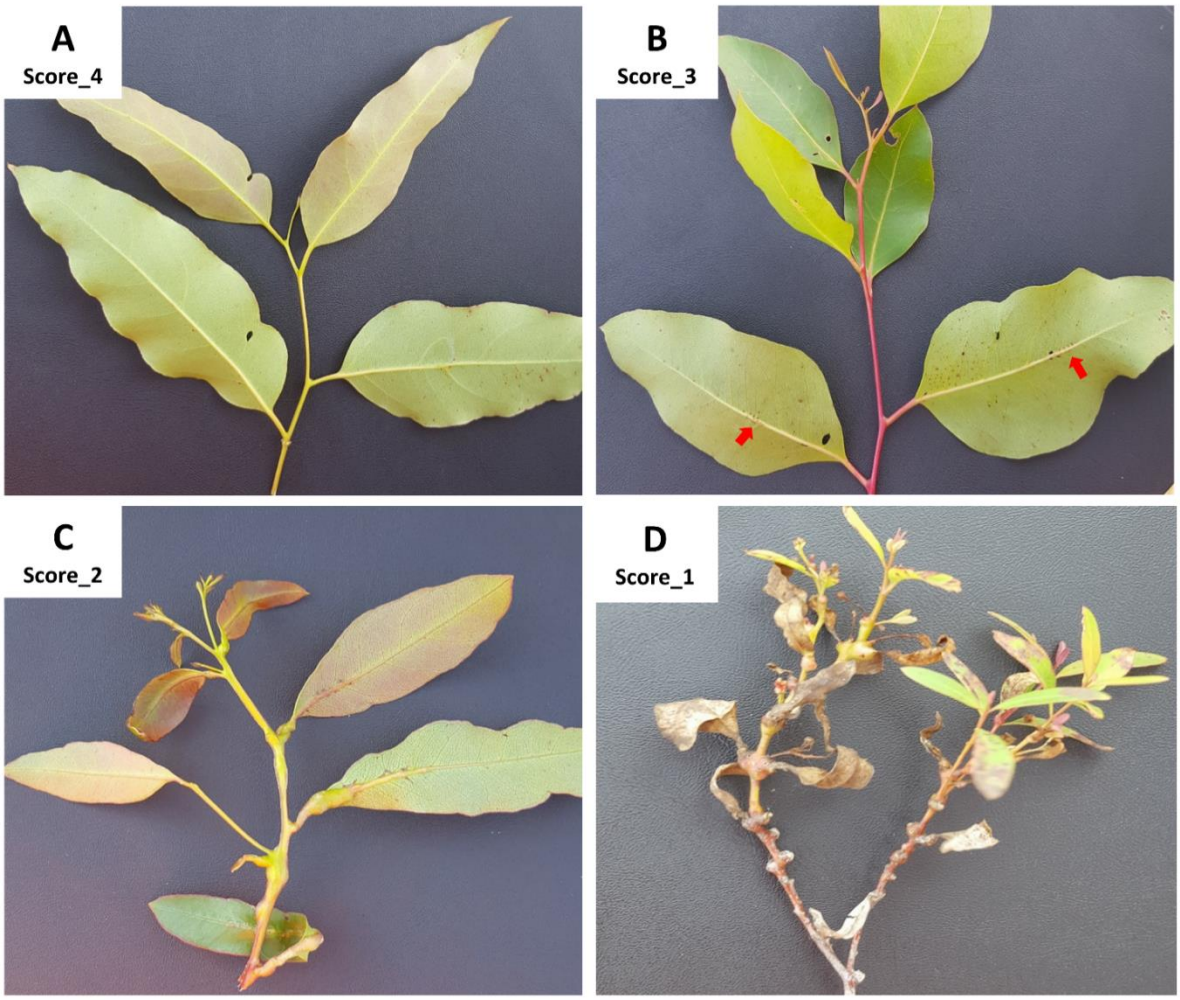


FIGURE 3.3 | Symptoms and incidence scores of *Leptocybe invasa* (*Lepto*). (A) Score 4 – No evidence of an attack on the leaf midrib or petiole, **(B)** Score 3 – Evidence of attack on the leaf midrib or petiole without galls (indicated by red arrows), **(C)** Score 2 – Evidence of attack on the leaf midrib or petiole with galls and **(D)** Score 1 – Evidence of a lethal outcome of an attack on the leaf midrib or petiole with galls.

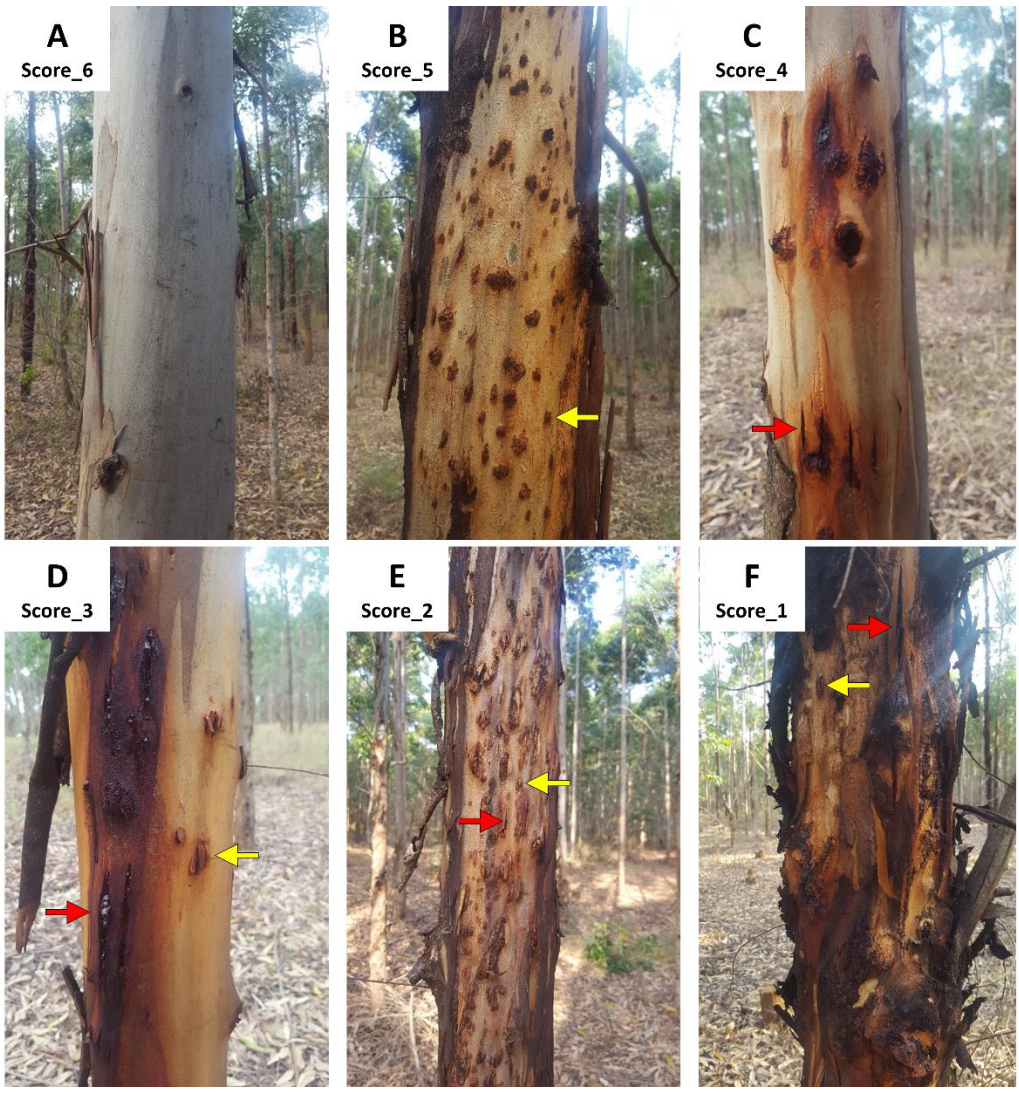


FIGURE 3.4 | Symptoms and incidence scores for *Botryosphaeria / Teratosphaeria* stem disease complex (*BotryoTera*). (A) Score 6 represents trees with no spots/cracks or redness. (B) Score 5 represents trees with *T. zuluense* spots with redness. (C) Score 4 is given for trees with *B. dothidea* cracks with redness. (D) Score 3 shows a tree with *T. zuluense* spots and *B. dothidea* cracks with redness. (E) Score 2 represents trees with heavy *T. zuluense* spots and *B. dothidea* cracks with redness. (f) Score 1 represents trees with heavy *T. zuluense* spots and *B. dothidea* cracks with redness and cankers.

3.3.2 Genotyping of the study population

DNA was extracted from leaves using the NucleoSpin DNA extraction kit (Machery-Nagel, Germany). The *Eucalyptus* (EUChip60K) SNP chip, as described by Silva-Junior (Silva-Junior et al., 2015) available from GeneSeek (Neogen, Lansing, MI, USA), was used to genotype 964 trees across the families and trials **(Table 3.1)**. Of the 95 families in the trials,

93 contained a subset of 964 genotyped trees ranging from 2 to 24 trees per family. The two second-generation families were not genotyped. An average of four trees per family were genotyped of the unrelated families. For the third generation, fifteen trees per family were genotyped, while in the fourth generation, fourteen per family were genotyped. Of the 64,639 markers on the SNP chip (Silva-Junior et al., 2015), a total of 14,347 informative SNP markers with GenTrain scores ranging from 0.37 to 0.93 were retained markers with call rates of above 90% and a minor allele frequency above 0.05. The SNP genotype frequencies of the 14,347 markers were AA (0.307), GG (0.283), AG (0.270), CC (0.068), AC (0.065) and 0.007 missing. The number of SNP markers distributed on linkage groups ranged from 1,018 (Chromosome 1) to 1,877 (Chromosome 10). The SNP marker frequencies and distribution analysis were performed with the *synbreed* 0.10-2 R package (Wimmer et al., 2012) and the imputing of the missing SNP data was based on the allelic distribution, assuming Hardy–Weinberg equilibrium.

3.3.3 Mixed model analysis

Linear mixed models were fit to estimate variance components and solve mixed model equations to obtain solutions for fixed and random effects. The matrix notation for the linear mixed models used is as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{\varepsilon} \qquad (1)$$

where $\boldsymbol{y}$ is a vector of phenotypes, $\mathbf{X}$ is the design matrix for the fixed effects (site), $\boldsymbol{\beta}$ is the vector of the fixed effect coefficients (intercept site), $\mathbf{Z}$ is an incidence matrix for the random effects of individual trees and $\boldsymbol{u}$ is the vector of random effect coefficients (genotype, genotype by site interaction, replication effect nested in site effect), and $\boldsymbol{\varepsilon}$ is the vector of residual effect coefficients. The expectations of $\boldsymbol{y}$, $\mathbf{u}$ and $\boldsymbol{e}$ are $\boldsymbol{E}(\boldsymbol{y}) = \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{E}(\boldsymbol{\mu}) = \sim\boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}})$ and $\boldsymbol{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and the variances are $\boldsymbol{Var}(\boldsymbol{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$, $\mathbf{Var}(\boldsymbol{\varepsilon}) = \mathbf{R} =$

$\sim \boldsymbol{N}(\mathbf{0}, \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}})$, and $\mathbf{Var}(\boldsymbol{u}) = \mathbf{G} = \mathbf{A}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}$, respectively, where $\mathbf{A}$ is the relationship matrix of the random effects, $\boldsymbol{\sigma}_{\boldsymbol{\epsilon}}^{\mathbf{2}}$ is the variance associated with the residuals and $\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}}$ is the variance associated with the random effects. The assumptions of residual matrix $\mathbf{R}$ were relaxed to have a heterogeneous error variance across the environments. Similarly, the assumptions of the $\mathbf{G}$ matrix were relaxed to model full G×E and heterogeneous genetic variances at each site ($\boldsymbol{s}$ + 1 variance parameters), where $\boldsymbol{s}$ is the number of environments (Isik et al., 2017). Empirical breeding value prediction for the half-sibs was performed by solving the mixed model equations.

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \boldsymbol{A}^{-1}\boldsymbol{\lambda} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\boldsymbol{y} \\ \mathbf{Z}'\boldsymbol{y} \end{bmatrix} \qquad (2)$$

where $\boldsymbol{A}^{-1}$ is the inverted additive genetic relationship matrix derived from the pedigree and $\boldsymbol{\lambda} = \frac{\sigma_e^2}{\sigma_u^2}$ is the shrinkage factor. The genomic relationship matrix $\mathbf{G}$ from the genotyped trees was computed as described in (VanRaden, 2008):

$$\boldsymbol{G} = \frac{(\boldsymbol{Z}-\boldsymbol{P})(\boldsymbol{Z}-\boldsymbol{P})'}{\mathbf{2}\sum \boldsymbol{pi}(\mathbf{1}-\boldsymbol{p_i})} \qquad (3)$$

where $\boldsymbol{Z}$ and $\boldsymbol{P}$ are two matrices of dimension $\boldsymbol{n}$ (individuals) × $\boldsymbol{p}$ (markers). The base pair calls were transformed into gene content values of the minor alleles at each SNP loci in each individual in matrix $\boldsymbol{Z}$, with elements -1 (homozygote major allele), 0 (heterozygote), and 1 (homozygote minor allele). The frequencies of the genotypes were 0.584, 0.338 and 0.078, respectively. The allele frequencies in matrix $\boldsymbol{P}$ are presented as $\mathbf{2}(\boldsymbol{p_i} - \mathbf{0.5})$, where $\boldsymbol{p_i}$ is the observed allele frequency at the marker $\boldsymbol{i}$ for all individuals. The $\mathbf{2}\sum \boldsymbol{pi}\,(\mathbf{1} - \boldsymbol{p_i})$ is the variance of alleles summed across all the loci. A ssGBLUP model was fitted using a blended relationship ($\boldsymbol{H}$) matrix, incorporating the ($\boldsymbol{G}$) matrix of genotyped trees that are linked to the non-genotyped trees by the half-sib pedigree ($\boldsymbol{A}$) matrix (Legarra et al., 2009; Aguilar et al., 2010; Christensen and Lund, 2010).

The $\boldsymbol{H}$ matrix used in the ssGBLUP was formulated as follows: where $\boldsymbol{u}$ as a vector of genetic effects with variances $\boldsymbol{Var(u) = A\sigma_u^2}$. Within the genetic effects ($\boldsymbol{u}$) there are non-genotyped and ($\boldsymbol{u_1}$) and genotyped ($\boldsymbol{u_2}$), individuals partitioned in the $\boldsymbol{A}$ matrix as:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A_{11}} & \boldsymbol{A_{12}} \\ \boldsymbol{A_{21}} & \boldsymbol{A_{22}} \end{bmatrix} \qquad \textbf{(4)}$$

where $\boldsymbol{A_{11}}$ is the relationship matrix of non-genotyped individuals, $\boldsymbol{A_{22}}$ is the relationship matrix for the genotyped individuals, and $\boldsymbol{A_{12}}$ and its transpose $\boldsymbol{A_{21}}$ are the covariances between the genotyped non-genotyped individuals. We then replaced the $\boldsymbol{u_2}$ genetic effects with the pedigree relationship of $\boldsymbol{A_{22}}$ with their $\boldsymbol{G}$ matrix as constructed in equation 3. The relationship between the non-genotyped and ($\boldsymbol{u_1}$) and genotyped ($\boldsymbol{u_2}$) individuals in $\boldsymbol{A_{12}}$ and $\boldsymbol{A_{21}}$ is then adjusted by the $\boldsymbol{G}$ matrix via the pedigree relationship of all other individuals in the $\boldsymbol{H}$ matrix (Legarra et al., 2009):

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{A_{11} + A_{12}A_{22}^{-1}(G - A_{22})\, A_{22}^{-1}A_{21}} & \boldsymbol{A_{12}A_{22}^{-1}G} \\ \boldsymbol{GA_{22}^{-1}A_{21}} & \boldsymbol{G} \end{bmatrix} \qquad \textbf{(5)}$$

The upper left corner of the $\boldsymbol{H}$ matrix is the variance of the $\boldsymbol{u_1}$ individuals. with $\boldsymbol{Var(u_1) = [A_{11} + A_{12}A_{22}^{-1}(G - A_{22})\, A_{22}^{-1}A_{21}]\sigma_A^2}$, and $\boldsymbol{Var(u_2) = G\sigma_A^2}$ and $\boldsymbol{Cov(u_1, u_2) = A_{12}A_{22}^{-1}G\sigma_A^2}$. The inverse of the $\boldsymbol{H}$ matrix is:

$$\boldsymbol{H^{-1}} = \boldsymbol{A^{-1}} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G^{-1} - A_{22}^{-1}} \end{bmatrix} \qquad \textbf{(6)}$$

Variance components from the ABLUP and ssGBLUP were estimated along with the heritability for diameter growth, *Lepto* and *BotryoTera* tolerance across and within the three sites.

3.3.4 Multivariate analysis

A multivariate linear mixed model was fitted to estimate additive genetic correlations between three pairs of traits as described in Isik et al. (2017), following the multivariate model general design:

$$\boldsymbol{y}_{n\times d} = \mathbf{X}_{n\times(p+1)}\boldsymbol{\beta}_{(p+1)\times d} + \mathbf{Z}_{n\times r}\boldsymbol{u}_{r\times d} + \boldsymbol{\varepsilon}_{n\times d} \qquad (7)$$

where $\boldsymbol{n}$ is the number of rows of individuals, and $\boldsymbol{d}$ is the number of dependent variables (traits). The design matrix $\mathbf{X}$ has the dimensions $\boldsymbol{n} \times (\boldsymbol{p} + \mathbf{1})$, where $\boldsymbol{p}$ is the number of fixed estimators which are replication nested in location for the traits, and the additional column is added for the intercept. $\boldsymbol{\beta}$ is the matrix of coefficients of fixed predictor effects to be estimated with dimensions $(\boldsymbol{p} + \mathbf{1}) \times \boldsymbol{d}$. The rows of $\boldsymbol{\beta}$ correspond to predictor variables, and the columns are response variables. The design matrix of $\mathbf{Z}$ has dimensions $\boldsymbol{n} \times \boldsymbol{r}$, where $\boldsymbol{r}$ is the number of random effects (individual trees) per trait, and $\boldsymbol{u}$ is a $\boldsymbol{r} \times \boldsymbol{d}$ matrix of the random effects.

The $\boldsymbol{G}$ and $\boldsymbol{R}$ variance-covariance matrices of the multivariate model were designed with the variances for the three traits on the diagonal and the covariances between the traits on the off-diagonals:

$$\boldsymbol{G} = \boldsymbol{A} \otimes \begin{bmatrix} \boldsymbol{\sigma}^2_{\boldsymbol{A11}} & \boldsymbol{\sigma}_{\boldsymbol{A12}} & \boldsymbol{\sigma}_{\boldsymbol{A13}} \\ \boldsymbol{\sigma}_{\boldsymbol{A21}} & \boldsymbol{\sigma}^2_{\boldsymbol{A22}} & \boldsymbol{\sigma}_{\boldsymbol{A23}} \\ \boldsymbol{\sigma}_{\boldsymbol{A31}} & \boldsymbol{\sigma}_{\boldsymbol{A32}} & \boldsymbol{\sigma}^2_{\boldsymbol{A33}} \end{bmatrix} \qquad (8)$$

$$\boldsymbol{R} = \boldsymbol{I}_{\boldsymbol{m}} \otimes \begin{bmatrix} \boldsymbol{\sigma}^2_{\boldsymbol{\varepsilon 11}} & \boldsymbol{\sigma}_{\boldsymbol{\varepsilon 12}} & \boldsymbol{\sigma}_{\boldsymbol{\varepsilon 13}} \\ \boldsymbol{\sigma}_{\boldsymbol{\varepsilon 21}} & \boldsymbol{\sigma}^2_{\boldsymbol{\varepsilon 22}} & \boldsymbol{\sigma}_{\boldsymbol{\varepsilon 23}} \\ \boldsymbol{\sigma}_{\boldsymbol{\varepsilon 31}} & \boldsymbol{\sigma}_{\boldsymbol{\varepsilon 32}} & \boldsymbol{\sigma}^2_{\boldsymbol{\varepsilon 33}} \end{bmatrix} \qquad (9)$$

Where the $\boldsymbol{G}$ matrix, is the direct product of the $\boldsymbol{A}$ matrix (pedigree relationship) for the ABLUP model and substituted with the $\boldsymbol{H}$ matrix for the ssGBLUP model with an

unstructured, heterogeneous variance and covariance structure, where each environment has a unique genetic variance, and each pair of the environments has a unique covariance, with a $\boldsymbol{s(s+1)/2}$ variance parameter (Isik et al., 2017). The $\boldsymbol{R}$ matrix is the direct product of the identity matrix ($\boldsymbol{I_m}$) with $\boldsymbol{m}$ dimensions, $\boldsymbol{m}$ is the number of genotypes with variance $\boldsymbol{\sigma^2_{\varepsilon 1}}$ for diameter growth, $\boldsymbol{\sigma^2_{\varepsilon 2}}$ for *BotryoTera* and $\boldsymbol{\sigma^2_{\varepsilon 3}}$ for *Lepto* and their covariances nested within.

The construction of the expected additive ($\boldsymbol{A}$ matrix) and the realised genomic ($\boldsymbol{G}$) were calculated using the package *synbreed* 0.10-2 (Wimmer et al., 2012) in the R environment v3.5.3. The blended genetic relationships and its inverse were obtained using scripts according to Isik et al. (2017). All the statistical models were performed using ASReml software v4.1 (Gilmour et al., 2015).

3.3.5 Expected direct and indirect genetic gains

The direct genetic gains for *diameter* growth, *Lepto*, and *BotryoTera* tolerance were calculated from the ABLUP and ssGBLUP models breeding value predictions. The selection differential was based on the top 10% of individuals for direct selection. The indirect responses of the remaining traits were calculated based on the ranking of the direct selections. The percentage expected genetic gains were calculated at the mean of the breeding value of the selected individuals over the population mean.

3.4 RESULTS

3.4.1 Genetic parameters

To assess the increased accuracy of the ssGBLUP model, we compared the heritability estimates from ssGBLUP with those from ABLUP analysis. The ssGBLUP model generally

produced lower heritability estimates compared to the ABLUP model for the three sites **(Table 3.2)**. The exception was the heritability estimates for *BotryoTera* tolerance in Kwambonambi and Nyalazi, which were higher for ssGBLUP (0.45 vs 0.29 and 0.11 vs 0.08, respectively). Overall, the Kwambonambi site produced the highest heritability values ranging from 0.29 to 0.63 (ABLUP) and from 0.45 to 0.70 (ssGBLUP) across the traits **(Table 3.2)**. In contrast, the heritability estimates for *Lepto* tolerance from the ABLUP and ssGBLUP models were the highest at 0.71 and second-highest at 0.38, respectively, in Nyalazi, while the estimates for diameter growth and *BotryoTera* tolerance at the Nyalazi site were reasonably low, ranging from 0.07 to 0.11 for the ABLUP and ssGBLUP models, respectively **(Table 3.2)**. The overall heritability estimates across sites were higher for the ABLUP model with *Lepto* tolerance moderately high at 0.54, diameter growth at 0.33 and *BotryoTera* tolerance at 0.23 **(Table 3.3)**. The heritability estimates with the ssGBLUP across sites were lower with *Lepto* tolerance at 0.36, diameter growth at 0.25 and *BotryoTera* tolerance at 0.23 **(Table 3.3)**. The heritability estimates for ssGBLUP may be more accurate due to the blended pedigree relationship matrix increased precision.

TABLE 3.2 | Site-specific variance components and genetic parameters estimated using the ABLUP and ssGBLUP mixed models for diameter growth, BotryoTera and Lepto tolerance. The residual variance ($\boldsymbol{\sigma_e^2}$), additive genetic variance ($\boldsymbol{\sigma_u^2}$), narrow-sense heritability ($\boldsymbol{h^2}$) and their standard errors ($\boldsymbol{se}$) are shown.

	$\boldsymbol{\sigma_u^2}$ $(\boldsymbol{se})$	$\boldsymbol{\sigma_e^2}$ $(\boldsymbol{se})$	$\boldsymbol{h^2}$ $(\boldsymbol{se})$
Diameter			
ABLUP			
Mtunzini	6.655 (0.281)	2.360 (0.655)	0.35 (0.092)
Kwambonambi	13.193 (0.662)	8.250 (1.945)	0.63 (0.129)
Nyalazi	11.928 (0.547)	0.884 (0.579)	0.07 (0.048)
ssGBLUP			
Mtunzini	6.670 (0.277)	1.620 (0.504)	0.24 (0.072)
Kwambonambi	13.592 (0.682)	7.852 (1.487)	0.58 (0.092)
Nyalazi	11.958 (0.552)	0.779 (0.582)	0.07 (0.048)
BotryoTera			
ABLUP			
Mtunzini	1.450 (0.055)	0.424 (0.115)	0.29 (0.0752)
Kwambonambi	2.334 (0.099)	0.115 (0.203)	0.29 (0.0823)
Nyalazi	1.411 (0.059)	0.109 (0.056)	0.08 (0.0393)
ssGBLUP			
Mtunzini	1.447 (0.053)	0.222 (0.077)	0.15 (0.052)
Kwambonambi	2.404 (0.110)	1.088 (0.227)	0.45 (0.083)
Nyalazi	1.418 (0.060)	0.154 (0.073)	0.11 (0.051)
Lepto			
ABLUP			
Mtunzini	0.454 (0.017)	0.161 (0.039)	0.35 (0.080)
Kwambonambi	0.762 (0.035)	0.524 (0.105)	0.69 (0.118)
Nyalazi	0.764 (0.037)	0.542 (0.112)	0.71 (0.125)
ssGBLUP			
Mtunzini	0.452 (0.016)	0.110 (0.026)	0.24 (0.055)
Kwambonambi	0.770 (0.033)	0.538 (0.070)	0.70 (0.072)
Nyalazi	0.744 (0.031)	0.281 (0.049)	0.38 (0.059)

TABLE 3.3 | Overall variance components and genetic parameters across the three sites for solving ABLUP and ssGBLUP mixed models for diameter growth, *BotryoTera* and *Lepto* tolerance. The residual variance ($\boldsymbol{\sigma_e^2}$), additive genetic variance($\boldsymbol{\sigma_u^2}$), narrow-sense heritability ($\boldsymbol{h^2}$) and its standard error ($\boldsymbol{se}$) are presented.

	$\boldsymbol{\sigma_u^2}$ $(\boldsymbol{se})$	$\boldsymbol{\sigma_e^2}$ $(\boldsymbol{se})$	$\boldsymbol{h^2}$ $(\boldsymbol{se})$
ABLUP			
Diameter	10.581 (0.314)	3.450 (0.720)	0.33 (0.063)
BotryoTera	1.732 (0.044)	0.407 (0.092)	0.23 (0.051)
Lepto	0.659 (0.021)	0.357 (0.059)	0.54 (0.077)
ssGBLUP			
Diameter	10.729 (0.313)	2.733 (0.469)	0.25 (0.040)
BotryoTera	1.755 (0.046)	0.396 (0.071)	0.23 (0.038)
Lepto	0.655 (0.017)	0.238 (0.024)	0.36 (0.032)

3.4.2 ssGBLUP additive and Type-B genetic correlations

The additive genetic correlations of diameter growth and *Lepto* tolerance estimated with the ssGBLUP model was high at 0.78 (**Table 3.4, Eq. 7**). In contrast, the additive genetic correlation of diameter growth and *BotryoTera* tolerance was moderate at -0.38. The additive genetic correlation for *BotryoTera* and *Lepto* tolerance was also moderate at -0.47 **(Table 3.4)**. These results suggest that tandem improvement of diameter growth and *Lepto* tolerance is possible, but they predict a negative response in *BotryoTera* tolerance, which presents a challenge to breeders. The overall Type-B genetic correlation **(Eq. 7)** was high, ranging from 0.77 to 0.81 for the three traits associated with small standard errors **(Table 3.5)**, suggesting low G×E interactions across the sites.

TABLE 3.4 | Additive genetic correlations (r_g) of diameter growth, *BotryoTera*, and *Lepto* tolerance based on ABLUP and ssGBLUP models. tandard errors are in the parenthesis.

	BotryoTera	*Lepto*
ABLUP		
Diameter	-0.46 (0.116)	0.81 (0.054)
BotryoTera		-0.47 (0.111)
ssGBLUP		
Diameter	-0.38 (0.106)	0.78 (0.055)
BotryoTera		-0.47 (0.089)

TABLE 3.5 | Overall Type-B genetic correlation (r_B) across sites for diameter growth, *BotryoTera* and *Lepto* tolerance based on ABLUP and ssGBLUP models. Standard errors are in the parenthesis.

	r_B (se)
ABLUP	
Diameter	0.90 (0.096)
BotryoTera	0.99 (0.000)
Lepto	0.87 (0.055)
ssGBLUP	
Diameter	0.80 (0.140)
BotryoTera	0.81 (0.147)
Lepto	0.77 (0.072)

3.4.3 Trait performance across site and generations

Diameter growth and the *Lepto* incidence scores resembled a normal distribution **(Figure 3.5).** *BotryoTera* incidence scores had a high frequency of score 6, representing uninfected stems, and Kwambonambi has a high frequency of score 3 **(Figure 3.5)**. The latter may be ascribed to the high susceptibility of the second generation families **(Figure 3.6 and 3.7B)**. The Kwambonambi site had the lowest mean *BotryoTera* tolerance compared to the Nyalazi and Mtunzini **(Figure 3.7E).** The average diameter growth improved by 3.2% from the third to the fourth-generation **(Figure 3.7A)**, whereas *Lepto* tolerance improved by 3.6% **(Figure 3.7C).** The improvement in diameter growth is driven by recurrent selection over the

generations, with *Lepto* tolerance benefiting from its strong additive genetic correlation with diameter growth **(Table 3.4)**. There was a 13.3% improvement of *BotryoTera* tolerance from the second to the third generation; however, it was unchanged from the third to the fourth-generation **(Figure 3.7B).** The apparent absence of genetic gain for *BotryoTera* tolerance from the third to the fourth generation is in part due to the moderately negative genetic correlation with diameter growth **(Table 3.4)**. The above results suggest that a revised breeding strategy is needed to improve the three traits simultaneously.

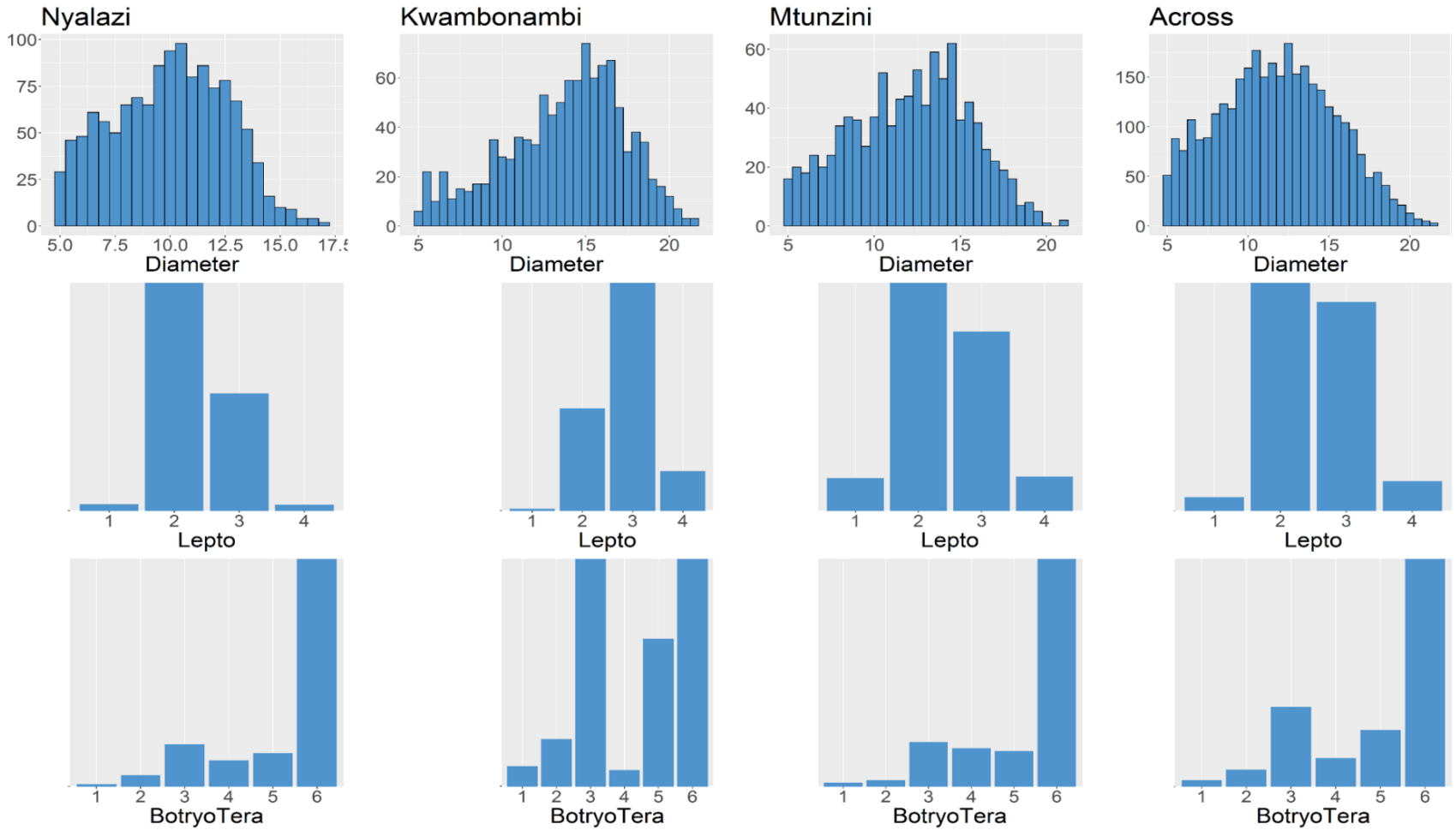


FIGURE 3.5 | Trait distribution plots of the three traits across the trial sites. The *top row* is the *diameter* growth histogram plot, the *middle row* is the histogram plot for *Lepto* incidence scores, and the *bottom row* is the *BotryoTera* incidence scores. The *last column* is the combined distribution of the measured traits across the three sites.

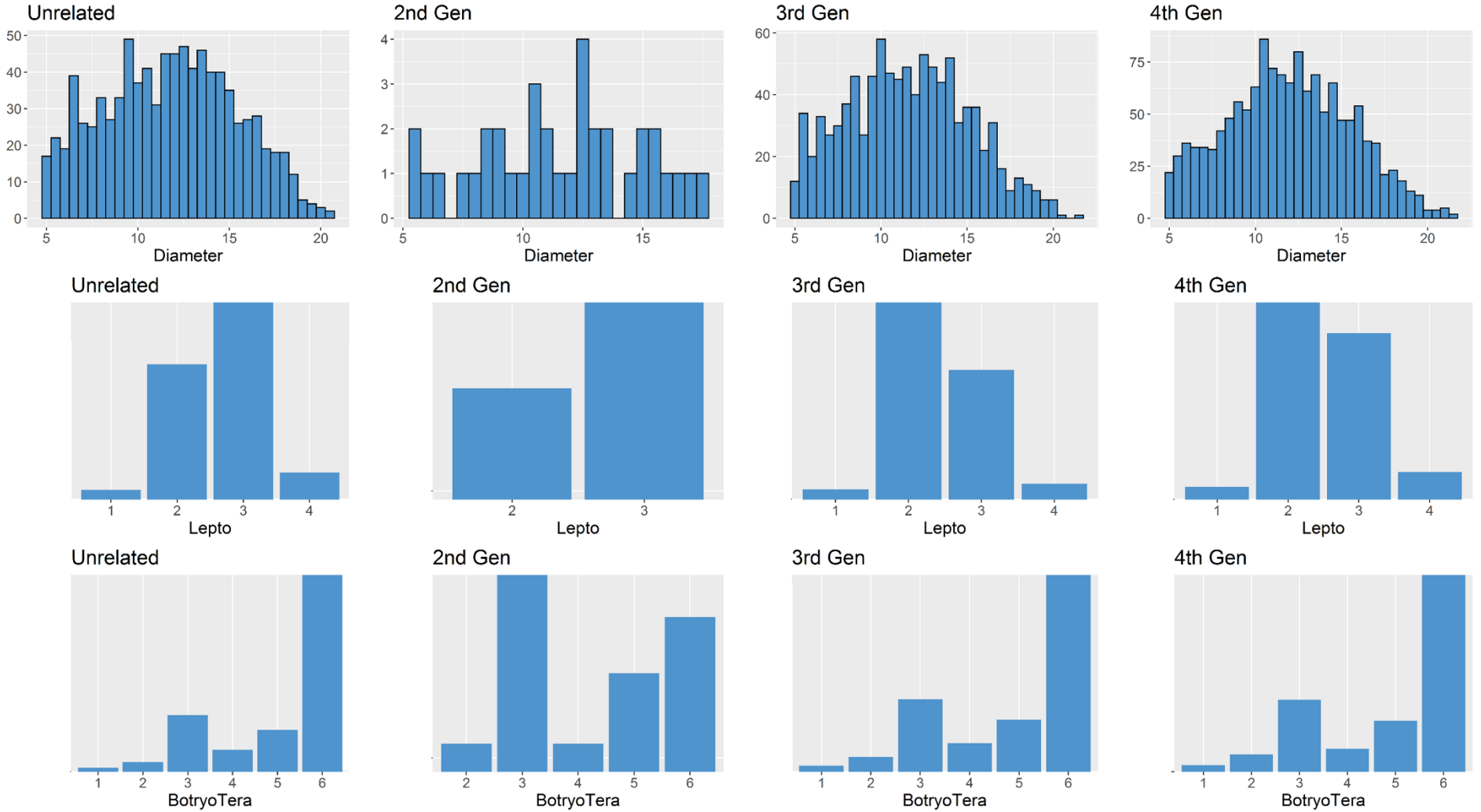


FIGURE 3.6 | Frequency plots across the representative pedigree linked generations and unrelated families. The *top row* is the diameter growth histogram plot, the *middle row* is the histogram plot for *Lepto* incidence scores, and the *bottom row* is the *BotryoTera* incidence scores.

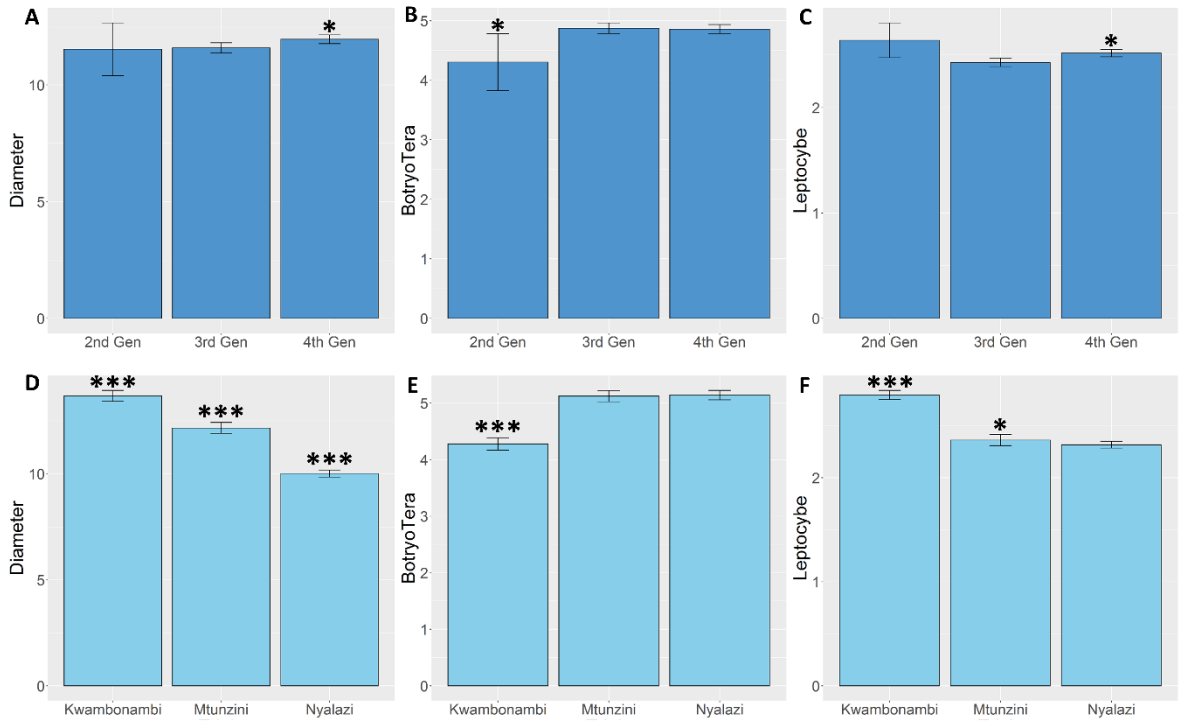


FIGURE 3.7 | Marginal trait means with error bars indicating the 95% confidence interval. (A) Mean diameter growth (cm) for families in the three sites. **(B)** The mean *Lepto* tolerance score for families in the three sites. **(C)** The mean *BotryoTera* tolerance score for families in the three sites. **(D)** Mean diameter growth (cm) for families in the three generations. **(E)** The mean *Lepto* tolerance score for families in the three generations. **(F)** The mean *BotryoTera* tolerance score for families in the three generations. Student *t*-test was performed to ass the significant difference between the means, *p*-value < 0.05 (*), *p*-value < 0.01 (**) and *p*-value < 0.001 (***).

3.4.4 Correlated response based on ssGBLUP breeding value

The direct genetic gains estimated for diameter growth and *Lepto* tolerance were 12.4% and 24.7%, respectively, with *BotryoTera* at 9.8% **(Table 3.6)**. There is an indirect loss of 3.4% in *BotryoTera* tolerance and a gain of 10.0% in *Lepto* tolerance when selecting for diameter growth. Direct selection for *BotryoTera* tolerance would result in an expected indirect loss of 5.6% for diameter growth and 6.5% for *Lepto* tolerance. However, a direct selection of *Lepto* tolerance would result in an expected gain of 6.0% for diameter growth and a loss of 3.8% in *BotryoTera* tolerance **(Table 3.6)**. Together these results illustrate the challenge of achieving

genetic gains for all three of these traits and the need for customised breeding strategies to deal with this challenge.

TABLE 3.6 | Expected genetic gains (%) based on the top 10% selected individuals in the study population. The table shows the indirect response in the expected genetic gains of the paired traits of Diameter growth, *BotryoTera* and *Lepto* tolerance. The bold diagonals are the direct response, with the off-diagonal as the indirect responses.

		Diameter	*BotryoTera*	*Lepto*
ABLUP				
Direct response		15.1	8.9	32.2
Indirect response	Diameter	**15.1**	-3.5	9.9
	BotryoTera	-6.1	**9.3**	-8.0
	Lepto	5.9	-1.9	**26.4**
ssGBLUP				
Direct response		12.4	9.8	24.7
Indirect response	Diameter	**12.4**	-3.4	10.0
	BotryoTera	-5.6	**10.1**	-6.5
	Lepto	6.0	-3.8	**20.9**

3.5 DISCUSSION

Pests and pathogens are significant risk factors in forest plantations (Wingfield et al., 2015). These risk factors are highlighted in African agroforestry systems affecting indigenous and natural forests (Graziosi et al., 2020). Mitigation of these risk factors will require recognising the parallels and synergies in management methods between pest and pathogen studies (Jactel et al., 2020), integration of system genetics and systems biology (Naidoo et al., 2019) in this genomic era (Naidoo et al., 2014). The continued improvement of economic traits such as volume growth, density, and pulp yield in the context of pest and pathogen challenges is vital. Here we combined phenotypic data for a large half-sib breeding trial with genotypic data for a subset of siblings in a single-step genomic BLUP approach to estimate genetic parameters and response to selection for diameter growth, *BotryoTera* and *Lepto* tolerance in the *E. grandis* breeding population. We also proposed a practical genomic selection breeding strategy that is

likely to improve all three traits in *E. grandis* **(Figure 3.8)**. One of the study strengths was the availability of replicated trials with *BotryoTera* infections and *Lepto* infestation across all three sites.

Furthermore, the study benefited from planting pedigree-linked families from three successive generations in the same space and time. Therefore, these trials provided an opportunity to evaluate the outcomes of a direct and indirect artificial selection regime applied in successive generations. A limitation was the inability to score *B. dothidea* and *T. zuluensis* infections separately, which we mitigated by developing a combined phenotypic score **(Figure 3.4)**. Diameter growth, *BotryoTera* and *Lepto* tolerance had moderate heritability estimates (0.25 to 0.36, **Table 3.3**). Diameter growth and *Lepto* tolerance had a strong positive additive genetic correlation. However, both were negatively correlated with *BotryoTera* tolerance, though the correlations were not strong. This presents a challenge to achieve genetic gains in all three traits simultaneously.

3.5.1 Genetic parameter for diameter growth, *Lepto* and *BotryoTera* tolerance

Coefficients of relationships from pedigree data are expectations and do not represent the actual genome shared between relatives, estimated from various allelic frequency parameters (Forni et al., 2011). Forest trees with deep full-sib pedigrees have estimated coefficients of relationships that are much closer to the actual genetic relationships (Bartholome et al., 2016; Chen et al., 2018). However, more precise coefficients of relationships are estimated using DNA markers such as SNPs (Habier et al., 2007; Hayes et al., 2009). When expected genetic relationships are combined with the genome estimated relationships, this precision can be extrapolated to the ***A*** matrix with the blended ***H*** matrix used in ssGBLUP analyses (Legarra et al., 2009; Aguilar et al., 2010). Half-sib pedigree relationships do not include cryptic genetic

relationships in the population, in some instances leading to biased estimation of additive genetic variances (Ratcliffe et al., 2017).

This study generally observed lower heritability estimates from ssGBLUP compared to ABLUP **(Table 3.3)**. Lower additive genetic correlation estimates were also observed for ssGBLUP compared to ABLUP **(Table 3.4)**. Luo et al. (2014) presented heritability estimates of *Lepto* tolerance in *E. camaldulensis* and *E. tereticornis* breeding populations in China of 0.54 and 0.52, respectively. da Silva et al. (2020) also presented heritability estimated from multiple *Eucalyptus* species ranging from 0.27 to 0.68, with *E. grandis* at 0.58. These heritability estimates are similar to what we obtained in our study at 0.54 for *E. grandis* **(Table 3.3)**. The *Lepto* tolerance scores in the study by Luo et al. (2014) were based on the proportion of the canopy affected, with score 0 indicating no symptoms on the canopy and score 4 meaning greater than 75% of the canopy affected (Thu et al., 2009).

In contrast, our scoring system was not based on canopy proportions but rather the severity of gall formation, with score 4 indicating no evidence of gall formation and score 1 indicating lethal outcome from gall formation in both mid-ribs and petioles of the leaves **(Figure 3.3)**. Luo et al. (2014) reported a moderately negative genetic correlation between tree height (at nine months) and *Lepto* susceptibility in *E. camaldulensis* at -0.33 and for *E. tereticornis* at -0.47. Due to the inverted scores used in our study, we report a positive genetic correlation (0.78) with diameter growth at 48 months **(Table 3.4)**. These results suggest that vigorous tree growth is positively related to tolerance to *L. invasa*. Plant growth regulators are well-characterised phytohormones involved in influencing plant development and abiotic stress responses (Wani et al., 2016) and pest tolerance (Harun-Or-Rashid and Chung, 2017). There is

evidence to suggest that the microbiome of the maternal environment may affect the performance of their progeny and tolerance to pathogens in *E. grandis* (Vivas et al., 2017). A study to characterise the relationship of maternal and/or progeny microbiomes, phytohormones, and their interactions, on superior tree growth and health, is warranted.

3.5.2 Genotype-by-environment interaction and trait performance

The mean annual precipitation of the three sites in the subtropical region of South Africa decreases from South to North, tracking the increase in the mean annual temperature maximum **(Figure 3.2)**. Therefore, Nyalazi in the North is on average warmer and drier compared to Mtunzini in the South, which is on average colder and wetter, whereas Kwambonambi has mid-ranged environmental conditions **(Table 3.1)**. The pairwise Type-B genetic correlation for diameter growth, *Lepto* and *BotryoTera* tolerance across the sites ranged from 0.77 to 0.81 **(Table 3.5)**, indicating low G×E interaction. The Nyalazi trial was surrounded by commercial stands of an *E. grandis* × *E. camaldulensis* (G×C) clone that was highly susceptible to *L. invasa*. The G×C hybrid genotype has been shown in the literature to be susceptible to *L. invasa* (Thu et al., 2009; Luo et al., 2014). The G×C clone planted in the Nyalazi site had an increased infestation of *L. invasa*, translating into the high frequency of *Lepto* tolerance score 2 in the trial and a much lower frequency of *Lepto* tolerance score 3 and 4 **(Figure 3.5)**. In Mtunzini, there was also an increased frequency of *Lepto* score 2 **(Figure 3.5)**; however, the trial was surrounded by a tolerant *E. grandis* × *E. urophylla* (G×U) clone. Above-average actively growing shoots in Mtunzini due to its favourable environmental conditions **(Table 3.1)**. These actively growing shoots are targets for *L. invasa* infestation. The heritability estimates of *Lepto* tolerance in Mtunzini and Nyalazi were adjusted lower from 0.35 to 0.24 and 0.71 to 0.38, respectively, by the ssGBLUP model **(Table 3.2)**. It is not clear why the heritability correction in Nyalazi was so significant compared to that in Mtunzini.

In Kwambonambi, the mid-range environmental conditions to Mtunzini and Nyalazi, which was also surrounded by a tolerant G×U clone, *Lepto* tolerance showed similar heritability estimates between ABLUP (0.69) and ssGBLUP (0.70) and for diameter growth ABLUP (0.63) and ssGBLUP (0.58) **(Table 3.2)**. The similar heritability estimates in Kwambonambi of diameter growth and *Lepto* tolerance may result from their relatively high positive additive genetic correlation. The estimated marginal means for diameter growth and *Lepto* tolerance in Kwambonambi further support this relationship **(Figure 3.7D and F)**.

There is an increased incidence of *BotryoTera* tolerance score 3 in Kwambonambi **(Figure 3.5)**, resulting from the increased susceptibility from the second-generation families **(Figure 3.6)**. *BotryoTera* appeared as a fungal stem disease in the mid to late 1990s, which means that the first-generation parents (second-generation families) were selected in the absence of the *BotryoTera* disease explaining the higher susceptibility of the second generation families. The environmental conditions at the Kwambonambi site are optimal for diameter growth, and, due to the negative correlation with *BotryoTera* tolerance, there was high susceptibility to *BotryoTera* in Kwambonambi **(Figure 3.7E)**. Diameter growth, *Lepto*, and *BotryoTera* tolerance in the Kwambonambi site, the mid-range environmental conditions of Nyalazi and Mtunzini seem to reflect the trait performances, corresponding with their additive genetic correlation.

3.5.3 Generational performance for diameter growth, *Lepto* and *BotryoTera* tolerance

Recurrent selection in tree breeding ensures the gradual improvement of target economic traits over generations. Such efforts are under threat from pest and pathogen pressures as well as climate change (Wingfield et al., 2015). Reversing the decline of *E. grandis* in the subtropical

region of South Africa due to *L. invasa* gall wasp and the co-occurrence of *B. dothidea* and *T. zuluensis* fungal stem disease is vital. *BotryoTera* fungal stem disease was discovered and described in South Africa early to mid-1990s (Smith et al., 1994; Wingfield et al., 1996). This meant that selections or evaluations in the government landrace breeding populations did not involve *BotryoTera* tolerance until the first-generation in the 2nd trial series and onwards in the private breeding population **(Figure 3.1)**. Evidenced by the high *BotryoTera* incidence score 3 **(Figure 3.6)** of the second-generation families in particular, in the Kwambonambi site **(Figure 3.5)**. Evaluation for *BotryoTera* tolerance in the second-generation resulted in the increased tolerance in the third-generation and maintained in the fourth-generation **(Figure 3.7B)**. When looking at the high frequency of *BotryoTera* score 6 in **Figure 3.5** and **3.6**, it does suggest that the evaluation strategy has had a limited role to play in improving *BotryoTera* tolerance because this trait seems to have plateaued in the last generations. The limitation of the evaluation strategy for *BotryoTera* tolerance is that selection was only performed within families already selected for diameter growth and further compounded by the fact that *BotryoTera* tolerance is negatively correlated with diameter growth.

L. invasa was reported in South Africa in 2007 (Neser et al., 2007), coinciding with the third-generation tested in the 4th trial series **(Figure 3.1)**. *Leptocybe* appeared when the trial series was at age five. The canopies were already inaccessible for scoring and selecting *Lepto* tolerance for the fourth generation **(Figure 3.1)**. The indirect improvement of *Lepto* tolerance from the third to the fourth-generation is owed to the strong positive additive genetic correlation with diameter growth **(Figure 3.7C)**. This study showed that the recurrent selection strategy successfully improved diameter growth and indirectly improved *Lepto* tolerance, with limited impact on *BotryoTera* tolerance.

3.5.4 Proposed selection strategies for diameter growth and *Lepto* and *BotryoTera* tolerance

Eucalypts, including *E. grandis*, are currently experiencing a decline, mainly due to pest and pathogen pressures for commercial deployment and breeding populations such as *Puccinia psidii* (Silva et al., 2013), *L. invasa* (da Silva et al., 2020), *T. zuluensis* (Wingfield et al., 1996; Aylward et al., 2019), and *B. dothidea* (Smith et al., 1996; Marsberg et al., 2017). This study offers opportunities to revise historical evaluation and selection strategies to improve diameter growth and *BotryoTera* and *Lepto* tolerance. Testing all these pedigree-linked *E. grandis* generations in the same space and time has highlighted the successes and challenges of traditional evaluation and selection strategies and their direct and indirect impact on economic traits over the generations as new pests and pathogens emerge. First, pests and pathogens may appear during a growth stage within a breeding cycle when trees cannot be effectively scored and selected. Second, pests and pathogens affect different parts of the tree, young leaves (early in the growth cycles) and stem (later in the growth cycles); therefore, the correct timing of scoring is crucial. Third, although present, pests and pathogens may differ in their infestation and infection severity due to many factors, leading to highly varying levels of challenge and incomplete expression of tolerance or susceptibility. Fourth, the emergence of pests and pathogens sometimes may reveal inadequacies of already established selection strategies, thereby requiring revision, as is the case for *BotryoTera*.

A multivariate approach to deal with these challenges requires an understanding of the additive genetic correlations of traits. Such a strategy would require turning over a generation in which all three traits were measured on each tree to estimate their between and within family breeding values. The challenge with field trials is that there are often difficult to score pest and pathogen

tolerance accurately, as discussed. Breeders may adopt a multivariate approach to primarily select for diameter growth and indirectly for *Lepto* tolerance and then only consider selecting *BotryoTera* tolerant individuals from high ranked families **(Figure 3.8A)**.

Circumventing field trials and the inconsistency of pest infestations or pathogen infections, tree breeders may consider a proposed serial selection strategy with genomic selection and controlled pollination in potted trials **(Figure 3.8B)**. This approach would require the integration of nursery and field phenotypes to develop a more accurate GS model. Such an approach was demonstrated in *Populus deltoids* for tree height to accelerating its breeding strategies (Alves et al., 2020). The proposed GS approach in this involves challenging potted families with *L. invasa* and scoring *Lepto* tolerance 6 months after potting and then advancing the most tolerant individuals across families for *BotryoTera* tolerance scoring at 12 months after potting. The best individuals from the top *Lepto* and *BotryoTera* tolerant families are then cloned to validate the pest and pathogen tolerance (Set A).

Meanwhile, the second set of ramets from the same clones (Set B) is planted in field trials to validate the expected correlated diameter growth response, while the third set of ramets (Set C) are subjected to flower induction to produce control-pollinated next-generation families are. The clonal phenotypic data can be used together with genome-wide genotyping to train a genomic selection model for implementation **(pink arrows in Figure 3.8B)**. Genomic estimated breeding values and genomic relationship matrices will inform the control pollination (diallel in the potted orchard) (Munoz et al., 2014; Li et al., 2019). This approach should increase the selection intensity and reduce the need for costly controlled pest and pathogen challenges. Thereby fast-tracking clonal tests and producing next-generation control-pollinated

(CP) seedlings (with breeding value predictions for all three traits) to improve gains per unit time over what can be achieved in a traditional open-pollinated (OP) field testing approach.

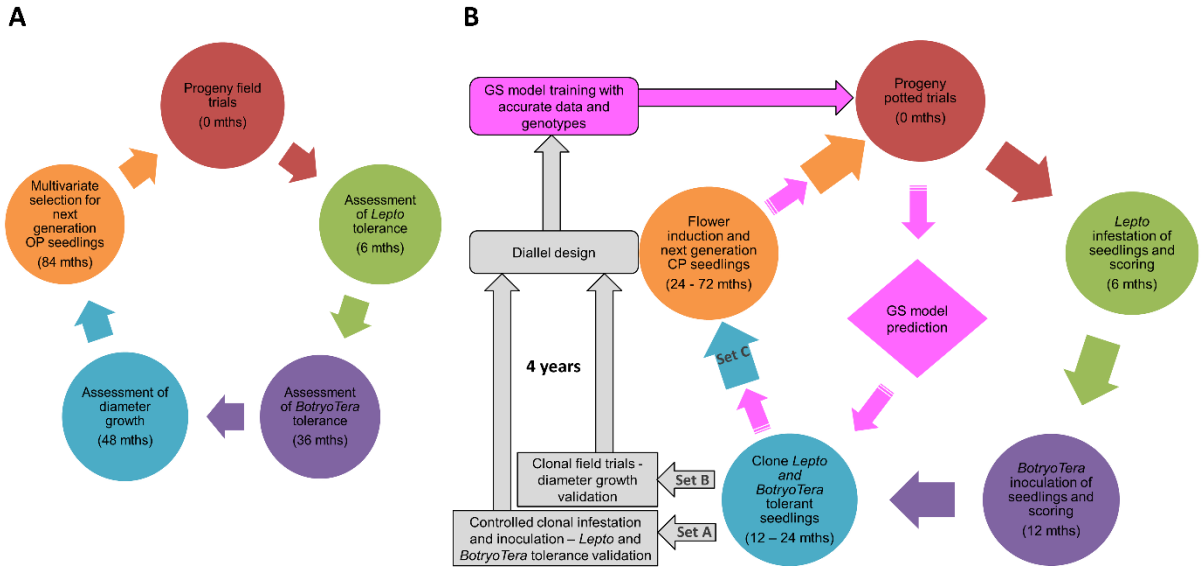


FIGURE 3.8 | Proposed breeding strategies to improve diameter growth under pest and pathogen pressures. (A) Traditional field-based multivariate selection strategy whereby diameter growth (genetically correlated with *Lepto* tolerance) is the target trait. *BotryoTera* tolerance selections are made within top-ranked diameter growth families to produce open-pollinated (OP) families for the next generation. **(B)** Proposed non-field based serial selection strategy in which *Lepto* tolerance and *BotryoTera* tolerance are scored after successive (6 months and 12 months) controlled infestation and inoculations, respectively. Candidate seedlings from within these tolerant families are cloned and used for flower induction (Set C) and generation of CP families for the next generation. Another set of candidate clones is used to validate the *Lepto* and *BotryoTera* tolerance (Set A). The thirds set is then planted in field clonal trials for diameter growth (Set B). Accurate phenotypes from the clonal material and genome-wide genotyping of the clones create an opportunity to train a genomic selection model that can reduce (*pink arrows*) the need for expensive pest and disease phenotyping in the next generation.

3.6 CONCLUSION

Diameter growth and pest and pathogen tolerance are essential components of sustainable plantation forestry. Therefore, a multivariate selection approach informed by their additive genetic correlations is key to improving genetic gains in these traits simultaneously. This study shows that evaluation and selection strategies implemented for *E. grandis* over the past three

generations have succeeded in improving diameter growth and indirectly *Lepto* tolerance, while limited gain was achieved for *BotryoTera* tolerance. We proposed an alternative to the traditional field-based multivariate strategy, which has many challenges mainly limited by the reliability of assessing pest infestations and pathogen infections in the field. The proposed serial genomic selection strategy involves controlled infestations with *Lepto* and inoculations with *BotryoTera* of cloned families in pots to achieve validated and accurate tolerance scores and diameter growth measurements from clonal field trials. This approach will ensure reliable multivariate genomic selection training and development to exploit the additive genetic correlations void phenotyping challenges with field trials. The proposed genomic selection strategy, possibly via ssGBLUP (Misztal et al., 2013), would be a feasible approach to improve diameter growth, *Lepto* and *BotryoTera* tolerance in *E. grandis*.

3.7 REFERENCE

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: A unified approach to utilise phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 92**,** 743-752.

Alves, F. C., Balmant, K. M., Resende Jr, M. F. R., Kirst, M., and de Los Campos (2020). Accelerating forest tree breeding by integrating genomic selection and greenhouse phenotyping. *The Plant Genome* 13**,** e20048.

Aylward, J., Roets, F., Dreyer, L. L., and Wingfield, M. J. (2019). Teratosphaeria stem canker of *Eucalyptus*: two pathogens, one devastating disease. *Molecular Plant Pathology* 20**,** 9-19.

Bartholome, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., and Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. *BCM Genomics* 17**,** 604.

Bihon, W., Slippers, B., Burgess, T. I., Wingfield, M. J., and Wingfield, B. D. (2011). Sources of *Diplodia pinea* endophytic infections in *Pinus patula* and *P. radiata* seedlings in South Africa. *Forest Pathology* 41**,** 370-375.

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

Burgess, T. I., Tan, Y. P., Garnas, J., Edwards, J., Scarlett, K. A., Shuttleworth, L. A., Daniel, R., Dann, E. K., Parkinson, L. E., Dinh, Q., Shivas, R. G., and Jami, F. (2019). Current status of the *Botryosphaeriaceae* in Australia. *Australasian Plant Pathology* 48**,** 35-44.

Bush, S. J., Dittrich-Schroder, G., Neser, S., Gevers, C., Baffoe, K. O., Slippers, B., and Hurley, B. P. (2018). First record of *Quadrastichus mendeli*, a parasitoid of *Leptocybe invasa*, in South Africa. *Southern Forests: a Journal of Forest Science* 80**,** 275-277.

Cappa, E. P., De Lima, B. M., Da Silva-Junior, O. B., Carcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Science* 284**,** 9-15.

Cetinkol, O. P., Smith-Moritz, A. M., Cheng, G., Lao, J., George, A., Hong, K., Henry, R., Simmons, B. A., Healzwood, J. L., and Holmes, B. M. (2012). Structural and chemical characterisation of hardwood from tree species with applications as bioenergy feedstocks. *PLOS One* 7**,** e52820.

Chang, R. L., Arnold, R. J., and Xhou, X. D. (2012). Association between enzyme activity levels in *Eucalyptus* clones and their susceptibility to the gall wasp, *Leptocybe invasa,* in South China. *Journal of Tropical Forest Science* 24**,** 256-264.

Chen, S. F., Pavlic, D., Roux, J., Slippers, B., Xie, Y. L., Wingfield, M. J., and Zhou, X. D. (2011). Characterisation of *Botryosphaeriaceae* from plantation-grown *Eucalyptus* species in South China. *Plant Pathology* 60**,** 739-751.

Chen, Y., and Lubberstedt, T. (2010). Molecular basis of trait correlation. *Trends in Plant Science* 15**,** 454-461.

Chen, Z-Q., Baison, J., Pan, J., Karlsson, B., Anderson, B., Westin, Rosario, M., Gil, G., and Wu, H. X. (2018). Accuracy of genomic selection for growth and wood quality traits in two control pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BCM Genomics* 19**,** 946.

Christensen, O. F., and Lund, M.S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 422**,** 2.

Cortinas, M.N., Koch, N., Thain, J., Wingfield, B. D., and Wingfield, M. J. (2004). First record of the *Eucalyptus* stem canker pathogen, *Coniothyrium zuluense* from Hawaii. *Australasian Plant Pathology* 33**,** 309-312.

Crous, P. W., Groenewald, J. Z, Summerell, B. A., Wingfield, B. D., and Wingfield, M. J. (2009). Co-occurring species of *Teratosphaeria* on *Eucalyptus*. *Persoonia: Molecular Phylogeny and Evolution of Fungi* 22**,** 38-48.

Da Silva, P. H. M., Junqueira, L. R., De Araujo, M. J., Wilcken, C. F., Moraes, M. L. T., and De Paula, R. C. (2020). Susceptibility of eucalypt taxa to a natural infestation by *Leptocybe invasa*. *New Forests* 51**,** 753-763.

Devappa, R. K., Rakshit, S. K., and Dekker, R. F. H. (2015). Forest biorefinery: Potential of poplar phytochemicals as value-added co-products. *Biotechnology Advances* 33**,** 681-716.

Dittrich-Schroder, G., Harney, M., Neser, S., Joffe, T., Bush, S., Hurley, B. P., Wingfield, M. J., and Slippers, B. (2014). Biology and host preference of *Selitrichodes neseri:* A potential biological control agent of the *Eucalyptus* gall wasp, *Leptocybe invasa*. *Biological Control* 78**,** 22-41.

Dittrich-Schroder, G., Wingfield, M. J., Hurley, B. P., and Slippers, B. (2012). Diversity in *Eucalyptus* susceptibility to the gall-forming wasp *Leptocybe invasa*. *Agricultural and Forest Entomology* 4**,** 419-427.

Durand, N., Rodrigues, J. C., Mateus, E., Boavida, C., and Branco, M. (2011). Susceptibility variation in *Eucalyptus* spp. in relation to *Leptocybe invasa* and *Ophelimus maskelli* (Hymenoptera: Eulophidae), two invasive gall wasps occurring in Portugal. *Silva Lusitana* 19**,** 19-31.

Epstein, L., Sukhwinder, K., and Vandergheynst, J. S. (2008). *Botryosphaeria*-related die-back and control investigated in noncoastal California grapevines *Califonia Agriculture* 62**,** 161-166.

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to quantitative genetics.* United Kingdom: Longman Group.

Forni, S., Aguilar, I., and Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* 43.

Gezahgne, A., Roux, J., and Wingfield, M. J., (2003). Diseases of exotic *Eucalyptus* and *Pinus* species in Ethiopia plantations. *South African Forestry Journal* 99**,** 29-33.

Gezahgne, A., Roux, J., Slippers, B., and Wingfield, M. J. (2004a). Identification of the causal agent of *Botryosphaeria* stem canker in Ethiopian *Eucalyptus* plantations. *South African Journal of Botany* 70**,** 241-248.

Gezahgne, A., Roux, J., Thu, P. Q., and Wingfield, M. J. (2004b). *Coniothyrium* stem canker of *Eucalyptus*, new to Argentina and Vietnam. *South African Forestry Journal* 99**,** 578-588.

Gilmour, A. R., Gogel, B.J., Cullis, B.R., Welham, S.J., and Thompson, R. (2015). "ASReml guide release 4.1 functional specifications". 4.1 ed. (Hemel Hempstead, HP1 1ES, UK: VSN International Ltd (www.vsni.co.uk)).

Graziosi, I., Tembo, M., Kuate, J., and Muchugi, A. (2020). Pests and diseases of trees in Africa: A growing continental emergency. *Plants People Planet* 2**,** 14-28.

Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., and Su, S. (2014). Comparison of single-trait and multiple-trait genomic prediction models. *BCM Genetics* 15**,** 30.

Habier, D., Fernando, R. L., and Dekker, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177**,** 2389-2397.

Harun-or-Rashid, M., and Chung, Y. R. (2017). Induction of systemic resistance against insect herbivores in plants by beneficial soil microbes. *Frontiers in Plant Science* 8**,** 1816.

Hayes, B. J., Visscher, P.M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realised relationship matrix. *Genetics Research* 91**,** 47-60.

Hill, W. G., and Weir, B. S. (2010). Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetic Resources* 93**,** 47-64.

Isik, F., Holland, J., and Maltecca, C. (2017). *Genetic data analysis for plant and animal breeding.* Springer International Publishing. ISBN 978-3-319-55177-7.

Jactel, H., Desperz-Loustau, M-L., Battisti, A., Brockerhoff, E., Santini, A., Stenlid, J., Bjorkman, C., Branco, M., Dehnen-Schumutz, K., Douma, J. C., Drakulic, J., Drizou, F., Eschen, R., Franco, J. C., Gossner, M. M., Green, A., Kenis, M., Klapwijk, M. J., Liebhold, A. M., Orazio, C., Prospero, S., Robinet, C., Schroedeer, M., Slippers, B., Stoev, P., Sun, J., Van Den Dool, R., Wingfield, M. J., and Zalucki, M. P. (2020). Pathologists and entomologists must join forces against forest pest and pathogen invasions. *NeoBiota* 58**,** 107-127.

Jimu, L., Chen, S., Wingfield, M. J., and Mwenje, E. (2015). Three genetic groups of the *Eucalyptus* stem canker pathogen *Teratosphaeria zuluensis* introduced into Africa from an unknown source. *Antonie van Leeuwenhoek* 109**,** 1.

Kim, I.-K., Mendel, Z., Protasov, A., Blumberg, D., and Salle, J. L. (2008). Taxonomy, biology, and efficacy of two Australian parasitoids of the eucalyptus gall wasp, *Leptocybe invasa* Fisher & La Salle (Hymenoptera: Eulophidae: Tetrastichinae). *Zootaxa* 1910**,** 1-20.

Klapste, J., Dungey, H. S., Graham, N. J., and Telfer, E. J. (2020). Effect of trait's expression level on single-step genomic evaluation of resistance to *Dothistroma* needle blight. *BCM Plant Biology* 20**,** 205.

Klapste, J., Suontanna, M., Dungey, H. S., Telfer, E. J., Graham, N. J., Low, C. B., and Stovold, G. T. (2018). Effect of hidden relatedness on single-step genetic evaluation in an advanced open-pollinated breeding program. *Journal of Heredity* 109**,** 802-810.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92**,** 4656-4663.

Legarra, A. C., O. F., Aguilar, I., and Misztal, I. (2014). Single Step, a general approach for genomic selection. *Livestock Science* 166**,** 54-65.

Li, Y., Klapste, J., Telfer, E., Wilcox, P., Graham, N., Macdonald, L., and Dungley, H. S. (2019). Genomic selection for non-key traits in radiata pine when the documented pedigree is corrected using DNA marker information. *BCM Genomics* 20**,** 1026.

Lourenco, D. A. L., Tsuruta, S., Fragomeni, B. O., Masuda, Y., Aguilar, I., Legarra, A., Bertrand, J. K., Amen, T. S., Wang, L., Moser, D. W., and Misztal, I. (2015). Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *American Society of Animal Science* 93**,** 2653-2662.

Luo, J., Arnold, R., Lu, W., and Lin, Y. (2014). Genetic variation in *Eucalyptus camaldulensis* and *E. tereticornis* for early growth and susceptibility to the gall wasp *Leptocybe invasa* in China. *Euphytica* 196**,** 397-411.

Ma, P., Lund, M. S., Nielsen, U. S., Aamand, G. P., and Su, G. (2015). Single-step genomic model improved reliability and reduced the bias of genomic prediction in Danish Jersey. *Journal of Dairy Science* 98**,** 9026-9034.

Malan, F. S. (1993). The wood properties and qualities of three South African-grown eucalypt hybrids. *South African Forestry Journal* 167**,** 35-44.

Marsberg, A., Kemler, M., Jami, F., Nagel, J. H., Postma-Smidt, A., Naidoo, S., Wingfield, M. J., Crous, P, W., Spatafora, J. W., Hesse, C. N., Robbetse, B., and Slippers, B. (2017). *Botryosphaeria dothidea*: a latent pathogen of global importance to woody plant health. *Molecular Plant Pathology* 18**,** 477-488.

Mendel, Z., Protasov, A., Fisher, N., and La Salle, J. (2004). Taxonomy and biology of *Leptocybe invasa* gen. & sp. n. (Hymenoptera: Eulophidae), an invasive gall inducer on *Eucalyptus*. *Australian Entomology* 43**,** 101-113.

Mhoswa, L., O'Neill, M. M., Mphahlele, M. M., Oates, C. N., Payn, K. G., Slippers, B., Myburg, A. A., and Naidoo, S. (2020). A Genome-wide association study for resistance to the insect pest *Leptocybe invasa* in *Eucalyptus grandis* reveals genomic regions and positional candidate defence genes. *Plant and Cell Physiology* 61**,** 1285-1296.

Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full and genomic information. *Journal of Dairy Science* 92**,** 4648-4655.

Mohali, S. A., Slippers, B., and Wingfield, M. J. (2007). Identification of *Botryosphaeriaceae* from *Eucalyptus, Acacia* and *Pinus* in Venezuela. *Fungal Diversity* 25**,** 103-125.

Munoz, P. R., Resende Jr, M. F. R., Huber, D. A., Quesada, T., Resende, M. D. V., Naele, D. B., Wegrzyn, J. L., Kirst, M., and Peter, G. F. (2014). Genomic relationship matrix for correcting pedigree errors in breeding populations: Impact on genetic parameters and genomic selection accuracy. *Crop Science* 54**,** 115-1123.

Myburg, A. A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D.,Goodstein, D.M., Dubchak, I., Poliakov, A., Mizrachi, E., Kullan, A.R.K., Hussey, S.G., Pinard, D., Van Der Merwe, K., Signh, P., Van Jaarsveld, I., Silva-Junior, O.B., Togawa, R.C., Pappas, M.R., Faria, D.A., Sansaloni, C.P., Petroli, C.D., Yang, X., Ranjan, P., Tschaplinski, T.J., Ye, C-Y., Li, T., Sterck, L., Vanneste, K., Murat, F., Soler, M., San Clemente, H., Saidi, N., Cassan-Wanh, H., Dunand, C., Hefer, C.A., Bornberg-Bauer, E., Kersting, A.R., Vining, K., Amarasinghe, V., Ranik, M., Naithani, S., Elser, J., Boyd, A.E., Liston, A., Spatafora, J.W., Dharmwardhana, P., Rija, R., Sullivan, C., Romanel, E., Alves-Ferreira, M., Kulheim, C., Foley, W., Carocha, V., Paiva, J., Kudrna, D., Brommonschenkel, S.H., Pasquali, G., Byrne, M., Rigault, P., Tibbits, J., Spokeviciuos, A., Jones, R.C., Steane, D.A., Vaillancourt, R.E., Potts, B.M., Joubert, F., Barry, K., Pappas Jr, G.J., Strauss,

S.H., Jaiswal, P., Grima-Pettenati, J., Salse, J., Van De Peer, Y., Rokhsar, D.S., and Schmutz, J. (2014). The genome of *Eucalyptus grandis*. *Nature* 510**,** 356-375.

Naidoo, S., Kulheim, C., Zwart, L., Mangwanda, R., Oates, C. N., Visser, E. A., Wilhen, F. E., Mamni, T. B., and Myburg, A. A. (2014). Uncovering the defence responses of *Eucalyptus* to pests and pathogens in the genomics age. *Tree Physiology* 34**,** 931-943.

Naidoo, S., Slippers, B., Plett, J. M., Coles, D., and Oates, C. N. (2019). The road to resistance in forest trees. *Frontiers in Plant Science* 10**,** 273.

Neser, S., Prinsloo, G. L., and Neser, O. C. (2007). The eucalypt leaf, twig and stem (sic) galling wasp, *Leptocybe invasa*, now in South Africa. *Plant Protection News* 72**,** 1-2.

Nugnes, F., Gebiola, M., Monti, M. M., Gualtieri, L., Giorgini, M., Wang, J., and Bernardo, U. (2015). Genetic diversity of the invasive gall wasp *Leptocybe invasa* (Hymenoptera: Eulophidae) and of its *Rickettsia* Endosymbiont, and associated sex-ratio differences. *PLOS One* 10**,** e0124660.

Nyeko, P. (2005). The cause, incidence and severity of a new gall damage on *Eucalyptus* species at Oruchinga refugee settlement in Mbarara district, Uganda. *Uganda Jurnal of Agricultural Science* 11**,** 47-50.

Nyeko, P., Mutitu, K. E., Otieno, B. O., Ngae, G. N., and Day, R . K. (2010). Variations in *Leptocybe invasa* (Hymenoptera: Eulophidae) population intensity and infestation on eucalyptus germplasms in Uganda and Kenya. *International Journal of Pest Management* 56**,** 137-144.

Perez, C.A., Altier, N., Simeto, S., Wingfield, M. J., Slippers, B., and Blanchette, R. A. (2008). *Botryosphaeriaceae* from *Eucalyptus* and native Myrtaceae in Uruguay. *Agrociencia* 12**,** 19-30.

Perlack, R. D., Turhollow, W. L .L., Af, G. R. L., Stokes, B. J., Erbach, D. C. (2005). "Biomass as a feedstock for a bioenergy and bioproduct industry: the technical feasibility of a billion-ton annual supply", (ed.) *O.R.N. Laboratory*. (USA: Department of Energy).

Prabhu, J. S. T. (2010). Susceptibility of *Eucalyptus* species and clones to gall wasp, *Leptocybe invasa* Fisher and La Salle (Eulophidae: Hymenoptera) in Karnataka. *Karnataka Journal of Agricultural Science* 23**,** 220-221.

Ratcliffe, B., El-Dien, O. G., Cappa, E . P., Porth, I., Klápste, J., Chen, C., and El-Kassaby, Y. A. (2017). Single-step BLUP with varying genotyping effort in open-pollinated *Picea glauca*. *Genes Genomics Genetics* 7**,** 935-942.

Raymond, C. A. (2002). Genetics of *Eucalyptus* wood properties *Annual Forestry Science* 59**,** 525-531.

Rodas, C. A., Slippers, B., Gryzenhout, M., and Wingfield, M. J. (2009). *Botryosphaeriaceae* associated with *Eucalyptus* canker diseases in Colombia. *Forest Pathology* 39**,** 110-123.

Roux, J., Coutinho, T. A., Wingfield, M. J., and Bouillet, J.-P. (2000). Diseases of plantation *Eucalyptus* in the Republic of Congo. *South African Journal of Science* 96**,** 454-456.

Sangtongpraow, B., Charensom, K., and Siripatanadilok, S. (2011). Longevity, fecundity and development time of *Eucalyptus* gall wasp, *Leptocybe invasa* Fisher and La Salle (Hymenoptera: Eulophidae) in Kanchanaburi province, Thailand *Thai Journal of Agricultural Science* 44**,** 155-163.

Silva-Junior, O. B., Faria, D.A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206**,** 1527-1540.

Silva, P. H. M., Miranda, A. C., Moraes, M. L. T., Furtado, E. L., Stape, J. L., Alvares, C. A., Sentelhas, P. C., Mori, E. S., and Sebbenn, A. M. (2013). Selecting for rust (*Puccinia psidii*) resistance in *Eucalyptus grandis* in Sao Paulo State, Brazil. *Forest Ecology and Management* 303**,** 91-97.

Slippers, B., Burgess, T., Pavlic, D., Ahumada, R., Maleme, H., Mohali, S., Rodas, C., and Wingfield, M. J. (2009). A diverse assemblage of Botryosphaeriaceae infects *Eucalyptus* in native and non-native environments. *Southern Forests* 71**,** 101-110.

Smith, H., Kemp, G. H. J., and Wingfield, M. J. (1994). Canker and die-back of *Eucalyptus* in South Africa caused by *Botryosphaeria dothidea*. *Plant Pathology* 43**,** 1031-1034.

Smith, H., Wingfield, M. J., and Petrini, O. (1996). *Botryousphaeria dothidea* endophytic *Eucalyptus grandis* and *Eucalyptus nitens* in South Africa. *Forest Ecology and Management* 89**,** 189-195.

Stafford, W., De Lange, W., Nahman, A., Chunilall, V., Lekha, P., Andrew, J., Johakimu, J., Sithole, B., Trotter, D. (2020). Forestry biorefineries. *Renewable Energy* 154**,** 461-475.

Thu, P. Q., Dell, B., and Burgess, T. I. (2009). Susceptibility of 18 Eucalypt species to the gall wasp *Leptocybe invasa* in the nursery and young plantations in Vietnam. *ScienceAsia* 35**,** 113-117.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91**,** 4414-4423.

Viggiani, G., Laudonia, S., and Bernardo, U. (2000). The increase of insect pests in *Eucalyptus*. *Informatore Agrario* 58**,** 86-87.

Vivas, M., Kemler, M., Mphahlele, M. M., Wingfields, M. J., and Slippers, B. (2017). Maternal effects on phenotype, resistance and the structuring of fungal communities in *Eucalyptus grandis*. *Environmental and Experimental Botany* 140**,** 120-127.

Wani, S. H., Kumar, V., Shriram, V., and Sah, S. K. (2016). Phytohormones and their metabolic engineering for abiotic stress tolerance in crop plants. *The Crop Journal* 4**,** 162-176.

Wiley, J., And Skelly, P. (2008). Pest Alert: A *Eucalyptus* pest, *Leptocybe invasa* Fisher and LaSalle (Hymenoptera: Eulophidae), genus and species new to Florida and North America. *Florida Department of Agriculture and Consumer Services* 23**,** 359-374.

Wimmer, V., Albrecht, T., Auinger, H-J., and Schon, C-C. (2012). synbreed: a framework for the analysis of genomic predictions data using R. *Bioinformatics* 28**,** 2086-2087.

Wingfield, M. J., Brocherhoff, E. G., Wingfield, B. D., and Slippers, B. (2015). Planted forest health: The need for a global strategy. *Science* 349**,** 832-836.

Wingfield, M. J., Crous, P. W., and Coutinho, T. A. (1996). A serious canker disease of *Eucalyptus* in South Africa caused by a new species of *Coniothyrium*. *Mycopathologia* 136**,** 139-145.

Zhang, M., Zhou, C., Song, Z., Weng, Q., Li, M., Ji, H., Mo, X., Huang, H., Lu, W., Luo, J., Li, F., and Gan, S. (2018). The first identification of genomic loci in plants associated with resistance to galling insects: a case study in *Eucalyptus* L'Hér. (Myrtaceae). *Scientific Report* 8**,** 2319.

Zheng, X-L., Huang, Z-Y., Dong, D., Guo, C-H., Li, J., Yang, Z-D., Yang, X-H., and Lu, W. (2016). Parasitoids of the eucalyptus gall wasp *Leptocybe invasa* (Hymenoptera: Eulophidae) in China. *Parasite* 23**,** 58.

Zhu, F-L., Ren, S-X., Giu, B-L., Huang, Z., and Peng, Z-Q. (2012). The abundance and population dynamics of *Leptocybe invasa* (Hymenoptera: Eulophidae) galls on *Eucalyptus* spp. in China. *Journal of Integrative Agriculture* 11**,** 2116-2123.

CHAPTER 4

COMBINING GENOMIC SELECTION AND ACCELERATED FLOWERING STRATEGY IN *EUCALYPTUS* PLANTATION SPECIES

Makobatjatji M. Mphahlele[1,2], Fikret Isik[3], Gary R. Hodge[3,4], and Alexander A. Myburg[2*]

[1]Mondi Forests, Research and Development Department, Trahar Technology Centre - TTC, P.O. Box 12, Hilton, 3245, South Africa
[2]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa
[3]Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, NC 27695, USA
[4]Camcore, North Carolina State University, P.O. Box 7626, Raleigh, NC 27695, USA

Author contributions:

This data chapter has been prepared for publication in the format of a manuscript for the international ISI journal, Frontiers Plant Science. I carried out the experimental design, data collection, data analysis, and drafting of this paper as part of his PhD thesis. F.I. assisted with the modelling and scripts of the data analysis tools. F. I., G. H., and A. A. M. supervised the study and helped with data interpretation and the drafting of the manuscript. All authors read, edited and approved the final manuscript.

4.1 ABSTRACT

A significant benefit of genomic selection (GS) in forest trees is to reduce breeding cycle times and increase gains per unit time. Flowering time (seed-to-seed) of many forest tree species is a crucial factor contributing to the length of breeding cycles. In this study, we investigated the benefits of combining GS with transgenic *Flowering locus T* (FT)-scion for early floral induction (GS-FT) transgrafting approach in the precocious *Eucalyptus grandis* and non-precocious *E. dunnii*, two economically important plantation species. We simulated traditional breeding (TB), GS and GS-FT in the species and compared their transitional cost ratio and their benefit-cost ratio in terms of genetic gain. Implementation of GS and GS-FT compared to TB strategy was 10 to 13.8 fold more expensive for both species. Whereas the transition from GS to GS-FT strategy in both species is 1.2 fold more expensive. This resulted in a punitive benefit-cost ratio implementing GS and GS-FT strategies compared to the TB strategies. The implementation of the GS-FT, compared to the GS strategy, had a profitable benefit-cost ratio. With the adoption of the GS-FT strategy, which effectively removes flowering time as a barrier, similar breeding cycle times are achieved in both species, but *E. dunnii* benefits more due to the significant reduction in breeding cycle time. Our simulation provides a first indication of the potential benefits of implementing FT floral induction in conjunction with GS breeding strategies in plantation forest species.

4.2 INTRODUCTION

Tree breeding requires the application of genetics, reproductive biology, economics and management principles to achieve improved performance of forest trees. These genetic improvements, ranging from yield increases to environmental adaptation, pest and pathogen tolerance and improved wood quality, are achieved by increasing genetic gains per unit of time and cost. A typical forest tree breeding program starts with the selection of plus trees from planted progeny trials. Plus trees are multiplied for commercial deployment based on the appropriate reproductive system either via seedlings (from seed orchards) or rooted cuttings (from vegetative propagation). The operational efficiency of tree breeding strategies often involves adopting approaches such as grafting to propagate superior trees in species that are difficult to multiply using rooted cuttings and/or moving superior genotypes into environments more conducive to flowering. Knowledge of reproductive biology is critical for conducting breeding operations and shortening the breeding cycle (seed-to-seed).

Reproductive maturity in trees can be defined by the presence of fully developed reproductive organs. The development of reproductive maturity is regulated and modulated by environmental conditions through integrating complex regulatory networks and signal pathways (Mouradov et al., 2002; Cho et al., 2017). Flowering time can be affected by many factors, including the dosage of a stimulus, plant developmental stage, nutrient availability, ambient temperature, drought, salinity, exogenously applied hormones and chemicals, and pathogenic microbes (Cho et al., 2017) and biotic and abiotic stresses (Kazan and Lyons, 2016). Photoperiod is one of the critical factors controlling seasonal flowering time via signal transduction pathways involved in circadian clock function (Singh et al., 2017). Vernalisation, which is also subject to genetic influences, is another process whereby flowering time is induced by exposure to low temperatures (Kim and

Sung, 2014). A similar process was studied in *Eucalyptus nitens*, whereby chill units resulted in floral induction (Gardner and Bertling, 2005). High altitude with low temperature, thinning trees early and planting edge trees can also influence flowering (Graca, 1987).

Breeders would like to have some control over the development of reproductive maturity (referred to in this manuscript as "flowering time") to reduce the time required to complete breeding cycles. Ionescu et al. (2017) has reviewed less-studied metabolites and exogenously applied chemicals that may influence the transition to flowering and chemical compounds stimulating flower opening. The growth regulator paclobutrazol has been used successfully in *Eucalyptus* species to shorten breeding cycles by promoting precociousness in target species. Paclobutrazol reduces vegetative growth and enhances flower-bud production of several species with low and infrequent flower production, such as *E. globulus* (Griffin et al., 1993), *E. nitens* (Moncur and Hasan, 1994; Gardner and Bertling, 2005), *E. smithii* (Gardner et al., 2016) and *E. dunnii* (Arnold and Dongyun, 2003). The flowering of *E. dunnii* has no association with tree age. Paclobutrazol treatment alone or in combination with fertiliser has been used to stimulate flower buds and increase bud numbers per tree (Arnold and Dongyun, 2003). However, where environmental conditions did not favour natural flower induction of *E. dunnii*, no flowers were observed regardless of chemical treatment, highlighting the need to place selections into environments conducive to floral induction.

First developed in China before 2000 BCE, grafting was well established by 500 BCE and practised in agriculture and horticulture (Mudge et al., 2009). Grafting artificially conjoins different vascular systems (i.e., rootstock and scion) for multiple purposes, such as asexual

propagation, to increase pest, disease and abiotic tolerance (using tolerant rootstocks) and to alter plant vigour, architecture and precocity (Aloni et al., 2010). The earliest experiments with grafting for floral induction in a non-woody plant was with soybean (Heinze et al., 1942), and in woody plants was in pine (Righter, 1939). There is overwhelming evidence of signal molecule movement essential for plant development, such as proteins and RNAs, via the phloem (Song et al., 2015). Some of these molecules (e.g., water, hormones, sugars, nutrients, RNAs, and proteins) contribute to rootstock-scion interactions, including physiological interactions (Martinez-Bellesta et al., 2010). Such interactions were shown to influence fruit quality (Goncalves et al., 2006; Tietel et al., 2020). Grafting non-flowering scion onto flowering rootstock was shown to have a genotype-dependent effect on flowering in scions in Cassava (Adeyemo et al., 2017) and sweet potato (Mubayiwa et al., 2016).

One of the well-studied proteins/peptides that function as a signal molecule is *Flowering locus* T (FT) produced in leaves to regulate plant flowering and tuber development (Matsuda et al., 2009; Zhang et al., 2010; Navarro et al., 2011; Song et al., 2013; Klocko et al., 2016; Adeyemo et al., 2017; McGarry et al., 2017; Kinoshita and Richter, 2020). FT is produced in the phloem companion cells of young shoot leaves under the transcription factor CONSTANS regulation. FT is then translocated to adjacent sieve tube elements and transported to the apical meristem where it combines with bZIP transcription factor *Flowering locus* D (FD) (Abe et al., 2005) to form an FT-FD transcription factor module which then interacts with APETALA1 and LEAFY to convert vegetative buds to floral buds (Pineiro and Coupland, 1998; An et al., 2004; Abe et al., 2005; Eriksson et al., 2006; Corbesier et al., 2007; Li et al., 2019).

Transgrafting refers to transgenic molecules or their bio-products moving across the graft union, transmitting transgenic influences from scion or rootstock to non-transgenic rootstock or scion, respectively. Transgrafting ensures that the transgraft product such as seeds are free of the *FT* transgenes since only the signal molecule (FT) movement induces flowering in the non-transgenic transgraft plant parts. Therefore, there is a high potential for commercial and breeding exploitation of precocious floral induction without *FT* transgene biosafety risks. FT protein expressed in transgenic scion was shown to move across the graft union in tomatoes (Lifschitz et al., 2006) and potatoes (Navarro et al., 2011) to induce flowering of the rootstock. *FT* expressed in transgenic *Poplar* rootstock was unable to induce flowering of the scion (Zhang et al., 2010), suggesting some difficulty of FT translocation from scion to rootstock floral induction process in woody plants. FT translocation from rootstock to scion would require active transportation (Yoo et al., 2013). However, recently citrus rootstock expressing the endogenous FT3 protein under an *Arabidopsis* phloem specific promoter (*AtSUC2*) induced flower development translocating from rootstock to non-transgenic scion (Soares et al., 2020). The result suggests that a specific endogenous *FT* gene orthologue under tissue-specific promoters is essential for rootstock to scion translocation of FT floral induction.

Transgenic approaches to stimulate flowering, particularly *FT* overexpression in *Eucalyptus* plants, resulted in very early flower development within months in potted plants (Klocko et al., 2016). There are numerous examples in woody plants, in which expressing *FT* or *FT*-like transgenes promoted flowering, in apple (Trankner et al., 2010), cassava (Adeyemo et al., 2017), blueberry (Song et al., 2013), pears (Matsuda et al., 2009), kiwifruit (Moss et al., 2018), citrus (Soares et al., 2020), and tree species such as *Populus* (Zhang et al., 2010; Hsu et al., 2011). Myburg et al. (2019) postulated that *FT*-constructs

introduced into elite parental selections used for intraspecific or interspecific hybrid breeding can accelerate flowering in tree breeding and bioengineering strategies. However, the time gained by accelerated flowering may be offset by the control pollination process to segregate the transgenic construct and deploy non-transgenic products.

Precocious flowering is critical to shortening breeding cycles, and in conjunction with GS breeding strategies, it would result in improved genetic gains per unit time in *Eucalyptus* (Mphahlele et al., 2020). This study focuses on two *Eucalyptus* plantation species, *E. grandis*, a precocious flowerer and *E. dunnii*, a non-precocious flowerer. The study uses a simulated approach to investigate the benefits and costs associated with the practical breeding steps required for implementing a traditional breeding (TB) strategy, genomic selection (GS) strategies, and a combination of GS and FT-scion transgraft induction (GS-FT) to accelerate flowering and further reduce breeding cycle times. We report on the benefits of genetic gain per unit time and cost for each of the three breeding strategies in *E. grandis* and *E. dunnii*. The outcomes of this study will inform breeders and managers of the benefits, costs and practical implications to maximise genetic gains using the GS-FT approach.

4.3 MATERIALS AND METHODS

4.3.1 Introducing E. grandis and E. dunnii

E. grandis (Rose gum) and *E. dunnii* (Dunn's white gum) originate in coastal ranges of Australia, mainly New South Wales and Queensland. Both are crucial commercial plantation forest species planted worldwide. *E. dunnii* is cold and drought-tolerant compared to *E. grandis* (Arnold and Dongyun, 2003). Another distinction is their flowering behaviour in the South African environment. *E. grandis* can flower under trial spacing (typically 3m × 2m),

starting from two years onwards for some individuals until rotation age at eight years. *E. dunnii* does not flower under typical trial spacing and duration. It requires grafting from the trial into an orchard spacing (typically 6m × 5m) and a conducive environment, where some individuals can start to flower from 4 years onwards. This means that within a typical trial planting cycle, a tree breeder can turn over a breeding generation for *E. grandis*. In contrast, with *E. dunnii,* the tree breeder would need to graft the superior trees from the trials into a clonal seed orchard (CSO) located in a conducive flowering environment before turning over the generation. This adds several years to the breeding cycle time of *E. dunnii.*

4.3.2 Study simulation parameters

The detailed operational steps and their assigned costs involved in the nursery and field were considered for the breeding cycles of *E. grandis* **(Appendix A, Figure 4.1)** and *E. dunnii* **(Appendix B, Figure 4.2)** over a set time and summarized in **(Table 4.1)**. Nursery operations included seed sowing, rooting of cuttings, potting of plants, grafting, and transgrafting. Field operations included progeny and clonal trials, diameter and wood quality assessment. Other nursery operations included the GS model prediction of seedlings. A selection index model of diameter growth, density and pulp yield are the traits of interest used in both species (Ceron-Rojas et al., 2015). We assumed a 5% compound increase in genetic gains for every completed breeding cycle (seed-to-seed) across the different breeding strategies in the study. We considered zero cost to infrastructure and human resources in the cost simulation parameters and the development and acquisition of the *FT*-overexpressed scion transgenic material. We also considered a discounted cost for compliance with the Biosafety standards when handling *FT*-scion overexpressing transgrafted plants. The *Eucalyptus* (EUChip60K) SNP chip, as described by (Silva-Junior et al., 2015) available from GeneSeek (Neogen, Lansing, MI, USA), was used to present the current genotypic cost. We further compare the TB, GS and the GS-FT breeding strategies for *E. grandis* and

E. dunnii with the diminishing cost of the *Eucalyptus* (EUChip60K) SNP chip from 100% at the current market value to 25%.

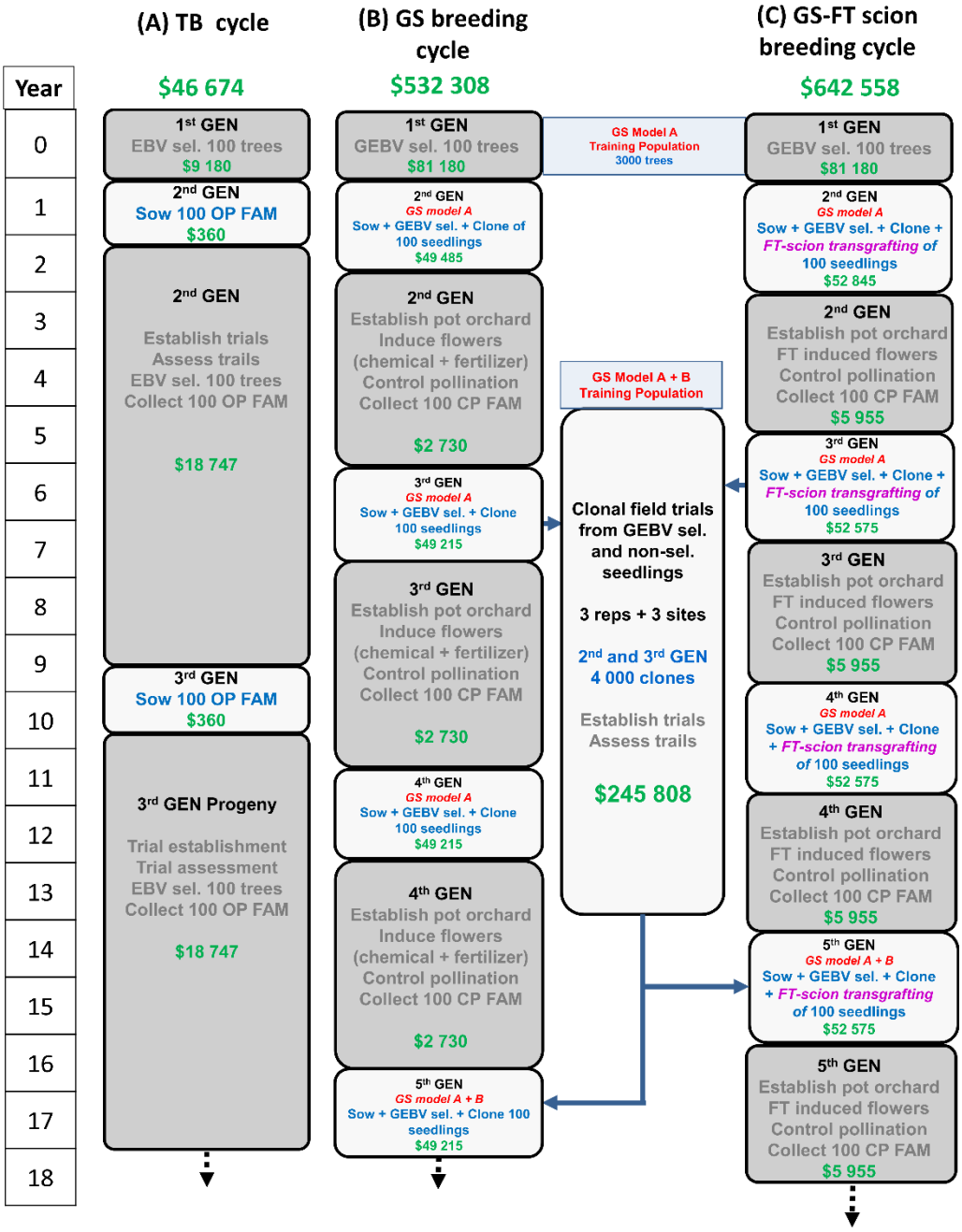


FIGURE 4.1 | Comparison of operational breeding steps and associated cost between breeding strategies for *E. grandis*. (A) Traditional, (B) genomic selection (GS) and (C) genomic selection with FT-scion transgraft (GS-FT) breeding cycles over 17 years. The comparison includes a clonal trials series to update the GS model A to GS model A + B. The costs reflect 100% of the marker value for genotyping with the *Eucalyptus* EUChip60K SNIP chip.

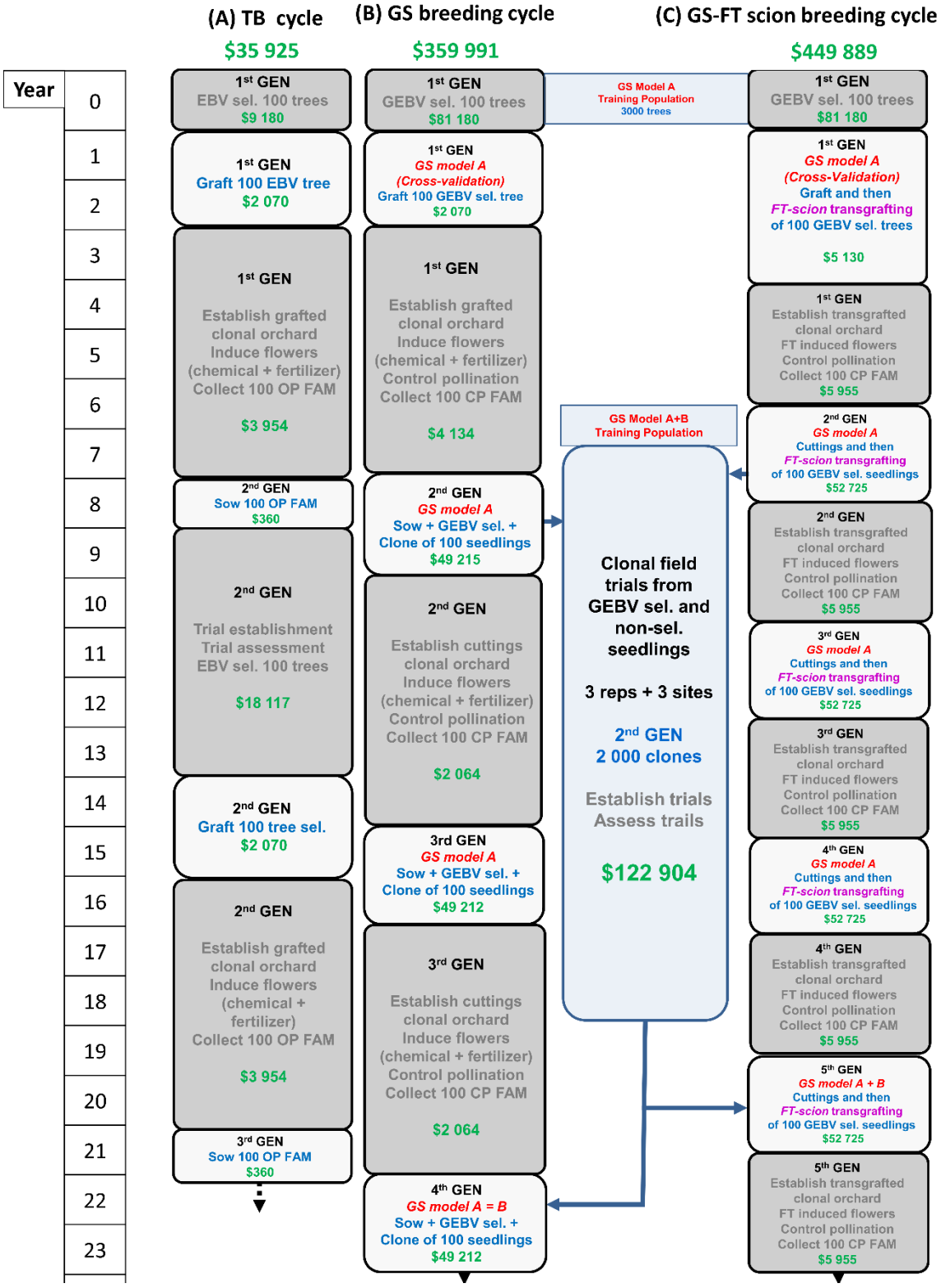


FIGURE 4.2 | Comparison of operational breeding steps and associated cost between breeding strategies for *E. dunnii*. (A) Traditional, (B) genomic selection (GS) and (C) genomic selection with FT-scion transgraft (GS-FT) breeding cycles over 23 years. The comparison includes a clonal trials series to update the GS model A to GS model A + B. The costs reflect 100% of the marker value for genotyping with the *Eucalyptus* EUChip60K SNIP chip.

TABLE 4.1 | Summary of the operational breeding steps and the associated costs between breeding strategies. The proportional costs between genotyping and operations are indicated. The cost ratio of the breeding strategies are also indicated.

Breeding strategies	Breeding strategy cost				Years
Cost of genotyping from curret value	**100%**	**75%**	**50%**	**25%**	
E. grandis					
TB strategy	$ 46 674	$ 46 674	$ 46 674	$ 46 674	17
GS breeding strategy	$ 532 308	$ 466 308	$ 400 308	$ 334 308	17.5
Operational expenses	*50%*	*58%*	*67%*	*80%*	
Genotyping expenses	*50%*	*42%*	*33%*	*20%*	
GS-FT strategy	$ 642 558	$ 578 250	$ 512 250	$ 446 250	18
Operational expenses	*59%*	*66%*	*74%*	*85%*	
Genotyping expenses	*41%*	*34%*	*26%*	*15%*	
Cost ratio: GS vs TB	*11.4*	*10.0*	*8.6*	*7.2*	
Cost ratio: GS-FT vs TB	*13.8*	*12.4*	*11.0*	*9.6*	
Cost ratio: GS-FT vs GS	*1.2*	*1.2*	*1.3*	*1.3*	
E. dunnii					
TB strategy	$ 35 925	$ 35 925	$ 35 925	$ 35 925	20.5
GS breeding strategy	$ 359 991	$ 305 991	$ 251 991	$ 197 991	23
Operational expenses	*40%*	*47%*	*57%*	*73%*	
Genotyping expenses	*60%*	*53%*	*43%*	*27%*	
GS-FT strategy	$ 449 889	$ 383 889	$ 317 889	$ 251 889	23.5
Operational expenses	41%	48%	58%	74%	
Genotyping expenses	59%	52%	42%	26%	
Cost ratio: GS vs TB	*10.0*	*8.5*	*7.0*	*5.5*	
Cost ratio: GS-FT vs TB	*12.5*	*10.7*	*8.8*	*7.0*	
Cost ratio: GS-FT vs GS	*1.2*	*1.3*	*1.3*	*1.3*	

2.7.1 Studied families and genomic selection parameters

It is common practice that *E. grandis* and *E. dunnii* progeny trials are established as half-sibs progeny. Therefore, the proposed simulated three breeding strategies for both species were initiated as half-sib progeny trials (first generation) of 300 families replicated across three sites. The training of the GS model A involved 3000 progeny from families across the three trials. In the subsequent generations, the 300 families are reduced to 100 whilst maintaining maximum genetic diversity. In the traditional breeding strategy, genetic diversity was maintained at the family level, tracking only the maternal lineage.

In contrast, with GS-based breeding strategies, genome-wide molecular markers are used to track allelic diversity (Rodriguez-Ramilo et al., 2015). All the genotyped seedlings and predicted with the GS model but un-selected are retained and planted in clonal trials to validate and update GS model A into GS model A+B. These clonal trials also serve to select candidate commercial clones for deployment.

4.3.3 Simulated *E. grandis* breeding strategies and cycles

E. grandis, in practice, has a reasonably less challenging breeding cycle. It flowers readily in field trial conditions. At rotation age which is around eight years, most of the canopies in the trials would have reached reproductive maturity to turn over the generation, which means that breeders can match trait maturity with reproductive maturity. A single breeding cycle (seed-to-seed) with nursery activities should take eight and a half years **(Figure 4.3A)**. Two traditional breeding cycles of *E. grandis* should be completed within eighteen years **(Figure 4.4A)**. In comparison, the GS breeding strategy does not require field trials. However, it still requires the induction of flowering with growth regulators such a paclobutrazol to accelerate flowering in non-trial conditions (potted orchard) **(Figure 4.3B)**. The shortest reasonable seed-to-seed time with this GS strategy for *E. grandis*, including nursery operations, is five years **(Figure 4.4B)**. To achieve an accelerated flowering time to the shortest biologically allowable breeding cycle, we proposed the transgrafting FT-scion in the GS strategy (GS-FT) **(Figure 4.3C)**. A single GS-FT breeding cycle should take about 4.5 years and include all nursery operations (sowing, cuttings, and transgrafting). Over eighteen years, four GS-FT breeding cycles should be completed **(Figure 4.4C)**. The traditional breeding strategy involved open-pollination families, whereas the GS and the GS-FT strategy involved control-pollinated families, similarly in the *E. dunnii* breeding strategies.

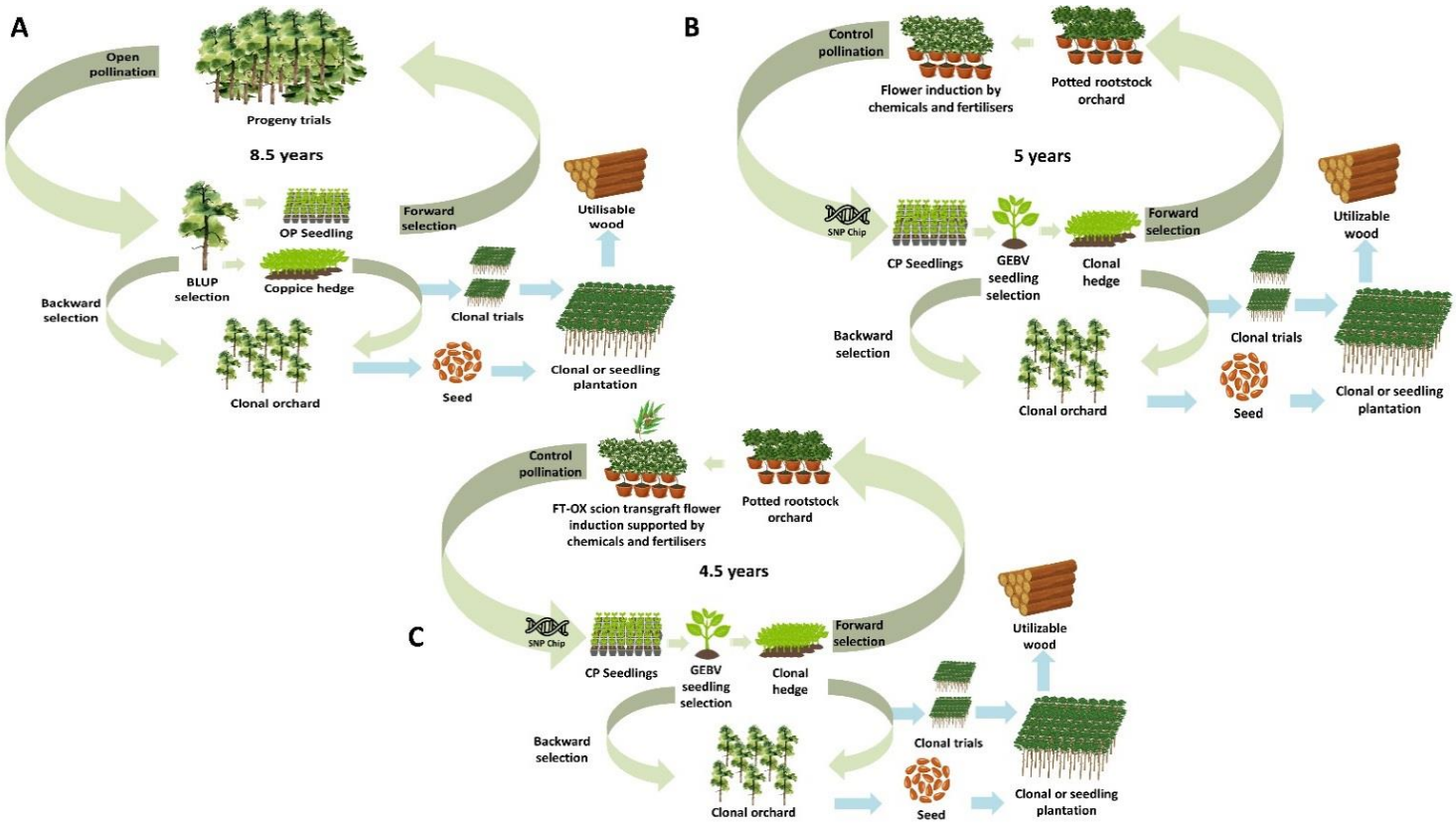


FIGURE 4.3 | *E. grandis* breeding cycle with commercial deployment operations. (A) Traditional breeding cycle. The breeding cycle starts with an open-pollinated progeny trial. The cycle takes 7.5 years of progeny trial and one year of nursery operations. **(B)** Genomic selection breeding cycle. Open-pollinated progeny trial replaced with a potted orchard to produce control pollinated seedlings. The cycle takes four years of flower induction of cuttings in potted clonal seed orchard and one year of nursery operations. **(C)** Genomic selection breeding cycle with *FT*-scion transgraft. The marker-based selected seedlings serve as rootstock for the *FT*-scion transgrafting, replacing the open-pollinated progeny trials. Transgraft potted orchard produces control pollination seeding. The cycle takes 2.5 years of *FT*-scion flower induction in transgrafted potted orchards and two years of nursery operations.

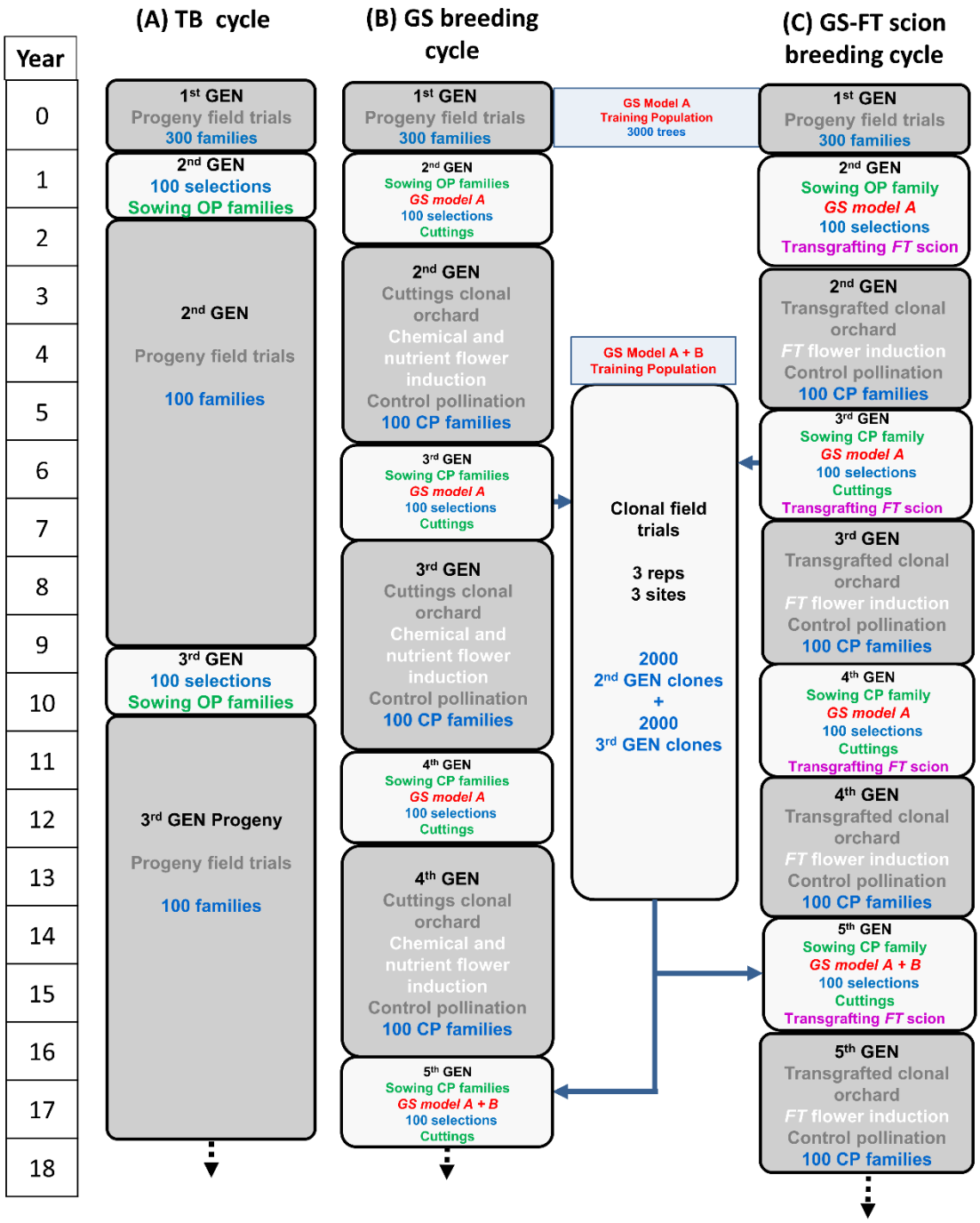


FIGURE 4.4 | Comparison of traditional, genomic selection (GS) breeding and genomic selection with FT-scion transgraft (GS-FT) breeding cycles over 18 years for *E. grandis*. **(A)** Two complete eight-year TB cycles and **(B)** three complete four-year GS breeding cycles, and **(C)** four complete GS-FT breeding cycles. The GS and GS-FT breeding cycles include full-sib clonal field trials to validate and update the current GS model A into GS model A+B. Genotyped and GS model A predicted seedlings (4000 of which 100 per generation were selected) from the 2nd and 3rd generation are retained and planted in a replicated to updated GS model A to GS model A + B.

4.3.4 Simulated *E. dunnii* breeding strategies and Mixed model analysis

E. dunnii is a shy flowerer and is challenging to breed compared to *E. grandis* as it does not flower under trial conditions. Its traditional breeding strategy involves planting a progeny trial, and after eight years, the plus trees are grafted into a CSO. There is a plausible expectation that all CSO selections would flower (graft-to-seed) within five years under favourable environmental conditions with effective fertigation and growth retardant treatment. The CSO serves two essential functions, the relocation of plus trees to a favourable flower-inducing environment and ensuring a pollen cloud between the plus trees. Seed sourced from the CSO will establish the next generation **(Figure 4.5A)**. Since *E. dunnii* does not flower under trial conditions, the breeding cycle can be reduced by grafting plus trees at five years as opposed to eight years. This strategy is conceivable because half rotation phenotypes in forest trees are strongly correlated with full rotation (Osorio et al., 2003). Furthermore, harvesting scion from the five-year-old plus trees for grafting will not compromise the trial integrity. Therefore there is an opportunity to assess full rotation data and re-evaluate any plus trees if necessary. This study considered the reduced five-year grafting cycle as opposed to the eight-year cycle to represent the traditional breeding strategy. A single traditional breeding cycle for *E. dunnii* will take about 13 years (**Figure 4.5A)**. This means that only one and a half breeding cycles can be completed within 23 years **(Figure 4.6A)**. Initially, the GS breeding strategy for *E. dunnii* involves similar steps to the TB strategy, in the form of grafting plus trees into the CSO. Instead of establishing field trials, seedlings from the CSO will be predicted with the GS model, and the GEBV plus seedlings are then cloned into potted orchards and clonal trials **(Figure 4.5B)**. A single GS breeding cycle (seed-to-seed) for *E. dunnii* with nursery operation should take seven years **(Figure 4.5B),** and over 23 years, two GS breeding cycles can be completed **(Figure 4.6B)**. Transgrafting of the FT-scion within the GS strategy would accelerate flowering time for *E. dunnii*, resulting in a single breeding cycle (seed-to-seed) of 4.5 years **(Figure 4.5C)**. There

are four breeding cycles completed over 23 years, including the nursery operations (sowing, cuttings and transgrafting) **(Figure 4.6C)**.

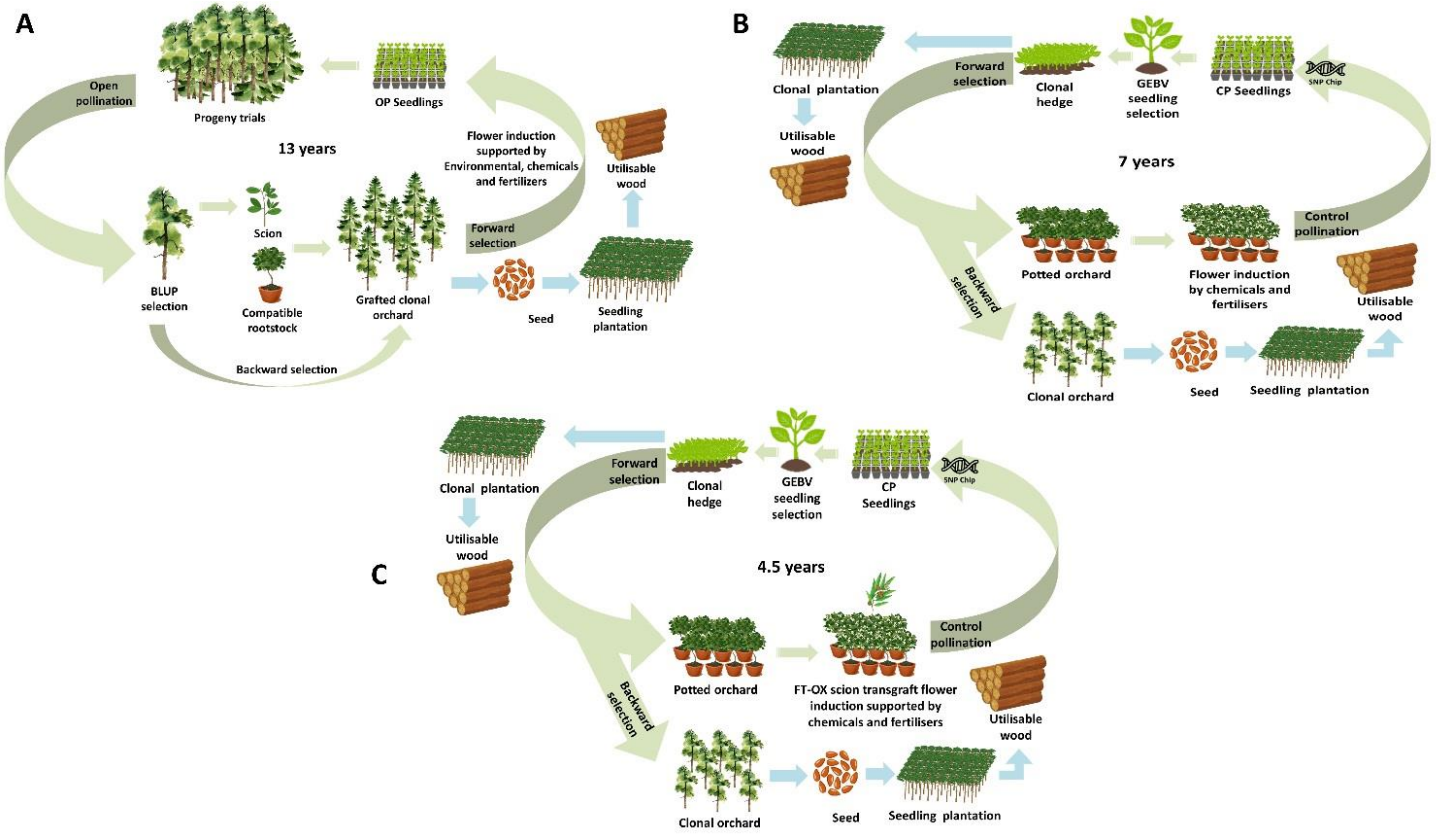


FIGURE 4.5 | E. dunnii breeding cycle with commercial deployment operations. (A) Traditional breeding cycle. The breeding cycle starts with an open-pollinated progeny trial. The cycle takes 13 years with five years of progeny trial, three years of grafting and five years of floral induction in the grafted orchard. **(B)** Genomic selection breeding cycle. Open-pollinated progeny trial replaced with a potted orchard to produce control pollinated seedlings. The cycle takes seven years with two years of nursery operations and five years of floral induction in the potted orchard. **(C)** Genomic selection breeding cycle with *FT*-scion transgraft. The marker-based selected seedlings serve as rootstock for the *FT*-scion transgrafting, replacing the open-pollinated progeny trials. Transgraft potted orchard produces control pollination seeding. The cycle takes 2.5 years of *FT*-scion flower induction in transgrafted potted orchards and two years of nursery operations.

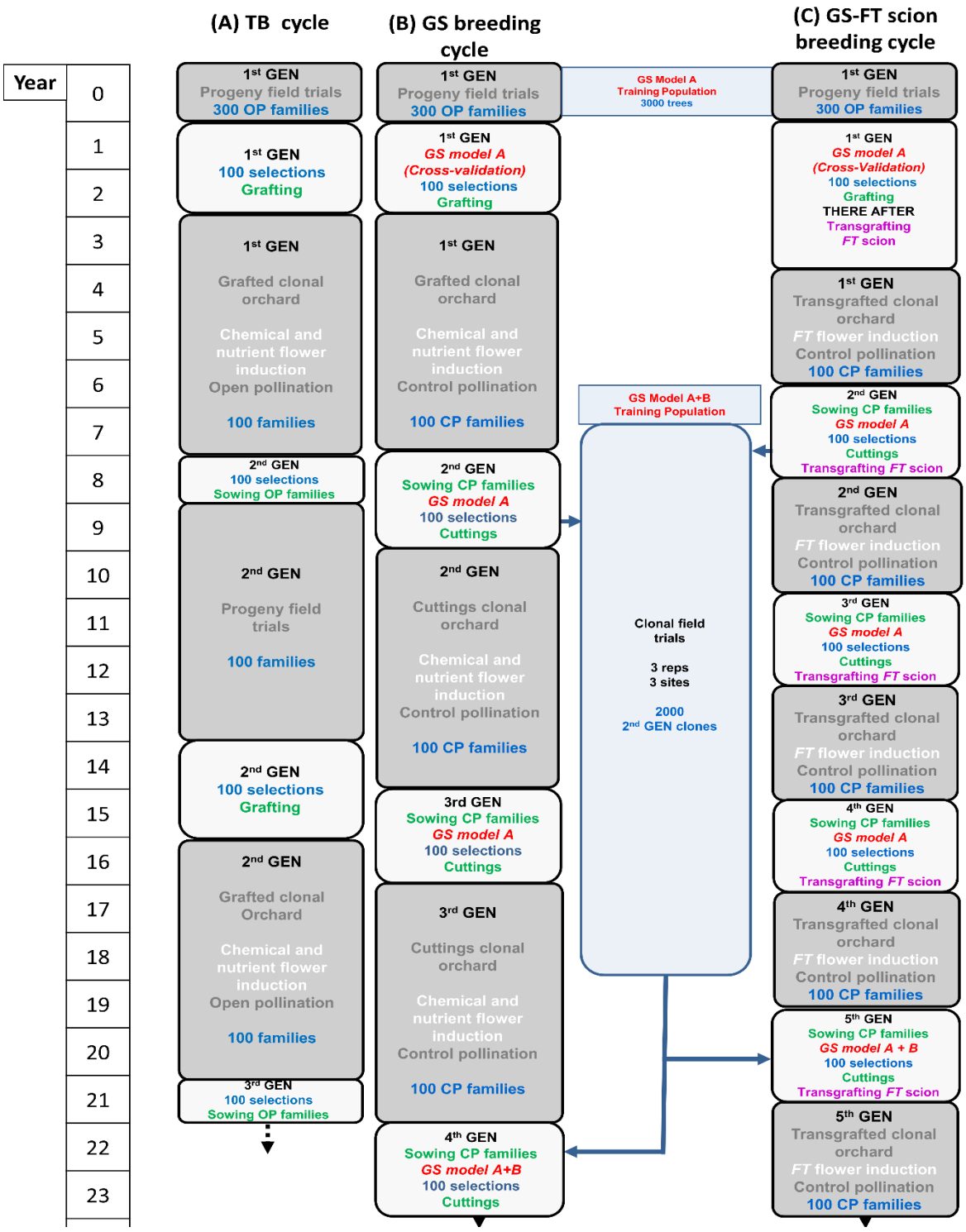


FIGURE 4.6 | Comparison of traditional, genomic selection (GS) breeding and genomic selection with FT-scion transgraft (GS-FT) breeding cycles over 23 years for *E. dunnii*. **(A)** A twelve-year TB cycle. **(B)** Two seven-year GS breeding cycles. **(C)** Four GS-FT breeding cycles. The GS and GS-FT breeding cycles include half-sib clonal field trials to validate and update the current GS model A into GS model A + B. Seedlings selected based on their GEBV were converted into clones from the 2nd generation.

4.4 RESULTS

Innovations in tree breeding strategies need to be realistic and sustainable, which means that their proposed benefits need to match or exceed the cost and time investment. These innovations may reduce time, but they require adjustment of operational efforts, and breeders have to account for the cost of implementing such innovations, hence the need to qualify and quantify where possible these input and output measures. Recent innovations in tree breeding include the ongoing application of genomic selection (GS, (Grattapaglia et al., 2011)) and the potential use of accelerated flowering technologies such as floral induction by FT overexpression (Klocko et al., 2016). In this study, we simulate and compare three breeding strategies, namely traditional breeding, GS and the combination of GS and FT induction in *E. grandis* and *E. dunnii*. For the comparison, we considered a standard period of 18 years in *E. grandis* and the number of traditional, GS and GS-FT cycles that can be fit into this period. We performed the same comparison for 23 years in *E. dunnii*. The genetic gains and costs of operational steps and genotyping were calculated for each species, including nursery and field operations. The combination of all this data enables the interrogation of genetic gain benefit over time and cost for these breeding strategies for *E. grandis* and *E. dunnii*.

4.4.1 The relative effect of genotyping cost on breeding strategies

Genotyping is an essential component of genomic selection breeding strategies. The genome-wide markers generated are used to obtain the GEBV of trees to effect selection replacing trial planting and phenotyping. The proportional cost of genotyping relative to the operational breeding steps is vital for tree breeders pursuing GS breeding strategies compared to TB strategies. Such an analysis would give breeders a relative idea of additional cost implications considering the adjusted breeding operations since trial planting and phenotyping are circumvented. There is a reasonable expectation that genotyping costs will, over time, reduce

with improved genotyping technologies with higher throughput and complexity reduction per unit cost (Perkel, 2008). It should be noted that operational costs are subject to inflation increases. Their relative importance compared to genotyping cost may therefore be understated over time.

There is a 50% split between genotyping and operational cost at 100% genotyping cost to implement the GS strategy in *E. grandis* **(Figure 4.7A)**. The proportion of genotyping cost reduces to 20% at 25% genotyping cost. The proportion of genotyping cost for *E. dunnii* GS breeding strategy is 60% at 100% genotyping cost, reducing to 27% at 25% genotypic cost **(Figure 4.7B)**. In contrast to the GS strategy, the GS-FT strategy involves additional transgrafting breeding steps to the operation breeding steps for *E. grandis*. This meant that the proportion of genotypic cost at 100% genotypic cost is 40% reducing to 15% at 25% genotypic cost for *E. grandis* **(Figure 4.7C)**. In the GS-FT strategy for *E. dunnii*, the proportion of genotyping cost is 59% and lowers to 27%, with the reduction of the genotyping cost to 25% **(Figure 4.7D)**. In the GS and GS-FT breeding strategy for *E. grandis* **(Figure 4.7A and C)**, there is an equal or lower proportion of genotypic costs regardless of the cost of genotyping. In contrast, with *E. dunnii* the GA and GS-FT strategy **(Figure 4.7B and D)**, the additional transgrafting floral induction steps mean that the proportion of genotypic cost starts above the operational cost, reaching a break-even point at approximately 75% of the genotyping cost and then becoming less than that operational cost.

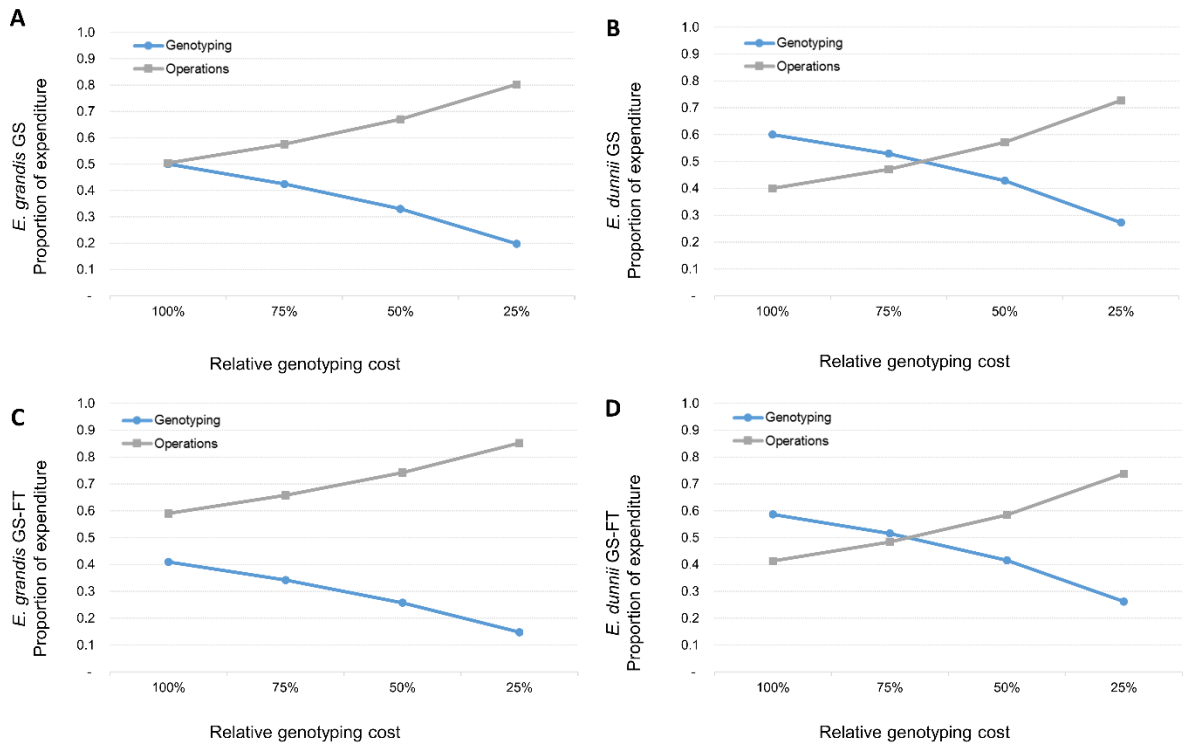


FIGURE 4.7 | Proportional contribution of genotyping and operational breeding steps to overall expenditure over 18 years (*E. grandis*) and 23 years (*E. dunnii*). The genomic selection strategy of *E. grandis* **(A)** and *E. dunnii* **(B).** The genomic selection strategy with *FT*-scion transgrafting floral induction in *E. grandis* **(C)** and *E. dunnii*. **(D).** The *y-axis* shows the proportion of expenditure, and the *x-axis* shows the relative cost of genotyping to the current prices.

4.4.2 Genetic gain benefit per unit time of the breeding strategy

Breeding cycle time is an important consideration in breeding strategies. The reduction of the breeding cycle while simultaneously improving genetic gains is essential to successful breeding strategies. The relative success of achieving this is more difficult in non-precocious species such as *E. dunnii*. Comparing all three strategies, we see that *E. dunnii* has a higher genetic benefit per unit time than *E. grandis* **(Figure 4.8)**. The adoption of the GS-FT compared to the TB strategy represents the highest benefit per unit time for both species **(Figure 4.8)**. Marker-based tree breeding approaches and accelerated floral induction are beneficial to tree

breeding by reducing the breeding cycle time, ensuring improved genetic gain benefit over time.

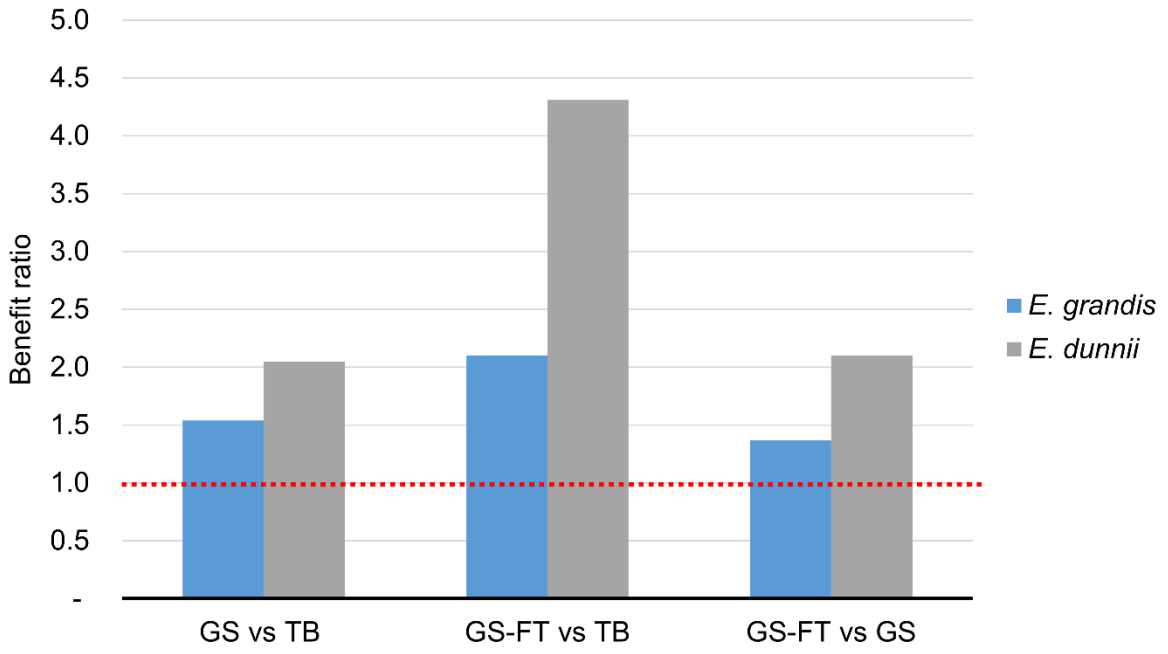


FIGURE 4.8 | Benefit ratio per unit time of the different breeding strategies over 18 years (*E. grandis*) and 23 years (*E. dunnii*). The y-axis shows the genetic gains benefit ratio, and the x-axis shows the compared breeding strategies, traditional breeding (TB), genomic selection (GS) and genomic selection with *FT*-scion transgrafting floral induction (GS-FT). The *red dotted line* indicates the benefit ratio break-even point at 1.0. Genotyping cost is set at 100% of current prices.

4.4.3 The cost ratio between the breeding strategies

The comparative cost of the different breeding strategies is an essential consideration for tree breeders. The questions answered are what are the additional resources that are needed in order to implement a new strategy compared to the current or another strategy. There is a 10 fold increase in cost to implement the *E. dunnii* GS strategy compared to its TB strategy with *E. grandis* at 11.4 fold **(Figure 4.9)**. The implementation of the GS-FT strategy compared to the TB strategy will result in a 13.8 and 12.5 fold increase in resources for *E. dunnii* and *E. grandis*, respectively **(Figure 4.9)**. In contrast, there is much of a likeness in cost ratio for implementing

the GS-FT strategy compared to the GS strategy for both species because the only difference is the transgrafting operation, which is relatively inexpensive **(Figure 4.8)**. This might be significantly higher should we consider GM regulatory compliant infrastructure and operational requirements.

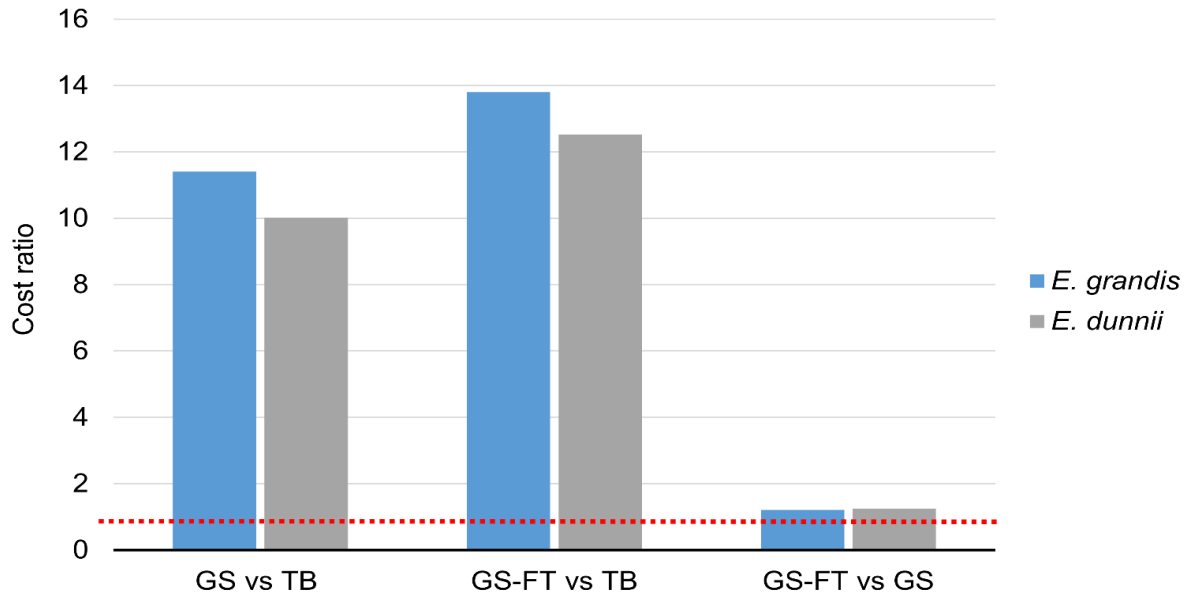


FIGURE 4.9 | Cost ratio of the different breeding strategies over 18 years (*E. grandis*) and 23 years (*E. dunnii*). The *y-axis* shows the cost ratio, and the *x-axis* shows the compared breeding strategies, traditional breeding (TB), genomic selection (GS) and genomic selection with *FT*-scion transgrafting floral induction (GS-FT). The *red dotted line* indicates the cost ratio break-even point at 1.0. Genotyping cost is set at 100% of current prices.

4.4.4 The relative genetic gain benefit to cost ratio of the breeding strategy

It is essential to consider the cost of any breeding strategy. However, the more pertinent question is the return on investment. What is the benefit of the additional cost associated with the breeding strategy compared to others? Implementing the GS vs TB strategy resulted in an unfavourable relative genetic gain benefit to cost ratio, similar to the GS-FT vs TB strategy **(Figure 4.10)**. In contrast, the GS-FT vs GS strategy implementation resulted in a favourable relative genetic gain benefit to the breeding strategies cost ratio **(Figure 4.10)**. In all the benefit-cost ratio analyses, *E. dunnii* had a better ratio than *E. grandis*.

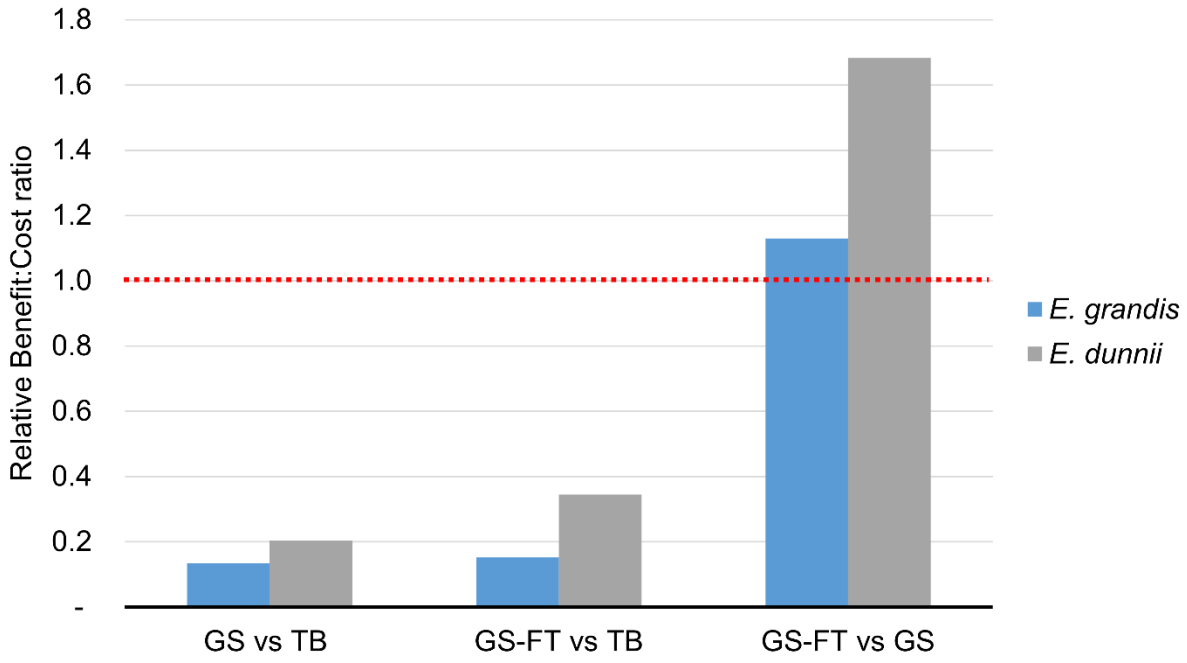


FIGURE 4.10 | The relative genetic gain benefit to cost ratio of the different breeding strategies over 18 years (*E. grandis*) and 23 years (*E. dunnii*). The *y-axis* shows the cost ratio, and the *x-axis* shows the compared breeding strategies, traditional breeding (TB), genomic selection (GS) and genomic selection with *FT*-scion transgrafting floral induction (GS-FT). The *red dotted line* indicates the cost ratio break-even point at 1.0. Genotyping cost is set at 100% of current prices.

4.5 DISCUSSION

Tree breeding is a technical operation in which breeding strategies are implemented, with operational and infrastructure adjustments including investing in innovative technologies such as GS (Resende et al., 2017; Mphahlele et al., 2020). This study measured the benefit of increased genetic gains by considering three breeding strategies and their associated genotyping and operational costs. The study simulated the three breeding strategies for *E. grandis* and *E. dunnii* and compared them against each other. For both species, their TB strategies were set as the base strategy. The GS breeding strategy reduced the breeding cycle time (seed-to-seed) by predicting individual genetic merits and eliminating field trial planting and phenotyping. The GS-FT breeding strategy also eliminated field trial planting and

phenotyping and incorporated FT-scion transgrafting to accelerated floral induction of the GS breeding strategy, further reducing the breeding cycle time.

The flowering precociousness of *E. grandis* ensures that there are two, three and four breeding cycles over 18 years for the TB, GS and the GS-FT strategies, respectively **(Figure 4.4)**. The non-precocious *E. dunnii*, particularly under trial conditions, needs to be grafted into suitable environments to promote flowering. Because of this additional step, one and a half, two and four breeding cycles over 23 years are completed for the TB, GS and GS-FT strategies, respectively **(Figure 4.6)**. Furthermore, with the accelerated breeding strategy the GS-FT, in which trial planting is exempt, and flowering is achieved within the shortened biological limit, 4.5 years is the most optimal breeding cycle for *E. grandis* **(Figure 4.3C)** and *E. dunnii* **(Figure 4.5C)**.

There are differences in the proportion of the genotyping cost between the species and the GS and GS-FT breeding strategies. The proportion of genotyping cost for *E. grandis* GS matches the operational cost **(Figure 4.7A)** and is lower in the GS-FT strategy because of the additional transgrafting of the FT-floral induction, increasing the proportion of operational breeding step costs **(Figure 4.7B)**. Meaning that at 100% of genotyping cost, the proportional operational cost of implementing the GS and GS-FT strategy for *E. grandis* is matched and higher than the genotypic proportion. There are fewer operational breeding steps with more waiting time for floral induction in the GS and GS-FT strategies for *E. dunnii*, which has a longer breeding cycle time because of its non-precociousness. Hence at 100% genotyping cost, the proportion of genotyping cost exceeds the operational cost **(Figure 4.7C and D)**. As genotyping costs reduce, the reverse is observed. As expected, the reduction in genotyping cost over time will

result in a lower proportion of the GS and GS-FT breeding strategies attributed to genotyping cost vs the cost of the operational breeding step.

Improving genetic gains is one of the core objectives of a breeding programme. The adoption of various breeding strategies and technologies do facilitate and maximise this objective. This is achieved in numerous ways, including the reduction of the breeding cycle. The TB as a base strategy for breeding cycle time was compared to the GS and the GS-FT strategies. There was a clear benefit in improved genetic gain per unit time for the GS and GS-FT strategies than the TB strategy for *E. grandis* and *E. dunnii* **(Figure 4.8)**. The benefit in genetic gain per unit time was more for *E. dunnii* than *E. grandis* because of the more significant reduction in the breeding cycle time for *E. dunnii* **(Figure 4.5)** compared to *E. grandis* over the set periods.

The relative cost of implementing one strategy over the other presents the opportunity to weigh the magnitude of the resources needed for implementing any of the breeding strategies. *E. dunnii* has a lower cost ratio than *E. grandis* in transitioning from the TB to the GS strategy. However, the transition requires more than 10 fold investment in both species **(Figure 4.9)**. A similar trend is observed with the transition to the GS-FT from the TB strategy, requiring above 12 fold investment for both species **(Figure 4.9)**. The GS strategy for *E. grandis* has three breeding cycles **(Figure 4.4B)**. In contrast, *E. dunnii* has two cycles with grafting **(Figure 4.6B)**. The cost of the GS breeding strategy for both species is proportionately similar. The GS-FT breeding cycle time for *E. grandis* **(Figure 4.3C)** and *E. dunnii* **(Figure 4.5C)** are similar at 4,5 years, resulting in the cost ratio between GS and GS-FT strategies for both species approximating to one. The two strategies are thus cost-neutral for both species **(Figure 4.9)**. The transition from TB strategies to genotyping-based breeding strategies in both species, such

as GS and GS-FT, requires considerable investment in both genotyping costs and adjustments to the operational breeding steps. This significant investment is even higher for the GS-FT strategy due to the transgrafting component. However, when the base breeding strategy is genotyping-based, there is virtually no additional cost burden transition from the GS to the GS-FT strategy in both species. Therefore, the additional investment in accelerated floral induction technology to further reduce the breeding cycle time is warranted when the base "*traditional*" breeding strategy is GS.

Next, we assessed if the cost ratio between the different breeding strategies penalises the benefit of the improved genetic gains. The significant investment needed for the TB strategy transition to the GS or the GS-FT strategy comes at a cost penalty to genetic gains in both species **(Figure 4.10)**. In contrast, the transition from the GS to the GS-FT strategy does not. *E. dunnii* has a higher relative benefit to cost ratio than *E. grandis* when comparing the transition between the three breeding strategies. The reason for this is because there is more breeding cycle time saving for *E. dunnii* when the transition from the TB (13 years) to the GS-FT strategy (4.5 years) **(Figure 4.5)**, compared to *E. grandis* TB (8 years) to the GS-FT strategy (4.5 years) **(Figure 4.3)**. It is essential to note that with the accelerated breeding strategy (GS-FT) for both species, there is no distinction between the breeding cycles **(Figure 4.3C and 4.5C)**. This highlights the importance of floral induction technologies in genotyping-based breeding strategies to maximise genetic gains per unit time **(Figure 4.8)** through breeding cycle reduction. Similarly, vegetative propagation, cuttings, grafting and transgrafting of FT transgenic scion are essential technologies that allow operational breeding cycle adjustments to reduce breeding times and achieve requisite efficiency in the breeding cycle time savings in both precocious and non-precocious *Eucalyptus* species.

4.6 CONCLUSION

Species precociousness is an essential consideration for their breeding strategy. It affects the operational breeding steps and the breeding cycle time. Genotyping-based breeding strategies such as GS can reduce the breeding cycle time by eliminating trials and phenotyping. However, genotyping-based strategies can complicate the operational breeding steps with the addition of genotyping costs. Accelerating floral induction to its biological limit using the *FT*-transgene within a GS breeding strategy will create more pressure and additional costs. The cost transition from the TB strategy to the genotyping-based strategies are prohibitive and come at a cost penalty to the genetic gains. Therefore, such a transition needs to be managed very well. There is a cost-neutral transition from the GS to the GS-FT strategy and no cost penalty to the benefit of increased genetic gain. This means that tree breeders need to find a cost-effective way to transition from the TB to the GS strategy before the transition to the GS-FT breeding strategy.

4.7 REFERENCE

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., Ichinoki, H., Notaguchi, M., Goto, K., and Araki, T. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309**,** 1052-1056.

Adeyemo, O. S., Chavarriaga, P., Tohme, J., Fregene, M., Davis, S. J., and Setter, T. L. (2017). Overexpression of Arabidopsis *FLOWERING LOCUS T (FT)* gene improves floral development in cassava (*Manihot esculenta*, Crantz). *PLOS One* 12**,** e0181460.

Aloni, B., Cohen, R., Karni, L., Aktas, H., and Edelstein, M. (2010). Hormonal signalling in rootstock–scion interactions. *Scientia Horticulturae* 127, 119-126.

An, H., Roussot, C., Suarez-Lopez, P., Corbesier, L., Vincent, C., Pineiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C., and Coupland, G. (2004). *CONSTANS* acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. *Development* 131**,** 3615-3626.

Arnold, R., and Dongyun, X. (2003). "Environmental and cultural influences on flowering of *Eucalyptus dunnii*", in *Eucalypts in Asia. Proceedings of an international conference.* (ed.) J.W. Turnbull. (Zhanjiang, Guangdong, People's Republic of China, 7-11 April: ACIAR Proceedings).

Ceron-Rojas, J. J., Crossa, J., Arief, V. N., Basford, K., Rustkoski, J., Jarguin, D., Alvarado, G., Beyene, Y., Semagn, K., and Delacy, I. (2015). A genomic selection index applied to simulated and real data. *Genes Genomics Genetics* 5**,** 2155-2164.

Cho, L-H., Yoon, J., and An. G. (2017). The control of flowering time by environmental factors. *The Plant Journal* 90**,** 708-719.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., and Coupland, G. (2007). FT protein movement contributes to long-distance signalling in floral induction of *Arabidopsis*. *Science* 316**,** 1030-1033.

Eriksson, S., Bohlenius, H., Moritz, T., and Nilsson, O. (2006). GA4 is the active gibberellin in the regulation of *LEAFY* transcription and *Arabidopsis* floral initiation. *The Plant Cell* 18, 2172-2181.

Gardner, R. A. W., and Bertling, I. (2005). Effect of winter chilling and paclobutrazol on floral bud production in *Eucalyptus nitens*. *South African Journal of Botany* 71**,** 238-249.

Gardner, R. A. W., Bartling, I., Savage, M. J., and Naidoo, S. (2016). Investigating optimal site conditions for flower bud production in *Eucalyptus smithii* orchards in South Africa. *Australian Forestry* 79**,** 137-146.

Goncalves, B., Mountinho-Pereira, J., Santos, A., Silva, A. P., Bacelar, E., Correia, C., and Rosa, E. (2006). Scion–rootstock interaction affects the physiology and fruit quality of sweet cherry. *Tree Physiology* 26**,** 93-104.

Graca, M. E. C. (1987). Evaluation of flowering and seed production potential of *Eucalyptus dunnii* in Brazil. *Boletim de Pesquisa Florestal* 14**,** 1-12.

Grattapaglia, D., Macos, D., and Resende, D.V. (2011). Genomic selection in forest tree breeding. *Tree Genetics &Genomes* 7**,** 241-255.

Griffin, A. R., Whiteman, P., Rudge, T., Burgess, I. P., and Moncur, M. (1993). Effect of paclobutrazol on flower-bud production and vegetative growth in two species of *Eucalyptus*. *Canadian Journal of Forest Research* 23**,** 640-647.

Heinze, P. H., Parker, M. W., and Borthwick, H. A. (1942). Floral initiation in Biloxi soybean as influenced by grafting. *Botanical Gazette* 103**,** 518-530.

Hsu, C-Y., Adams, J. P., Kim, H., No, K., Ma, C., Strauss, S. H., Drnevich, J., Vandevelde, L., Ellis, J. D., Rice, B. M., Wickett, N., Gunter, L. E., Tuskan, G. A., Brunner, A. M., Page, G. P., Barakat, A., Carlson, J. E., Depamphillis, C. W., Luthe, D. S., and Yuceer, C. (2011). *FLOWERING LOCUS T* duplication coordinates reproductive and vegetative growth in perennial poplar. *PNAS* 108**,** 10756-10761.

Ionescu, I. A., Moller, B. L., and Sanchez-Perez, R. (2017). Chemical control of flowering time. *Journal of Experimental Botany* 68**,** 369-382.

Kazan, K., and Lyons, R. (2016). The link between flowering time and stress tolerance. *Journal of Experimental Botany* 67**,** 47-60.

Kim, D-H., and Sung, S. (2014). Genetic and epigenetic mechanisms underlying vernalisation *The Arabidopsis Book* 12**,** e0171.

Kinoshita, A., and Richter, R. (2020). Genetic and molecular basis of floral induction in *Arabidopsis thaliana*. *Journal of Experimental Botany* 71**,** 2490-2504.

Klocko, A. L., Ma, C., Robertson, S., Esfandiari, E., Nilsson, O., and Strauss, S. H. (2016). *FT* overexpression induces precocious flowering and normal reproductive development in *Eucalyptus*. *Plant Biotechnology Journal* 14**,** 808-819.

Li, D., Zhang, H., Mou, M., Chen, Y., Xiang, S., Chen, L., and Yu, D. (2019). Arabidopsis class II TCP transcription factors integrate with the FT–FD module to control flowering. *Plant Physiology* 181**,** 97-111.

Lifschitz, E., Eviatar, T., Rozman, A., Shalit, A., Goldshmidt, A., Amsellem, Z., Alvarez, J. P., and Eshed, Y. (2006). The tomato *FT* ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *PNAS* 103**,** 6398-6403.

Martinez-Bellesta, M., Alcaraz-Lopez, C., Muries, B., Mota-Cadenas, C., and Carvajal, M. (2010). Physiological aspects of rootstock–scion interactions. *Scientia Horticulturae 127*, 112-118.

Matsuda, N., Ikeda, K., Kurosaka, M., Takashina, T., Isuzugawa, K., Endo, T., and Omura, M. (2009). Early flowering phenotype in transgenic pears (*Pyrus communis* L.) Expressing the *CiFT* Gene. *Journal of the Japanese Society for Horticultural Science* 78**,** 410-416.

Mcgarry, R. C., Klocko, A. L., Pang, M., Strauss, S. H., and Ayre, B. G. (2017). Virus-induced flowering: An application of reproductive biology to benefit plant research and breeding. *Plant Physiology* 173**,** 47-55.

Moncur, M. W., and Hasan, O. (1994). Floral induction in *Eucalyptus nitens*. *Tree Physiology* 14**,** 1303-1312.

Moss, S. A. M., Wang, T., Voogd, C., Brian, L. A., Wu, R., Hellens, R. P., Allan, A. C., Putterill, J., and Varkonyi-Gasic, E. (2018). AcFT promotes kiwifruit in vitro flowering when overexpressed and Arabidopsis flowering when expressed in the vasculature under its own promoter. *Plant Direct* 2**,** 1-12.

Mouradov, A., Cremer, F., and Coupland, G. (2002). Control of flowering time: Interacting pathways as a basis for diversity. *The Plant Cell* 14**,** S111-S130.

Mphahlele, M. M., Isik, F., Mostert-O'neill, M. M., Raynolds, S. M., Hodge, G. R., and Myburg, A. A. (2020). Expected benefits of genomic selection for growth and wood quality traits in *Eucalyptus grandis*. *Tree Genetics & Genome* 16**,** 49.

Mubayiwa, M., Mutasa, W., Gasura, and Mabasa, S. (2016). Grafting and 2.4-dichlorophenoxyacetic acid-induced flowering in non-flowering sweet potato in sub-tropics. *South African Journal of Botany* 106**,** 153-157.

Mudge, K., Janick, J., Scofield, S., and Goldschmidt. E. (2009). "The history of grafting," in *Horticulture reviews*. 449-475.

Myburg, A. A., Hussey, S. G., Wang, J. P., Street, N. R., and Mizrachi, E. (2019). Systems and synthetic biology of forest trees: A bioengineering paradigm for woody biomass feedstocks. *Frontiers in Plant Science* 10**,** 775.

Navarro, C., Abelenda, J. A., Cruz-Oro, E., Cuellar, C. A., Tamaki, S., Silva, J., Shimamoto, K., and Prat, S. (2011). Control of flowering and storage organ formation in potato by *FLOWERING LOCUS T*. *Nature* 478**,** 119-122.

Osorio, L. F., White, T. L., and Huber, D. A. (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 (Article). *Theoretical and Applied Genetics* 106**,** 735-743.

Perkel, J. (2008). SNP genotyping: six technologies that keyed a revolution. *Nature Methods* 5**,** 447-454.

Pineiro, M., and Coupland, G. (1998). The control of flowering time and floral identity in *Arabidopsis*. *Plant Physiology* 117**,** 1-8.

Resende, R. T., Resende, M. D. V., Silva, F. F., Azevero, C. F., Takahashi, E. K., Silva-Junior, O. B., and Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in *Eucalyptus* breeding with an additive-dominant model. *Heredity* 119**,** 245-255.

Righter, F. I. (1939). Early flower production among the Pines. *Journal of Forestry* 37**,** 935-938.

Rodriguez-Ramilo, S. T., Garcia-Cortes, L. A., and Rodriguez De Cara, M. A. (2015). Artificial selection with traditional or genomic relationship: consequences in coancestry and genetic diversity. *Frontiers in Genetics* 6**,** 127.

Silva-Junior, O. B., Faria, D. A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206**,** 1527-1540.

Singh, R. K., Svystun, T., Aldahmash, B., Jonsson, A. M., and Bhalerao, R. (2017). Photoperiod- and temperature-mediated control of phenology in trees – a molecular perspective. *New Phytologist* 213**,** 511-524.

Soares, J. M., Weber, K. C., Qui, W., Stanton, D., Mahmoud, L. M., Wu, H., Huyck, P., Zale, J., Al Jasim, K., Grosser, J. W., and Dutt, M. (2020). The vascular targeted citrus *FLOWERING LOCUS T3* gene promotes noninductive early flowering in transgenic Carrizo rootstocks and grafted juvenile scions. *Nature* 10**,** 21404.

Song, G-Q., Walworth, A. E., and Loescher, W. H. (2015). Grafting of genetically engineered plants. *Journal of American Society of Horticultural Science* 140**,** 203-213.

Song, G-Q., Walworth, A. E., Shoa, D., Jiang, N., and Hancock, J. F. (2013). The *Vaccinium corymbosum FLOWERING LOCUS T*-like gene (VcFT): A flowering activator reverses the photoperiodic and chilling requirements and enables early and continuous flowering in blueberry. *Plant Cell Reports* 32**,** 1759-1769.

Tietel, Z., Srivastava, S., Fait, A., Tel-Zur, N., Carmi, N., and Raveh, E. (2020). Impact of scion/rootstock reciprocal effects on metabolomics of fruit juice and phloem sap in grafted *Citrus reticulata*. *PLOS One* 15**,** e0227192.

Trankner, C., Lehmann, S., Hoenicka, H., Hanke, M-V., Fladung, M., Lenhardt, D., Dunemann, F., Gau, A., Schlangen, K., Malnoy, M., and Flachowsky, H. (2010). Over-expression of an FT-homologous gene of apple induces early flowering in annual and perennial plants. *Planta* 232**,** 1309-1324.

Yoo, S-C., Chen, C., Rojas, M., Daimon, Y., Ham, B-K., Araki, T., and Lucas, W. J. (2013). Phloem long-distance delivery of *FLOWERING LOCUS T (FT)* to the apex. *The Plant Journal* 75**,** 456-468.

Zhang, H., Harry, D. E., Ma, C., Yuceer, C., Hsu, C-Y., Vikram, V., Shevchenko, O., Etherington, E., and Strauss, S. H. (2010). Precocious flowering in trees: the *FLOWERING LOCUS T* gene as a research and breeding tool in *Populus*. *Journal of Experimental Botany* 61**,** 2549-2560.

CHAPTER 5

CONCLUDING REMARKS

5.1 INTRODUCTION

Plants are significant contributors to global permanent carbon biomass in the form of woody biomass (Bar-On et al., 2018), covering over 4 billion hectares, with 0.9 billion hectares of forests and woodlands (Bastin et al., 2019). Over 277 million hectares globally of forest plantations (Carle et al., 2008; Payn et al., 2015), of which over 20 million hectares constitutes *Eucalyptus* plantations (Iglesias-Trabado and Wilstermann, 2009), support many manufacturing industries that produce biomaterials, biochemicals and bioenergy (Malan, 1993; Perlack et al., 2005; Shepherd et al., 2011; Cetinkol et al., 2012; Stanturf et al., 2013; Devappa et al., 2015; Stafford et al., 2020). The sustainable production of woody biomass from plantation forestry is vital to mitigate pressures from a growing human population, decreasing land suitable for plantation forestry, increased demand for wood products and increased pressures from biotic and abiotic stressors (Wingfield et al., 2008; Wingfield et al., 2015) many of which are associated with rapid climate change (Irland et al., 2001; Booth et al., 2013). Faster and more agile breeding approaches are therefore needed to ensure the future sustainability of plantation forestry.

This PhD study aimed to evaluate the practical application of genomic-assisted prediction models contrasting them with traditional breeding (TB) approaches in *E. grandis* in South Africa. I first investigated the efficiency of genomic selection (GS) breeding strategies for growth and wood quality traits maximising genetic gains per unit breeding cycle time. I then used multivariate single-step GBLUP to investigate the simultaneous improvement of diameter growth and tolerance to *Leptocybe invasa* leaf gall wasp (*Lepto*) and the co-occurring fungal stem disease complex of *Botryosphaeria dothidea* and *Teratosphaeria zuluensis* (*BotryoTera*). This resulted in a proposed GS breeding strategy that mitigated some of the inherent difficulties

associated with phenotyping pest and pathogen incidences in field trials. The proposal includes more accurate phenotyping of pest and pathogen tolerance under controlled conditions (greenhouse or tunnel) to train and develop a multi-trait GS model for the simultaneous improvement of diameter growth and *Lepto* and *BotryoTera* tolerance. Last, I explored the feasibility of combining GS with accelerated flowering techniques such as floral induction via transgenic expression of the *Flowering locus T* (*FT*) gene followed by graft transmission (transgrafting) of the FT protein to non-transgenic scion or rootstock plants in precocious and non-precocious plantation *Eucalyptus* species. In these Concluding Remarks, I consolidate the main findings of the PhD study and relate these findings to current and future research opportunities in genome-assisted tree breeding models. I further discuss the practicality of integrating genome-assisted breeding strategies for operational *E. grandis* tree improvement programmes.

5.2 RELEVANCE TO CURRENT DEVELOPMENTS IN THE FIELD

Since the simulation study by Grattapaglia et al. (2011) of the most influential factors for GS in forest tree breeding, the list of traits included in GS studies has grown in part due to new phenotyping technologies (Finkel, 2009). These high-throughput techniques and methods deal with complex traits in breeding programmes leveraging multi-environment, multi-age, multi-trait, and even multi-omics information on the tree and its associated microbes (Cabrera-Bosquet et al., 2012; de Silva Fonseca et al., 2018; Juliana et al., 2019; Alves et al., 2020; Dastogeer et al., 2020; Moreira et al., 2020; Tong and Nikoloski, 2020). Grattapaglia et al. (2018) highlighted further factors needed to accelerate the adoption of GS in forest tree breeding such as cost-effective genotyping, model functional validation, model updating, multi-environmental forecasting, large-scale (internationally based) training populations and advanced computational models. These additional factors increase the complexity of

implementing GS in forest trees, especially when expensive phenomics and breeding objectives have to be coordinated in operational breeding programmes with limited resources.

Reproductive (flowering) age in *E. grandis* is one of the barriers that constrain GS breeding cycles to a minimum of four years. de Oliverira Castro et al. (2021) recently demonstrated an early floral induction of 6-months-old *Eucalyptus* hybrid seedlings, top-grafted on mature rootstocks treated with paclobutrazol as a possible solution to accelerated GS breeding cycles. Transgenic over-expression of the *FT* gene has been demonstrated as a potent early floral induction strategy in *Eucalyptus* (Klocko et al., 2016). The Forestry Stewardship Council (FSC) regulations restrict commercial testing and the use of transgenic material, but there is pressure to ease such restrictions (Strauss et al., 2019). To overcome the negative public perception of genetically modified trees and the FSC's commercial testing and use restrictions, a possible solution would be to use graft transmission (transgrafting) of the FT protein from transgenic rootstocks to non-transgenic scions (Song et al., 2015). This technology has been demonstrated to be feasible in plants such as blueberries (Song et al., 2019) and, recently, citrus (Soares et al., 2020), but has not been achieved in forest trees such as *Populus* (Zhang et al., 2010) and *Eucalyptus*. An alternative approach, whereby FT protein is transgrafted from transgenic scion to non-transgenic rootstock would be more practical and feasible in *Eucalyptus*, because it would be much easier to manage large numbers of non-transgenic genotypes as rooted plants (already in nurseries), rather than as harvested scion material. Such an approach would also allow for additional floral induction mechanisms such as vernalization and chemical induction (Gardner and Bertling, 2005; Ionescu et al., 2017) on the selected rootstocks complementing the FT induction as proposed in the accelerated GS-FT breeding strategy. If successful, the GS-FT breeding strategy will result in even greater genetic gains and benefit to cost ratios in non-precocious (late or shy flowering) *Eucalyptus* species because

of the greater breeding cycle time reduction in such species. A more comprehensive study is required to determine the real costs of implementing a GS-FT breeding strategy taking into consideration regulatory and biosafety measures and the type of facilities that will be required for large-scale production of GS-FT selected seedlings in an operational breeding programme.

The findings of my study are consistent with many other ongoing and completed GS studies in forest tree species over the past decade, in softwoods (Resende Jr et al., 2012; Zapata-Valenzuela et al., 2013; Isik et al., 2015; Bartholome et al., 2016; EL-Dien et al., 2016) and hardwoods (Zelener et al., 2005; Denis et al., 2013; Duran et al., 2017; Rambolarimanana et al., 2018; Tan et al., 2018; Suontama et al., 2019). These studies include advances such as the accurate dissection of additive and non-additive genetic components using GS models in tree species (Thavamanikumar et al., 2013; Luo et al., 2014; Cappa et al., 2019; de Lima et al., 2019). The benefits of GS are highest when phenotypic selection accuracy is poor, genetic variation is abundant and ranking is vital to decision making in breeding operations (Pegard et al., 2020).

5.3 APPLIED GENOMIC TREE BREEDING AND FUTURE PROSPECTS

Improving multiple traits simultaneously is another challenge in many breeding programmes. The use of GS index models has been shown to outperform univariate GS models to improve multiple traits simultaneously (Silva et al., 2021). Such GS index approaches can be used to simultaneously improve growth and wood quality traits as well as the growth and pest and disease tolerance to achieve selection efficiency and reasonable genetic gains across all of the traits. In the animal industry, decision-based GS models are implemented on a weekly basis in all areas of breeding such as which individuals to inseminate, which to use as embryo donors and which to sell or keep (VanRaden, 2019). Bayesian decision theory can be integrated into conventional GS strategies that consider genetic parameters, selection responses and the

genetic correlation of traits, thereby removing the need for breeders to consider individual GEBVs to make parental selections (Villar-Hernandez et al., 2021).

A unique approach of this PhD study is the simulation of GS breeding strategies in comparison to an active TB strategy in an operational *E. grandis* breeding programme. I show that a 4-year GS breeding cycle relative to the traditional 8-year breeding cycle is optimal for *E. grandis* to achieve good selection efficiencies for growth and wood quality traits. The 4-year GS breeding strategy is defined by the biological limits inherent to the *E. grandis* reproductive process. Such limitations vary across tree species and will dictate the factors to consider in deploying a GS breeding strategy. As shown by Lebedev et al. (2020), the economic viability of GS models need to be assessed with regards to the market value of the traits involved. GS approaches can positively influence the forestry value chain including the land expected value and projected forest value estimations (Chamberland et al., 2020). I show in this PhD study that practical adjustments to the breeding operations for *E. grandis* are needed to realise and empirically demonstrate the benefits of GS strategies. Implementing GS strategies is challenging particularly in industry because it involves convincing non-breeders (managers) to invest more resources into genotyping, phenotyping, data storage, data analysis and the development of training populations (Isik, 2014). Due to the long rotation cycles and the complex nature of the forestry business, the return on investment, particularly in a vertically integrated company, will only be realised when the trees are processed through the mills.

Even with the many advancements of simulation studies providing proof of concept and studies that generated empirical evidence for the viability of GS in forest trees, it remains a step-change for companies to adopt. To date, we have a reliable and relatively affordable SNP genotyping

platform (Silva-Junior et al., 2015). The cost is gradually declining with new higher-throughput SNP genotyping and DNA sequencing platforms coming to market. Data storage capacity is currently not limiting, but there is a scarcity of bioinformatics skills needed to access and analyse such data in industry. In addition, establishing suitable training populations is a critical challenge, in particular for companies that already have extensive breeding operations, sometimes for multiple species to deliver commercial outputs within limited resources. Establishing well-replicated training populations with relevant genotyped treatments and the subsequent phenotyping (from nursery to rotation age across all relevant trait interactions) will cause significant operational disruptions to current conventional breeding operations and will increase the operational breeding budget. Once GS model training is achieved, the company would need to review and resource the GS breeding operations with new infrastructure and skills. Ultimately, the adoption of GS within forestry companies will come with a step-change in the business model of their breeding programmes because the baseline cost will increase regardless of the reduction in genotyping cost. Current TB strategies rely on natural reproductive processes allowing breeders to focus on managing operational breeding steps, whereas GS and accelerated GS-FT breeding strategies would require that breeders manage both operational breeding steps and the reproductive processes within defined breeding cycle timeframes to realise the suggested gains per unit time.

There is currently still insufficient training data to support deep learning approaches that can extract nonlinear patterns and relationships of all relevant biological information to predict genetic merit and inform breeding decisions (Montesinos-Lopez et al., 2021). It is because of these gaps in relevant data and biological complexities that we can expect more simulations and empirical studies of GS in the near future to demonstrate the potential of GS in artificial intelligence-driven breeding strategies.

5.4 REFERENCES

Alves, F. V., Balmant, K. M., Resende Jr, M. F. R., Kirst, M., and De Los Campos, G. (2020). Accelerating forest tree breeding by integrating genomic selection and greenhouse phenotyping. *The Plant Genome* 13**,** e20048.

Bar-On, Y. M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. *PNAS* 115**,** 6506-6511.

Bastin, J.-F., Finegold, Y., Garcia, C., Mollicone, D., Rezende, M., Routh, D., Zohner, C. M., and Crowther, T. W. (2019). The global tree restoration potential. *Science* 365**,** 76-79.

Bartholome, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., and Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. *BCM Genomics* 17**,** 604.

Booth, T. H. (2013). Eucalypt plantations and climate change. *Forest Ecology and Management* 301**,** 28-34.

Cabrera-Bosquet, L., Crossa, J., Von Zitzewitz, J., Sorret, M. D., and Araus, J. L. (2012). High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *Journal of Integrative Plant Biology* 54**,** 312-320.

Cappa, E. P., De Lima, B. M., Da Silva-Junior, O. B., Carcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in *Eucalyptus* using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Science* 284**,** 9-15.

Carle, J., And Holmgren, P. (2008). Wood from planted forests a global outlook 2005 - 2030. *Forest Product Journal* 58**,** 6-18.

Cetinkol, O. P., Smith-Moritz, A. M., Cheng, G., Lao, J., George, A., Hong, K., Henry, R., Simmons, B. A., Healzwood, J. L., and Holmes, B. M. (2012). Structural and chemical characterization of hardwood from tree species with applications as bioenergy feedstocks. *PLOS One* 7**,** e52820.

Chamberland, V., Robichaud, F., Perron, M., Gelinas, N., Bousquet, J., and Beaulieu, J. (2020). Conventional versus genomic selection for white spruce improvement: a comparison of costs and benefits of plantations on Quebec public lands. *Tree Genetics & Genome* 16**,** 17.

Dastogeer, K. M. G., Tumpa, F. H., Sultana, A., Akter, M. A., and Chakraborty, A. (2020). Plant microbiome an account of the factors that shape community composition and diversity. *Current Plant Biology* 23**,** 100161.

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

de Oliverira Castro, C. A., Dos Santos, G. A., Takahashi, E. K., Nunes, A. C. P., Souza, G. A., and De Resende, M. D. V. (2021). Accelerating Eucalyptus breeding strategies through top grafting applied to young seedlings. Industrial Crops & Products 171, 113906.

de Silva Fonseca, E., Peixoto, R. S., Rosado, A. S., De Carvaldo Baliero, F., Tiedje, J. M., and De Costa Rachid, C. T. C. (2018). The microbiome of *Eucalyptus* roots under different management conditions and its potential for biological nitrogen fixation *Microbial Ecology* 75**,** 183-191.

Denis, M., and Bouvet, J-M. (2013). Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. *Tree Genetics & Genomes* 9**,** 37-51.

Devappa, R. K., Rakshit, S. K., and Dekker, R. F. H. (2015). Forest biorefinery: Potential of poplar phytochemicals as value-added co-products. *Biotechnology Advances* 33**,** 681-716.

Duran, R., Isik, F., Zapara-Valenzuela, J., Balocchi, C., and Valanzuela, A. (2017). Genomic predictions of breeding values in a cloned *Eucalyptus globulus* population in Chile. *Tree Genetics & Genome* 13.

El-Dien, O., G., Ratcliffe, B., Klapste, J., Porth, I., Chen, C., and El-Kassaby, Y. A. (2016). Implementation of the realized genomic relationship matrix to open-pollinated White Spruce family testing for disentangling additive from non-additive genetic effects. *Genes Genomics Genetics***,** 743-753.

Finkel, E. (2009). With 'Phenomics,' plant scientists hope to shift breeding into overdrive. *Science* 325**,** 380-381.

Gardner, R. A. W., and Bertling, I. (2005). Effect of winter chilling and paclobutrazol on floral bud production in *Eucalyptus nitens*. *South African Journal of Botany* 71**,** 238-249.

Grattapaglia, D., Macos, D., and Resende, D.V. (2011). Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7**,** 241-255.

Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Muller, B. S. F., Tan, B., Isik, F., Ratcliffe, D., and El-Kassaby, Y. A. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science* 9**,** 1693.

Iglesias-Trabado, G., and Wilstermann, D. (2009). *Global cultivated eucalypt forests map 2009* [Online]. http://git-forestry-blog.blogspot.com/2008/09/eucalyptus-global-map-2008-cultivated.html. [Accessed].

Ionescu, I. A., Moller, B. L., and Sanchez-Perez, R. (2017). Chemical control of flowering time. *Journal of Experimental Botany* 68**,** 369-382.

Irland, L. C., Adams, D., Alig, R., Betz, C.J., Chen, C-C., Hutchins, M., Mccarl, B.A., Skog, K., and Sohngen, B.L. (2001). Assessing socioeconomic impacts of climate change on US forests, wood-product markets, and forest recreation: The effects of climate change on forests will trigger market adaptations in forest management and in wood-products industries and may well have significant effects on forest-based outdoor recreation. *BioScience* 51**,** 753-764.

Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests* 45**,** 379-401.

Isik, F., Bartholome, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C., and Bouffier, L. (2015). Genomic selection in maratime pine. *Plant Science* 242**,** 108-119.

Juliana, P., Montesinos-Lopez, O. A., Crossa, J., Mondal, S., Perez, L. G., Poland, J., Huerta-Espino, J., Crespo-Herrera, L., Govindan, V., Dreisigacker, S., Shrestha, S., Perez-Rodriguez, P., Espinosa, F. P., and Singh, R. P. (2019). Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theoretical and Applied Genetics* 132**,** 177-194.

Klocko, A. L., Ma, C., Robertson, S., Esfandiari, E., Nilsson, O., and Strauss, S. H. (2016). *FT* overexpression induces precocious flowering and normal reproductive development in *Eucalyptus*. *Plant Biotechnology Journal* 14**,** 808-819.

Lebedev, V. G., Lebedeva, T. N., Chernodubov, A. I., and Shestibratov, K. A. (2020). Genomic selection for forest tree improvement: methods, achievements and perspectives. *Forests* 11**,** 1190.

Luo, J., Arnold, R., Lu, W., and Lin, Y. (2014). Genetic variation in *Eucalyptus camaldulensis* and *E. tereticornis* for early growth and susceptibility to the gall wasp *Leptocybe invasa* in China. *Euphytica* 196**,** 397-411.

Malan, F. S. (1993). The wood properties and qualities of three South African grown eucalypt hybrids. *South African Forestry Journal* 167**,** 35-44.

Montesinos-Lopez, O. A., Montesinos-Lopez, A., Perez-Rodriguez, P., Barron-Lopex, J. A., Martini, J. W. R., Fajardo-Flores, S. B., Gaytan-Lugo, L. S., Santana-Mancilla, P. C., and Crossa, J. (2021). A review of deep learning applications for genomic selection. *BCM Genomics* 22**,** 19.

Moreira, F. F., Oliveira, H. R., Volenec, J. J., Rainey, K. M., and Brito, L. F. (2020). Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. *Frontiers in Plant Science* 26**,** 681.

Payn, T., Carnus, J-M., Freer-Smith, P., Kimberley, M., Kollert, W., Liu, S., Orazio, C., Rodriguez, L., Silva, L. N., and Wingfield, M. J. (2015). Changes in planted forests and future global implications. *Forest Ecology and Management* 352**,** 57-67.

Pegard, M., Segura, V., Munoz, F., Bastien, C., Jorge, V., Sanchez, L. (2020). Favorable conditions for genomic evaluation to outperform classical pedigree evaluation highlighted by a proof-of-concept study in poplar. *Frontiers in Plant Science* 11**,** 581954.

Perlack, R. D., Turhollow, W. L .L., Af, G. R. L., Stokes, B. J., Erbach, D. C. (2005). "Biomass as a feedstock for a bioenergy and bioproduct industry: the technical feasibility of a billion-ton annual supply", (ed.) *O.R.N. Laboratory*. (USA: Department of Energy).

Rambolarimanana, T., Ramamonjisoa, L., Verhaegen, D., Tsy, J-M. L. P., Jacquin, L., Cao-Hakouanzi, G., and Bouvet, J-M. (2018). Performance of multi-trait genomic selection for *Eucalyptus robusta* breeding program. *Tree Genetics & Genome* 14**,** 71.

Resende Jr, M. F. R., Munoz, P., Resende, M.D.V., Garrick, D.J., Fernando, R.L., Davis, J.M., Jokela, E.J., Martin, T.A., Peter, G.F., and Kirst, M. (2012). Accuracy of genomic selection methods in standard data set of loblolly pine (*Pinus taeda* L.). *Genetics* 190**,** 1503-1510.

Shepherd, M., Bartle, J., Lee, D, J., Brawner, J., Bush, D., Turnbull, P., Macdonel, P., Simmons, B., and Henry, R. (2011). Eucalypts as a biofuel feedstock. *Biofuels* 2**,** 639-657.

Silva-Junior, O.B., Faria, D.A., and Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist* 206**,** 1527-1540.

Silva, L. A., Peixoto, M. A., Peixoto, L. A., Romero, J. V., and Bhering, L. L. (2021). Multi-trait genomic selection indexes applied to identification of superior genotypes. *Bragantia* 80**,** e3621.

Soares, J. M., Weber, K. C., Qui, W., Stanton, D., Mahmoud, L. M., Wu, H., Huyck, P., Zale, J., Al Jasim, K., Grosser, J. W., and Dutt, M. (2020). The vascular targeted citrus *FLOWERING LOCUS T3* gene promotes noninductive early flowering in transgenic Carrizo rootstocks and grafted juvenile scions. *Nature* 10**,** 21404.

Song, G.-Q., Walworth, A. E., and Loescher, W. H. (2015). Grafting of genetically engineered plants. *Journal of American Society of Horticultural Science* 140**,** 203-213.

Song, G.-Q., Walworth, A., Lin, T., Chen, Q., Han, X., Zaharia, L. I., and Zhong, G-Y. (2019). *VcFT*-induced mobile florigenic signals in transgenic and transgrafted blueberries. *Horticulture Research* 6**,** 109.

Stafford, W., De Lange, W., Nahman, A., Chunilall, V., Lekha, P., Andrew, J., Johakimu, J., Sithole, B., Trotter, D. (2020). Forestry biorefineries. *Renewable Energy* 154**,** 461-475.

Stanturf, J. A., Vance, E. D., Fox, T. R., and Kirst, M. (2013). *Eucalyptus* beyond its native range: Environmental issues in exotic bioenergy plantations. *International Journal of Forestry Research* 2013.

Strauss, S. H., Boerjan, W., Chiang, V., Costanza, A., Coleman, H., Davis, J. M., Lu, M-Z., Mansfield, S. D., Merkle, S., Myburg, A. A., Nilsson, O., Pilate, G., Powell, W., Seguin, A., and Valenzuela, S. (2019). Certification for gene-edited forests. *Science* 365**,** 767-768.

Suontama, M., Klápště, J., Telfer, E., Graham, N., Stovold, T., Low, C., Mckinley, R., and Dungey, H. (2019). Efficiency of genomic prediction across two *Eucalyptus nitens* seed orchards with different selection histories. *Heredity* 122**,** 370-379.

Tan, B., Grattapaglia, D., Wu, H. X., and Ingvarsson, P. K. (2018). Genomic relationships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Science* 267**,** 84-93.

Thavamanikumar, S., Southerton, S.G., Bossinger, G., and Thumma, B.R. (2013). Dissection of complex traits in forest tree: opportunities for marker-assisted selection. *Tree Genetics & Genomes* 9**,** 627-639.

Tong, H., Kuken, A., and Nikoloski, Z. (2020). Integrating molecular markers into metabolic models improves genomic selection for Arabidopsis growth. *Nature Communication* 11**,** 2410.

VanRaden, P. M. (2019). "Symposium review: How to implement genomic selection", in *Joint ADSA/Interbull session: Breeding and Genetics: Ten years of genomic selection.* (ADSA annual meeting, Cincinnati, Ohio).

Villar-Hernandez, B.J., Perez-Elizalde, S., Martini, J. W. R., Toledo, F., Perez-Rodriguez, P., Krause, M., Garcia-Calvillo, I. D., Covarrubias-Pazaran, G., Crossa, J. (2021). Application of multi-trait Bayesian decision theory for parental genomic selection. *Genes Genomics Genetics* 11**,** 2.

Wingfield, M. J., Brocherhoff, E.G., Wingfield, B.D., and Slippers, B. (2015). Planted forest health: The need for a global strategy. *Science* 349**,** 832-836.

Wingfield, M. J., Slippers, B., Hurley, B. P., Coutinho, T. A., Wingfield, B. D., and Roux, J. (2008). Eucalypt pest and diseases: Growing threats to plantation productivity. *Southern Forests* 70**,** 139-144.

Zapata-Valenzuela, J., Whetten, R. W., Naele, D., Mckeand, S., and Isik, F. (2013). Genomic estimated breeding values using genomic relationship matrices in a cloned population of Loblolly Pine. *Genes Genomics Genetics* 3**,** 909-916.

Zelener, N., Poltri, S. N. M., Bartoloni, N., Lopez, C. R., and Hopp, H. E. (2005). Selection strategy for seedling seed orchard design based on trait selection index and genomic analysis by molecular markers: a case study for *Eucalyptus dunnii*. *Tree Physiology* 25**,** 1457-1467.

Zhang, H., Harry, D. E., Ma, C., Yuceer, C., Hsu, C-Y., Vikram, V., Shevchenko, O., Etherington, E., and Strauss, S. H. (2010). Precocious flowering in trees: the *FLOWERING LOCUS T* gene as a research and breeding tool in *Populus*. *Journal of Experimental Botany* 61**,** 2549-2560.