University of São Paulo
"Luiz de Queiroz" College of Agriculture

Bridging genomics and quantitative genetics of *Eucalyptus*: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data

Bruno Marco de Lima

Thesis presented to obtain the degree of Doctor in Science. Program: Plant Genetics and Breeding

Piracicaba
2014

Bruno Marco de Lima
Forestry Engineer

Bridging genomics and quantitative genetics of *Eucalyptus*: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr.**ROLAND VENCOVSKY**

Co-advisor:
Prof. Dr. **DARIO GRATTAPAGLIA**

Thesis presented to obtain the degree of Doctor in Science. Program: Plant Genetics and Breeding

Piracicaba
2014

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Lima, Bruno Marco de
Bridging genomics and quantitative genetics of Eucalyptus: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data / Bruno Marco de Lima. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2014.
92 p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.

1. Melhoramento florestal 2. Seleção genômica 3. Marcador molecular 4. Herdabilidade I. Título

CDD 634.9734
L732b

This thesis is dedicated to my parents and family, for all the support during the last years of study and absence. They, even in harshtimes, always let me clear that I did the right choice (PhD). Sorry mom for not calling you in the phone, as frequently I should!

I offer the thesis to Giselle, my partner and love, who decided to spend her life by my side

A**CKNOWLEDGEMENTS**

I would like to thank the Escola Superior de Agricultura “Luiz de Queiroz” (“Luiz de Queiroz” College of Agriculture), in the University of São Paulo, and the Department of Genetics for all the structure and support, essential for my formation, since my undergraduate in Forestry Engineering until my PhD.

I thank the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; São Paulo Research Foundation) for my grant during this period, allowing me to develop my project internationally.

I thank International Paper Brasil for all the support, providing the *Eucalyptus* population used in this study. I really appreciate your help every time I needed: Carla Garcia, Adriano Almeira, Cristiane Lemos and Sebastião Oliveira. Also all the guys who helped me in the sampling procedures, thank you!

I am grateful to Prof. Dr. Roland Vencovsky and Prof. Dr. Dario Grattapaglia.

Professor Vencovsky was always more than a supervisor, he was a friend. Always helping me out, providing deep explanations on quantitative genetics and statistical analysis. I will always remember the classes you gave us in the lab and the questions you made me, about things I had never thought about (and I still don’t know the answers for some!). I am really proud to say to everyone that the greatest brazilian quantitative geneticist was my PhD supervisor.

Professor Grattapaglia or just Dario was also a very important person in my life. I have to say that if I am a tree breeder today, he is the responsible. Since my Master’sdegree, he always encouraged me to do my PhD in Brazil, in order to develop the *Eucalyptus* breeding in the country. I followed his advices and I must say that I have no regrets. Even by distance, we always developed the project together and he was always giving me the guidelines to achieve the results we were looking for. Forest genetics needs more people like you.

I also thank Prof. Dr. Shawn Mansfield. A great person, my host at the University of British Columbia (Vancouver, BC, Canada). He was always worried if I was happy and if my studies were doing well. He made my time in Canada so pleasant, showing me how great the Canadians are. Shawn, this thesis is yours too. Everything I know about wood analysis I owe you.

I thank my lab mates in Brazil. Fernando Toledo, Evandro Tambarussi and Bruna de Oliveira are big friends. We shared great moments and laughs. Also, this so diverse lab taught me a lot, from R scripts to breeding experiments analysis, passing through the use of molecular markers. Thank you guys and count on me every time you need.

Thank you, Danielle Assis Paiva, for the friendship during these years, coming and going to Brasília. It was great to meet you and I am sure we are still going to talk a lot about *Eucalyptus*.

I thank the folks from Embrapa-Cenargen lab, who helped me a lot during parts of my project done there. Special thanks to Dione, Rodrigo and Ediene. Also thanks to Maristela, for the friendship and for always hosting me at her apartment.

I thank Prof. Dr. Marcos Deon Resende (UFV), who has helped a lot with some complex GWS questions. Also thanks to Prof. Dr. Alexandre Coelho (UFG), for helping me during the wood predictions.

A special thanks to all my lab mates during my time in Vancouver. People, you are awesome. I will try a small list: Foster Hart, LiXi, Amanda Johnson, Sofyia Lazarova, Victoria Maloney, Sarah Barkwill, Ahmed Kenawy, Eliana Gonzales, Melissa Roach, Letitia Da Ros, Ilga Porth, Kent Chen, Yaseen Mottiar, Danyu Yao. Thank you for the friendship and for teach me so much Faride Unda, Grant McNair, Francis de Araujo, Charles Hefer and Alex Skyba.

I thank Professor Matias Kirst and Professor Patricio Muñoz from University of Florida and his students Leandro Neves and Marcio Resende Jr. for helping me in the genomic selection analysis.

I thank my roommates in Brazil: Fernando Piotto (Juds), Danilo Mizuta (Chuck), Gabriel Margarido (Palms), André Signoretti (Sinhá), Felipe Terra (Zélão), Fernado Egadi (Kju). And in Vancouver: Stéphanie Beauseigle (CAN), Barend Lotter (RSA), Ives (BRA), Claudio (BRA), Camila (BRA), Lucas (BRA), Marcia (BRA), João Fábio (BRA), Dawn (CAN), Tim (AUS), Tim (NED), Jean-Simmon (CAN), Sarah (MLT). You all made the last 4 years very happy and complete.

Finally, I thank my family. My parents (Cleide and Devair) for all the support the whole time, my brother (Kleber), sister (Carla), sister-in-law (Flavia), niece (Clara, also goddaughter), aunt (Claudia), uncle (Joilson), cousins (João and Juliana), grandma (Cida) for all the visits in Piracicaba when I was not able to go seeing you!

And thank you Giselle, for being by my side for the last years. You helped me a lot, sometimes just by staying with me. I love you and life is just starting for us!

“Man does not actually produce variability; he only unintentionally exposes organic beings to new conditions of life, and then nature acts on the organisation, and causes variability. But man can and does select the variations given to him by nature, and thus accumulate them in any desired manner. He thus adapts animals and plants for his own benefit or pleasure. He may do this methodically, or he may do it unconsciously by preserving the individuals most useful to him at the time, without any thought of altering the breed”

The origin of species
by **Charles Darwin**

SUMMARY

RESUMO

Conectando a genômica à genética quantitativa de*Eucalyptus*: predição genômica e estimação de parâmetros genéticos para crescimento e propriedades de madeira usando alta densidade de SNPs

A convergência da genética quantitativacom a genômica está se tornando a maneira pela qual a genética fundamental e aplicada serão conduzidas nas próximas décadas. Este estudo buscou conectar a genética defenótipos complexos de crescimento e propriedades de madeira às tecnologias genômicas, em uma abordagem inovadora para o melhoramento florestal. Florestas plantadas têm papel fundamental para satisfazer a crescente demanda mundial por produtos madeireiros e energia. O eucalipto,com sua alta produtividade e madeira versátil, é resultado de programas avançados de melhoramento associados à propagação clonal e silvicultura moderna. Apesar de seu rápido crescimento, ciclos de melhoramento ainda levam muitos anos e a avaliação detalhada de propriedades da madeira é limitada a apenas uma amostra das árvores em estágios avançados de seleção, devido aos altos custos de fenotipagem, não explorando assim toda a variação genética disponível. Neste estudo, examinamos quinze caracteres, incluindo crescimento e propriedades químicas e físicas da madeira, em 1000 indivíduos amostrados de uma população elite de melhoramento. Modelos de espectroscopia de infravermelho próximo (NIRS) foram desenvolvidos e utilizados para fenotipagem de altodesempenho de propriedades de madeira. Genotipagem de alta densidade com 29.090 SNPs foi utilizada para obter estimativas acuradas de componentes de variância, herdabilidades e correlações genéticas baseadas em uma matriz de parentesco realizado, ou seja,sem o uso de pedigree. Este é o primeiro estudo de que temos conhecimento a fazer isso em plantas. Predições NIRS foram precisas para caracteres químicos da madeira e densidade, e apresentaram sucesso variável para caracteres físicos. As herdabilidades foram médias para crescimento (0,34 a 0,44), altas para caracteres químicos de madeira (0,56 a 0,85) e variáveis para caracteres físicos da madeira (0,11 a 0,63). Altas correlações positivas entre caracteres de crescimento e negativas entre celulose e lignina foram observadas, enquanto correlações entre caracteres químicos e físicos da madeira foram baixas, porém significativas. Fenótipos e marcadores SNP foram em seguida utilizados na construção de modelos preditivos com a maior densidade de marcadores já utilizadaem estudos de seleção genômica em espécies florestais (1 SNP/21 kpb). Dois modelos de predição (RR-BLUP e LASSO Bayesiano)foram usados nas predições genômicas e comparados ao BLUP fenotípico. Os modelos apresentaram capacidades preditivas similares, fortemente correlacionadas às herdabilidades. Predições genômicas precisas foram obtidas para caracteres relacionados à lignina, densidade e crescimento, embora geralmente 15 a 25% menores do que as predições obtidas por BLUP fenotípico. Contudo, predições genômicas alcançaram coincidências acima de 70% na seleção das melhores 30 árvores ranqueadas pela seleção fenotípica para crescimento, densidade e relação S:G, e de 60% quando seleção em tandem foi aplicada. Os resultados deste estudo abrem enormes oportunidades para o uso combinado de fenotipagem NIRS e genotipagem com SNPs no melhoramento do eucalipto, permitindo estimativasacuradas de parâmetros genéticos e a predição de valores genéticos genômicos para plantas jovens ainda não fenotipadas. Estas aplicações deverão se tornar rotineiras nos programas de melhoramento florestal nos próximos anos, reduzindo significativamente a duração dos ciclos de seleção e, consequentemente, otimizando a alocação de recursos e a sustentabilidade do melhoramento.

Palavras-chave: Melhoramento florestal; Seleção genômica; Marcador molecular; Herdabilidade

ABSTRACT

Bridging genomics and quantitative genetics of *Eucalyptus*: genome-wide prediction and genetic parameter estimation for growth and wood properties using high-density SNP data

Convergence of quantitative genetics and genomics is becoming the way that fundamental genetics and applied breeding will be carried out in the next decades. This study bridges the quantitative genetics of complex growth and wood properties traits with genomic technologies towards a more innovative approach to tree breeding. Planted forests play a major role to fulfill the growing world demand for wood products and energy. Eucalypts stand out for their high productivity and versatile wood resulting from the advanced breeding programs associated to clonal propagation and modern silviculture. Despite their fast growth, breeding cycles still take several years and wood properties assessment is limited to a sample of trees in the late stages of selection due to the costs involved in wood phenotyping, not exploitingthe range of genetic variation in wood properties. In this study, we examined fifteen traits including growth and wood chemical and physical properties in 1,000 individuals sampled from an elite *Eucalyptus* breeding population. Near-infrared spectroscopy (NIRS) models were developed and used for high-throughput phenotyping of wood traits.High-density data for 29,090 SNPs was used to obtain accurate pedigree-record-free estimates of trait variance components, heritabilities, genetic and phenotypic correlations, based on a realized relationship matrix, comparing them to pedigree-based estimates. To the best of our knowledge, this is the first study to do this in plants. NIRS predictions were accurate for wood chemical traits and wood density, and variably successful for physical traits. Heritabilities were medium for growth (0.34 to 0.44), high for wood chemical traits (0.56 to 0.85) and variable for wood physical traits (0.11 to 0.63). High positive correlations among growth traits and negative between cellulose and lignin content were observed, while correlations between wood chemical and physical traits and between growth and wood quality traits were low although significant. Phenotypes and SNP markers were then used to build genomic predictive models using a marker density higher than any previous genomic selection study in trees (1 SNP/21 kbp). Two models (RR-BLUP and Bayesian LASSO) that differ regarding the assumed distribution of marker effects were used for genomic predictions. Predictions were compared to those obtained by phenotypic BLUP. Predictive abilities very similar by the two models and strongly correlated to the heritabilities. Accurate genomic-enabled predictions were obtained for wood chemical traits related to lignin, wood density and growth, although generally 15 to 25% lower than those achieved by phenotypic BLUP prediction. Nevertheless, genomic predictions yielded a coincidence above 70% in selecting the top 30 trees ranked by phenotypic selection for growth, wood density and S:G ratio, and 60% when tandem selection was applied. The results of this study open opportunities for an increased use of high-throughput NIRS phenotyping and genome-wide SNP genotyping in *Eucalyptus* breeding, allowing accurate pedigree-record-free estimation of genetic parameters and prediction of genomic breeding values for yet to be phenotyped trees. These applications should become routine in tree breeding programs for the years to come, significantly reducing the length of breeding cycles while optimizing resource allocation and sustainability of the breeding endeavor.

Keywords: Tree breeding; Genomic selection; Molecular marker; Heritability

1 INTRODUCTION

Wood products are present everywhere. The range of products and functions derived from wood is vast, from the most application to structures, furniture and complex biomaterials, to its use for energy, pulp and paper. To satisfy the needs of a growing population, planted forests play an essential role, alleviating human pressure on the existing, frequently weakened, natural forests in the world. However, most of the tree species have particular growth habits and, even planted extensively, would not be adequate to sustain the existing demands for wood products. The fact is that just a dozen tree species are effectively planted worldwide. Eucalypts stand out as one of the most productive planted trees, especially in tropical and subtropical areas, with productivities exceeding 60 $m^3.ha^{-1}.year^{-1}$ in some regions of Brazil, for example (REZENDE; RESENDE; ASSIS, 2014). In our country, eucalypts, which include species from the genus *Eucalyptus* and *Corymbia*, represent 76.6% of the planted forests, with 5.1 million hectares established (ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS - ABRAF, 2013), used mainly by the pulp and paper industry and for charcoal and firewood.

Native to Australia, Papua New Guinea and Indonesia with more than 700 described species (LADIGES; UDOVICIC; NELSON, 2003), eucalypts found in Brazil excellent conditions to reach their potential as fast growing species (ELDRIDGE et al., 1993). In order to achieve such high productivities, *Eucalyptus* has been extensively studied in the last decades. Silvicultural improvements and the clonal plantations, associated to competent breeding programs, were responsible for most of the gains. However, there is still room for improvement. Eucalypt breeding programs have focused almost exclusively on growth traits and adaptation, including disease resistance, frost and drought tolerance. Due to the versatile uses of eucalypt wood, however, wood properties have called the attention of tree breeders in recent years, and are now increasingly being included in the selection procedures. In addition, breeding cycles still take several years even for eucalypts. The selection of an elite clone is hardly completed in less than 12 to 18 years and efforts have been made to shorten these times by early selection.

Wood phenotyping is a challenging procedure. In eucalypt breeding the characterization of wood properties has to be done when trees are at least 3-years-old, in order to achieve adequate correlations with the wood at harvest or rotation age. Moreover, most measurement procedures are destructive and carried out in the late stages of selection, with a greater focus on wood density in most cases. Wood density is, in fact, a key trait, considerably

correlated to other important traits such as pulp yield and calorific power from wood, reason why this trait has been an important target of eucalypt breeding. Indirect measurement of wood properties by near-infrared spectroscopy (NIRS) has been used as a way to speed and reduce the cost of wood phenotyping (SCHIMLECK et al., 2000). This technique uses spectral data collected from NIRS readings of wood samples, which are also analyzed by wet chemistry procedures. These spectra together with the actual chemical measurements are then used to generate calibration curves using multivariate models, which in turn are subsequently used to predict the same wood chemical compounds in samples for which only spectral reads are taken. In eucalypt phenotyping, the procedure has been efficient and reliable to predict cellulose, lignin content and composition (syringyl to guaiacyl lignin ratio - S:G ratio), extractives, density and pulp yield (RAYMOND; SCHIMLECK, 2002; COSTA E SILVA et al., 2009; STACKPOLE et al., 2011).

Reliable wood characterization is just one of the current bottlenecks for the incorporation of wood properties assessment in eucalypt breeding. Genetic parameters estimation and genotypic correlations amongst such wood traits and growth is an essential component of the success of a breeding program. Knowledge of the genetic parameters such as heritability and genotypic correlation amongst traits usually leads to more accurate responses to selection and helps the breeder in the decision-making process along the breeding program. Estimation of such parameters has been performed based on pedigree information mainly in temperate *Eucalyptus* species such as *E. globulus* and *E. nitens*(RAYMOND; APIOLAZA, 2004; POKE et al., 2006; VOLKER; POTTS; BORRALHO, 2008; HAMILTON et al., 2009). To our knowledge, few published studies have been carried out with tropical species such as *E. grandis* and *E. urophylla* as well as with the most planted hybrid in Brazil (*E. grandis* x *E. urophylla*). Only recently more comprehensive reports have appeared in *E. urophylla* (HEIN et al., 2012; DENIS et al., 2013). Several such studies are routinely performed by private companies and for that reason they do not get published.

The estimation of genetic parameters such as heritability and genetic correlations rely on the estimation of variance and covariance components, which are derived from the expected genetic relatedness of sampled individuals. Genetic parameters of a trait are estimated by correlating the phenotypic resemblance among relatives with the proportion of the genome that they share identical by descent. In standard quantitative genetics experiments expected coefficient of relationship are used for this purpose in the expected numerator relationship matrix based on the supposedly correct pedigree, assuming unrelatedness of the

original parents involved in the crosses. Nevertheless, for eucalypts, accurate pedigree information is not always available and frequently open pollinated or polymix half-sib families are used to advance generations. Moreover, even when fully pedigreed populations are used the information is subject to errors and cryptic genetic relatedness may exist among individual trees selected in the original provenance/progeny trials where little is known about the actual structure of the introduced seed lots.

In the last decade, high-throughput genotyping platforms have become available for several plant and animal species, allowing fast and increasingly inexpensive genotyping of thousands of markers at the genome-wide scale. SNP (Single Nucleotide Polymorphisms) markers are the most used type of markers due to their abundance, assay reliability and genome coverage. These markers, when assayed at the genome-wide scale, allow the estimation of realized relatedness between pairs of individuals, by capturing the Mendelian segregation component, thus increasing the accuracy of parameters estimation in comparison to expected relatedness(HAYES; VISSCHER; GODDARD, 2009; LEE et al., 2010). In breeding when using the Henderson's animal model (HENDERSON, 1984) the traditional numerator relationship matrix is substituted for what is generally called the genomic or realized relationship matrix estimated using markers (ENDELMAN; JANNINK, 2012). By this procedure, even when the relationship between pairs of individuals is unknown, it can be fully estimated (EL-KASSABY; LSTIBUREK, 2009; PORTH et al., 2013). In plants, the development of large SNP panels with > 40,000 SNPs has been accomplished for important crops such as maize (GANAL et al., 2011) and soybean (SONG et al., 2013). For tree species SNP panels including 34,000 SNPs have been reported for poplar (GERALDES et al., 2013) and, more recently, a chip with 60,639 SNPs covering the reference genome at an average density of one SNP every ~10 kbp has been developed for *Eucalyptus* species (SILVA-JUNIOR et al., 2013). This powerful genotyping tool, now allows carrying out a number of detailed genetic analyses, including the precise estimation of relatedness and derived genetic parameters in eucalypt populations for which no pedigree information is available.

In the interface between genomics and quantitative genetics, the availability of large panels of genome-wide markers has also opened the way to genomic predictions of complex traits based exclusively on genotypic information provided by the markers. The idea of using markers to predict phenotypes is not recent (HALEY; VISSCHER, 1998; MEUWISSEN; HAYES; GODDARD, 2001). Nevertheless, its actual implementation only became feasible with the availability of genome-wideSNPs. The prediction of genomic breeding values opened the possibility to select individuals for which no phenotypes are available, based on

their genomic information, an approach that has now been called genome-wide selection (GWS) our just genomic selection (GS) (SCHAEFFER, 2006; GODDARD; HAYES, 2007). The approach puts into practice the old idea of marker-assisted selection (LANDE; THOMPSON, 1990) for quantitative traits. GS is now an established technology used routinely in dairy cattle breeding (PRYCE; DAETWYLER, 2012)with good perspectives in plant breeding as well(BERNARDO, 2008; GRATTAPAGLIA et al., 2009; JANNINK; LORENZ; IWATA, 2010), although the discussion is still open and contrasting views exist regarding its real prospects for annual crop plants (NAKAYA; ISOBE, 2012; JONAS; DE KONING, 2013). A general consensus exists, however, regarding the optimistic perspectives of genomic selection for perennial crops such as forest and fruit trees, in which considerable gains could be made per unit time by reducing the length of the breeding cycles, provided that genotyping costs are kept low (GRATTAPAGLIA; RESENDE, 2011; KUMAR et al., 2012). In forest tree breeding, deterministic simulations assessed the impact of linkage disequilibrium (LD), modeled by the effective population size and marker density, the size of the "training" population, trait heritability and number of QTL on prediction accuracies. Results indicated thatGS has the potential to radically improve the efficiency of tree breeding (GRATTAPAGLIA; RESENDE, 2011). Two experimental studies corroborated those findings, one in *Eucalyptus*(RESENDE M.D.V. et al., 2012) and another one in loblolly pine (RESENDE, M.F.R. et al., 2012), reporting accuracies of genomic models as good as or even higher than standard phenotypic selection. Still, these studies as the vast majority of the other published studies in plants, have only assessed the outlooks of genomic selection by cross validation within the same generation. Studies are therefore needed that will actually move from generating estimates of genomic prediction by cross validation to the actualrealized operational performance of genomic selection across generations.

The aim of this study was to provide a concrete demonstration of how to efficiently bridge the quantitative genetics of complex growth and wood properties traits with powerful genomic technologies, towards a more innovative approach to tree improvement in general and eucalypt breeding in particular. To this end, a large SNP-panel with more than 60,000 SNP markers was used in combination to advanced phenotyping procedures, to increase our understanding on the genetic control of fifteen growth and wood chemical and physical properties in an elite breeding population of *Eucalyptus*. NIRS model estimates were obtained and genetic parameters such as narrow-sense heritabilities and genotypic correlations amongst traits, were assessed using a realized relationship matrix. The combined dataset of SNPs and phenotypes was subsequently used to develop and validate genomic prediction models for the

wide range of traits that impact the downstream use of the eucalypt wood, and display different genetic control and expected response to selection. Predictions were carried out using one SNP every 20 kbp in the *Eucalyptus* genome, a marker density considerably higher than the one used to date in previous GS reports in forest trees. Predictions were generated by two analytical approaches that differ regarding the assumed distribution of marker effects. Nevertheless, predictive abilities observed with RR-BLUP and Bayesian LASSO were similar, although the distribution of marker effects varied and considerably more so depending on the trait under study. Extensive coincidence of ranking trees at the individual level was observed between genome based selection and standard BLUP phenotypic selection. The results presented in this study open vast opportunities for an increased use of SNP marker data in *Eucalyptus* breeding programs, allowing accurate pedigree-record-free estimation of genetic parameters and the prediction of genomic breeding values. These applications should become routine in tree breeding programs for the years to come, significantly reducing the length of breeding cycles and, consequently, optimizing resource allocation and sustainability of the breeding endeavor.

References

ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS.**Statistical yearbook 2013**:base year 2012. Brasília, 2013.148 p.

BERNARDO, R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. **Crop Science**, Madison, v. 48, n. 5, p. 1649-1664, 2008.

COSTA E SILVA, J.; BORRALHO, N.; ARAÚJO, J.; VAILLANCOURT, R.; POTTS, B. Genetic parameters for growth, wood density and pulp yield in Eucalyptus globulus. **Tree Genetics & Genomes**, Heidelberg, v. 5, n. 2, p. 291-305,2009.

DENIS, M.; FAVREAU, B.; UENO, S.; CAMUS-KULANDAIVELU, L.; CHAIX, G.; GION, J. M.; NOURRISIER-MOUNTOU, S.; POLIDORI, J.; BOUVET, J. M. Genetic variation of wood chemical traits and association with underlying genes in Eucalyptus urophylla. **Tree Genetics & Genomes**, Heidelberg, v. 9, n. 4, p. 927-942, 2013.

ELDRIDGE, K.; DAVIDSON, J.; HARWOOD, C.; VAN WYK, G. **Eucalypt domestication and breeding**. Oxford: Clarendon Press, 1993. 288 p.

EL-KASSABY, Y.A.; LSTIBUREK, M. Breeding without breeding. **Genetics Research**, New York, v. 91, n. 2, p. 111-120, 2009.

ENDELMAN, J.B.; JANNINK, J.L. Shrinkage estimation of the realized relationship matrix.**G3: Genes, Genomes, Genetics**, Bethesda, v. 2, n. 11, p. 1405-1413, 2012.

GANAL, M.W.; DURSTEWITZ, G.; POLLEY, A.; BERARD, A.; BUCKLER, E.S.; CHARCOSSET, A.; CLARKE, J.D.; GRANER, E.M.; HANSEN, M.; JOETS, J.; LE PASLIER, M.C.; MCMULLEN, M.D.; MONTALENT, P.; ROSE, M.; SCHON, C.C.; SUN, Q.; WALTER, H.; MARTIN, O.C.; FALQUE, M.A Large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. **PLoS One**, San Francisco, v. 6, n. 12, p. 1-15,2011.

GERALDES, A.; DIFAZIO, S.P.; SLAVOV, G.T.; RANJAN, P.; MUCHERO, W.; HANNEMANN, J.; GUNTER, L.E.; WYMORE, A.M.; GRASSA, C.J.; FARZANEH, N.; PORTH, I.; MCKOWN, A.D.; SKYBA, O.; LI, E.; FUJITA, M.; KLAPSTE, J.; MARTIN, J.; SCHACKWITZ, W.; PENNACCHIO, C.; ROKHSAR, D.; FRIEDMANN, M.C.; WASTENEYS, G.O.; GUY, R.D.; EL-KASSABY, Y.A.; MANSFIELD, S.D.; CRONK, Q.C.B.; EHLTING, J.; DOUGLAS, C.J.; TUSKAN, G.A. A 34K SNP genotyping array for *Populus trichocarpa*: design, application to the study of natural populations and transferability to other *Populus* species. **Molecular Ecology Resources**, Hoboken, v. 13, n. 2, p. 306-323, 2013.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**, Hoboken, v. 124, n. 6, p. 323-330, 2007.

GRATTAPAGLIA, D.; RESENDE, M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, Heidelberg, v. 7, n. 2, p. 241-255, 2011.

GRATTAPAGLIA, D.; PLOMION, C.; KIRST, M.; SEDEROFF, R.R. Genomics of growth traits in forest trees. **Current Opinion in Plant Biology**, London, v. 12, n. 2, p. 148-156, 2009.

HALEY, C.S.; VISSCHER, P.M. Strategies to utilize marker-quantitative trait loci associations. **Journal of Dairy Science**, New York, v. 81, p. 85-97, 1998.

HAMILTON, M.G.; RAYMOND, C.A.; HARWOOD, C.E.; POTTS, B.M. Genetic variation in *Eucalyptus nitens* pulpwood and wood shrinkage traits. **Tree Genetics & Genomes**, Heidelberg, v. 5, n. 2, p. 307-316, 2009.

HAYES, B.J.; VISSCHER, P.M.; GODDARD, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, New York, v. 91, n. 1, p. 47-60, 2009.

HEIN, P.R.G.; BOUVET, J.M.; MANDROU, E.; VIGNERON, P.; CLAIR, B.; CHAIX, G. Age trends of microfibril angle inheritance and their genetic and environmental correlations with growth, density and chemical properties in *Eucalyptus urophylla* ST Blake wood. **Annals of Forest Science**, Paris, v. 69, n. 6, p. 681-691, 2012.

HENDERSON, C.R. Estimation of variances and covariances under multiple trait models. **Journal of Dairy Science**, New York,v. 67, n. 7, p. 1581-1589, 1984.

JANNINK, J. L.; LORENZ, A. J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, Oxford, v. 9, n. 2, p. 166-177, 2010.

JONAS, E.; DE KONING, D.J. Does genomic selection have a future in plant breeding? **Trends in Biotechnology**, London, v. 31, n. 9, p. 497-504, 2013.

KUMAR, S.; CHAGNE, D.; BINK, M.C.A.M.; VOLZ, R.K.; WHITWORTH, C.; CARLISLE, C. Genomic selection for fruit quality traits in apple (*Malus domestica* Borkh.). **PLoS One**, San Francisco, v. 7, n. 5, p. e36674, 2012.

LADIGES, P.Y.; UDOVICIC, F.; NELSON, G. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. **Journal of Biogeography**, Hoboken, v. 30, n. 7, p. 989-998, 2003.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, Bethesda, v. 124, n. 3, p. 743-756, 1990.

LEE, S.H.; GODDARD, M.E.; VISSCHER, P.M.; VAN DER WERF, J.H.J. Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. **Genetics Selection Evolution**, London, v. 42, n. 22, p. 1-14,2010.

MEUWISSEN, T.H.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-29, 2001.

NAKAYA, A.; ISOBE, S.N. Will genomic selection be a practical method for plant breeding? **Annals of Botany**, Oxford, v. 110, n. 6, p. 1303-1316,2012.

POKE, F. S.; POTTS, B. M.; VAILLANCOURT, R. E.; RAYMOND, C. A. Genetic parameters for lignin, extractives and decay in Eucalyptus globulus. **Annals of Forest Science**, Paris, v. 63, n. 8, p. 813-821, 2006.

PORTH, I.; KLAPSTE, J.; SKYBA, O.; LAI, B.S.K.; GERALDES, A.; MUCHERO, W.; TUSKAN, G.A.; DOUGLAS, C.J.; EL-KASSABY, Y.A.; MANSFIELD, S.D. *Populus trichocarpa* cell wall chemistry and ultrastructure trait variation, genetic control and genetic correlations. **New Phytologist**, Hoboken, v. 197, n. 3, p. 777-790, 2013.

PRYCE, J.E.; DAETWYLER, H.D. Designing dairy cattle breeding schemes under genomic selection: a review of international research. **Animal Production Science**, Collingwood, v. 52, n. 2/3, p. 107-114, 2012.

RAYMOND, C.A.; SCHIMLECK, L.R. Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. **Canadian Journal of Forest Research/Revue Canadienne de Recherche Forestiere**, Ottawa, v. 32, n. 1, p. 170-176, 2002.

RAYMOND, C.R.; APIOLAZA, L.A. Incorporating wood quality and deployment traits in *Eucalyptus globulus* and *Eucalyptus nitens*. In: CARSON, M. (Ed.). **Plantation forest biotechnology for the 21st century**. Rotorua: Forest Research New Zealand, 2004. p. 87-99.

RESENDE, M.D.V.; RESENDE, 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, G.J.; KILIAN, A.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, Hoboken, v. 194, n. 1, p. 116-128, 2012.

RESENDE, M.F.R.; MUNOZ, P.; RESENDE, M.D.V.; GARRICK, D.J.; FERNANDO, R.L.; DAVIS, J.M.; JOKELA, E.J.; MARTIN, T.A.; PETER, G.F.; KIRST, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). **Genetics**, Bethesda, v. 190, n. 4, p. 1503-1510, 2012.

REZENDE, G.D.S.P.; RESENDE, M.D.V.; ASSIS, T.F. *Eucalyptus* breeding for cloanal forestry. In: FENNING, T. (Ed.). **Challenges and opportunities for world's forests in the 21st century**.Edinburgh:Springer, 2014. p. 393-424.

SCHAEFFER, L.R. Strategy for applying genome-wide selection in dairy cattle. **Journal of Animal Breeding and Genetics**, Hoboken, v. 123, n. 4, p. 218-223, 2006.

SCHIMLECK, L.R.; RAYMOND, C.A.; BEADLE, C.L.; DOWNES, G.M.; KUBE, P.D.; FRENCH, J. Applications of NIR spectroscopy to forest research. **Journal of the Technical Association of the Australian and New Zealand Pulp and Paper Industry**, Bundoora, v. 53, p. 458-464, 2000.

SILVA-JUNIOR, O.B.; FARIA, D.A.; TOGAWA, R.C.; GRATTAPAGLIA, D. *Eucalyptus* genotyping taken to the next level: development of the " EucHIP60k.br" based on large scale multi-species SNP discovery and ascertainment. In: IUFRO TREE BIOTECHNOLOGY CONFERENCE 2013, 2013, Asheville.**Proceedings…** 2013.

SONG, Q.J.; HYTEN, D.L.; JIA, G.F.; QUIGLEY, C.V.; FICKUS, E.W.; NELSON, R.L.; CREGAN, P.B. Development and evaluation of soySNP50K, a high-density genotyping array for soybean. **Plos One**, San Franscisco, v. 8, n. 1, p. 1-12,2013.

STACKPOLE, D.J.; VAILLANCOURT, R.E.; ALVES, A.; RODRIGUES, J.; POTTS, B.M. Genetic variation in the chemical components of *Eucalyptus globulus*wood. **G3: Genes, Genomes, Genetics**, Bethesda, v. 1, n. 2, p. 151-159, 2011.

VOLKER, P.; POTTS, B.; BORRALHO, N. Genetic parameters of intra- and inter-specific hybrids of *Eucalyptus globulus* and *E. nitens*. **Tree Genetics & Genomes**, Heidelberg, v. 4, n. 3, p. 445-460, 2008.

2 PEDIGREE-AGNOSTIC ESTIMATES OF GENETIC PARAMETERS FOR GROWTH, CHEMICAL AND PHYSICAL WOOD PROPERTIES IN *Eucalyptus* USING DATA FROM 29,000 GENOME-WIDE SNPS

Abstract

A better understanding of the heritability and multiple genetic and phenotypic correlations amongst wood and growth traits in species of *Eucalyptus* is of paramount importance to better exploit its natural variation and to advance breeding programs. Rare are the reports on genetic parameters for wood properties for tropical *Eucalyptus* species as the ability to non-destructively assess such traits for many trees is still a major bottleneck. In our study, we examined fifteen traitsincluding growth and wood chemical and physical properties in 1,000 individuals involving 45 full-sib families sampled from a hybrid *Eucalyptus grandis* by *E. urophylla* breeding trial.We used the power of near-infrared spectroscopy (NIRS) as a prediction tool for high-throughput phenotyping of wood traits, coupled to high-density genome-wide SNP data to obtain pedigree-record-free estimates of trait variance components, heritabilities and genetic and phenotypic correlations based on a realized relationship matrix. High NIRS prediction power was obtained for wood chemical traits such as lignin content and composition (correlations between 0.83 and 0.93) while variable predictions were observed for physical traits with the exception of wood density, where NIRS models provided a correlation of 0.73. Realized relationships based on 29,090 polymorphic SNP markers revealed variable levels of genetic relatedness between pairs of individuals in the population and extensive genome sharing between families, therefore providing more accurate estimation of genetic parameters when compared to the pedigree-based numerator relationship matrix. The estimated narrow-sense heritabilities were medium to high for growth traits (0.34 to 0.44), high for wood chemical traits (0.56 to 0.85) and from 0.11 to 0.63 for wood physical traits. High positive correlations among growth traits and negative between cellulose and lignin content were observed while correlations between wood chemical and physical traits and between growth and wood quality traits in general were low, although significant. By using the combined power of NIRS and high-density genome-wide genotyping this is the first study in forest trees and in plantsin general to demonstrate the value of using high-density SNP data to support quantitative genetics estimations that significantly impact the expected gains from selection. Furthermore the genetic parameters reported for a number of key traits in tropical *Eucalyptus* should provide useful guidelines for ongoing *Eucalyptus* breeding efforts.

Keywords: Wood quality; Heritability; Tree breeding; GBLUP; Molecular markers; Genotypic correlation

2.1 Introduction

Eucalyptus is the most extensively planted genus of hardwood trees worldwide. In Brazil, eucalypt forests comprise 5.1 million hectares, corresponding to 76.6% of fast growing forest plantations (ABRAF, 2013). Several species of the genus display fast growth and

extensive adaptability in a wide range of tropical and subtropical areas and soil types. Eucalypt wood is highly versatile, used for energy, solid wood products pulp and paper.

Wood phenotyping in the context of *Eucalyptus* breeding programs has been a challenging task. These programs usually deal with thousands of individuals and some wood traits can only be measured when trees are at least 3-years-old. Moreover, standard wood analysis methods generally need large amounts of sample, including the whole tree for some specific measurements. The procedure is expensive and destructive, reason why it is usually carried out only in the late stages of the breeding selection cycle past progeny tests and during final clonal trials. The main and frequently the only wood trait phenotyped at this stage is wood density since it is easy to measure and correlated, to a certain extent, with other important wood traits such as lignin content and pulp yield. However, the possibility of directly measuring such traits has now become relevant in breeding programs, especially as different wood characteristics are pursued, depending on the wood final use.

Wood phenotyping can be expensive and time consuming, currently representing a recognized bottleneck in breeding programs. To reduce time and cost of wood phenotyping methods that can predict wood traits based on near-infrared reflectance spectroscopy (NIRS) have become increasingly appealing (SCHIMLECK et al., 2000). The near-infrared reflectance-based modeling uses the spectra of a sample to predict its compounds. It has been successfully used in poplar for predictions of S:G ratio (ROBINSON; MANSFIELD, 2009), calorific value, specific gravity (MARANAN; LABORIE, 2007), wood decay and density (STIRLING et al., 2007) and of microfibril angle in *Pinustaeda*(SCHIMLECK et al., 2007). In *Eucalyptus*, several studies applied NIRS to predict wood properties traits. *Eucalyptus globulus* had NIRS calibrations for cellulose content with strong correlations, varying from 0.82 to 0.94, in different sites (RAYMOND; SCHIMLECK, 2002). Wood chemical components were also predicted in another study of *E. globulus*(STACKPOLE et al., 2011), in order to observe the natural genetic variation for S:G ratio, extractives, cellulose content and pulp yield. Genetic parametersof Pilodyn penetration under barkwere also estimated for *E. globulus* based on NIRS predictions to indirectly measure density and pulp yield (COSTA E SILVA et al., 2009).

The breeder’s equation, which infers the response to selection, is a product of the selection differential and heritability, this last one the genetic parameter that indicates the proportion of the phenotypic variance explained by genetic differences (FALCONER;

MACKAY, 1996). The breeder defines selection differential while the trait heritability has to be estimated. A better understanding of the genetic parameters and correlations amongst *Eucalyptus* growth and wood quality traits is of paramount importance to increase the response to selection and thus carry out a successful breeding program. Most of the studies to date have focused on growth traits and the few published studies on wood quality traits were mainly done for temperate species such as *E. globulus* and *E. nitens*(RAYMOND; APIOLAZA, 2004; POKE et al., 2006; VOLKER; POTTS; BORRALHO, 2008; HAMILTON et al., 2009; STACKPOLE et al., 2011). Reports on genetic parameters for tropical *Eucalyptus* are rare. Two studies reported on genetic parameters for wood quality and growth traits in progeny trials of *E. urophylla* using NIRS derived phenotypes (HEIN et al., 2012; DENIS et al., 2013) and correlations for a small number of traits were estimated for a small bi-parental mapping population in the context of a QTL mapping study (GION et al., 2011). To our knowledge there are no published reports describing the extent of genetic variation, heritabilities, genetic and phenotypic correlations amongst wood quality and growth traits for the widely planted *E. grandis x E. urophylla* hybrid.

Accurate assessment of genetic parameters such as heritability and genetic correlations rely on the estimation of variance and covariance components, which in turn are derived from the genetic relatedness of individuals sampled. Genetic control of a trait is estimated by correlating the phenotypic resemblance with the proportion of the genome that two relatives share identical by descent. In standard quantitative genetics experiments an expected coefficient of relationship of 0.25 is generally assumed for half-siblings and 0.5 for full sibs, based on the pedigree information and presumed unrelatedness of parents involved in the crosses. However, in *Eucalyptus* with a mixed mating system and small hermaphrodite flowers it is known that a number of factors can lead to higher average relatedness even in breeding populations. A small but frequent proportion of selfing is common in open pollinated half-sib families and pollen contamination or identification errors can always occur during controlled crosses, notwithstanding potential cryptic relatedness of parents involved especially in elite breeding populations. The availability of large numbers of molecular markers, particularly SNPs (Single Nucleotide Polymorphisms), covering the whole genome, now allows obtaining the so called realized relationship matrix which provides the actual marker-estimated relatedness amongst all individuals involved in the experimental population used to estimate the genetic parameters instead of the conventional numerator relationship matrix (HAYES; VISSCHER; GODDARD, 2009). The use of realized relationship matrices

reveals complex connections between individuals, sharing different parts of the genome, being analogous to the random variance-covariance matrix in Henderson's BLUP analysis (HENDERSON, 1984), called GBLUP (Genomic-BLUP) when the realized matrix is used.

The value of using marker realized relationship matrices for genetic parameter has been demonstrated in forest trees. In a Scots pine open pollinated progeny trial marker data for nine microsatellites were used in combination to pedigree information (KORECKÝ et al., 2013), while in *Populus* an array of 9,432 SNP markers was used in a common garden plot of 334 trees representing a latitudinal range of natural populations (PORTH et al., 2013). In *Eucalyptus*, while molecular markers have been extensively used to characterize genetic variation in populations, fingerprint individuals, estimate outcrossing rates and check or infer parentage (GRATTAPAGLIA et al., 2012), no study has yet examined the use of a marker-based realized relationship matrix to estimate genetic parameters. A tentative effort was reported by using population level average outcrossing rates obtained with isozyme markers to adjust the additive relationship matrix in *E. cladocalyx*(BUSH et al., 2011). Results showed that the commonly applied coefficient of relationship of 1/2.5 was suitable for correcting variance component and heritability estimates. Although several microsatellites are available for *Eucalyptus*, precise estimatimation of relationships requires genotyping a large number of them which would be labor intensive. We have developed high-throughput marker platforms for *Eucalyptus* in the last few years starting with DArT markers (SANSALONI et al., 2010) and low density SNPs (GRATTAPAGLIA et al., 2011). Recently, we have developed a high-density SNP chip that provides robust SNP data for up to 60,000 markers evenly distributed across the whole *Eucalyptus* reference genome at an average density of 1 SNP every 10kbp (SILVA-JUNIOR et al., 2013), providing an unprecedented resolution for genetic analysis. In this study, we employed this powerful genotyping platform to construct a genome-wide relationship matrix amongst 1,000 individuals sampled in a progeny trial elite breeding population involving 45 full-sib families of first and second-generation *E. grandis*, *E. urophylla*hybrids. This matrix was used to generate highly accurate estimates of heritabilities and genetic and phenotypic correlations amongst 15 key growth and wood traits, including chemical and physical properties that largely impact the wood use for pulp, paper and energy.

2.2 Development

2.2.1 Material and Methods

2.2.1.1 *Eucalyptus* population

The *Eucalyptus* population used in this study belongs to the breeding programof International Paper Brazil. The population is a full-sib progeny test of *Eucalyptus grandis*, *E. urophylla, E. camaldulensis* and hybrids (*E. grandis* x *E. urophylla*), composed of 58 full-sib families, comprising 2,784 trees, planted in 2006, in Brotas, São Paulo State, Brazil (22°S). Trees were measured for diameter at breast height (DBH) and plant height, and wood volume estimated thereof. Analyses were carried out using SELEGEN-REML/BLUP (RESENDE; OLIVEIRA, 1997) and a subset of 1,000 trees out of the 2,784 in the trial was selected based on growth (wood volume), according to the mixed model:

$$\mathbf{y = Xr + Zg + Wp + e}$$

where **y** is the vector of phenotypic data (wood volume), **r** is the vector of fixed effects (i.e. mean and experimental effects), **g** is the vector of random additive genetic effects of individuals, **p** is the vector of fixed plot effects and **e** is the random residual effect. **X**, **Z** and **W** are incidence matrices for each of the effects. The selected *Eucalyptus* population of 1,000 trees included 45 full-sib families and 46 different parents, in which 25 families were *E. grandis* x *E. urophylla* hybrids, totaling 610 trees (61%). *E. camaldulensis* was involved in only one cross with *E. grandis*, with 24 full-sibs (2%). Second-generation hybrid families involving *E. grandis* and *E. urophylla* were also sampled (table 2.1). From this point on the population will be addressed as *E.grandis* x *E. urophylla* hybrid population, since the population is formed primarily by hybrids between these two species.

Table 2.1 – Number of trees and families used according to the cross type, considering the species/hybrids used on the crosses

Cross type	Crossing species	Number of families	Number of trees
GxU	*E. grandis*x *E. urophylla*	25	610
GxC	*E. grandis*x *E. camaldulensis*	1	24
H_1xG	(*E. grandis*x *E. urophylla*)[1] x *E. grandis*	5	105
H_1xU	(*E. grandis*x *E. urophylla*)[1] x *E. urophylla*	3	26
H_2xG	(*E. grandis*x *E. urophylla*)[2] x *E. grandis*	1	16
H_2xDH	(*E. grandis*x *E. urophylla*)[2] x (*E. grandis*x *E. urophylla*)[3]	2	37
DHxG	(*E. grandis*x *E. urophylla*)[3] x *E. grandis*	6	126
DHxU	(*E. grandis*x *E. urophylla*)[3] x *E. urophylla*	2	56
Total		45	1,000

[1]hybrid which female tree is *E. urophylla*; [2]hybrid which female tree is *E. grandis*; [3]hybrid which male and female trees are *E. grandis*x *E. urophylla*hybrids

2.2.1.2 Wood traits phenotyping

Wood sampling was performed when plants were 5-year-old. Wood dust samples were taken using a driller (10 mm diameter) at breast high, always in a north-south direction (figure 2.1b). The wood was stored in paper envelopes (figures 2.2a and 2.2b), dried at room temperature and ground using a Willey mill (figure 2.2c). The 40-60 mesh portion of the sample was used in the analyses (figure 2.2d). Near-infrared spectra of wood dust samples were obtained using a NIRSystems 5000 equipment (FOSS, Hillerød, Denmark), reading every second wavelength, from 1,100 to 2,500 nm. Each sample was read 16 times, using the average of each one of 700 wavelengths.

In order to reduce the number of samples to be analyzed by wet chemistry lab procedures, a representative subset of 350 samples was selected to carry out the chemical and physical wood analysis, constituting a calibration set. Samples were selected by Kennard and Stone (1969) sampling algorithm, based on Euclidean distances of samples spectra aiming at maximizing sample variation and representativeness.

2.2.1.2.1 Wood chemical traits

Cell wall carbohydrates and total lignin:Ground wood (40-60 mesh) was extracted with acetone in Soxhlet apparatus for 12 hours. Cell wall carbohydrates (cellulose and hemicellulose) and total lignin (acid-soluble and -insoluble lignin, in combination forming the

total lignin) were determined as described by Huntley et al. (HUNTLEY et al., 2003), using the Klason method, with small modifications. Cell wall carbohydrates were quantified with a high-performance liquid chromatography system (HPLC) using a Dionex (DX-600, Sunnyvalle, CA, USA), equipped with a PA1 (Dionex) column, detector with a gold electrode and SpectraAS3500 auto injector (Spectra-Physiscs, Santa Clara, CA, USA). Carbohydrates amounts were quantified relative to monomeric cell wall-associated carbohydrates (glucose, xylose, mannose, galactose, rhamnose and arabinose) as standards (PORTH et al., 2013). The amounts of Klason lignin and cell wall sugars were obtained in percentages, relative to the initial weight of dry wood sample analyzed.

S:G ratio: Lignin monomer subunits (syringil:guayacil) proportion was determined from acetone-extracted ground wood as described by Robinson and Mansfield (2009) and analyzed via gas chromatography (GC) on a Hewlett Packard 5890 series II equipment (Agilent Technologies, Santa Clara, CA, USA), equipped with an autosampler, splitless injector, flame ionizing detector and a 30-m 5% diphenyl-95% dimethyl polysiloxane-coated RTX-5MS capillary column (inner diameter, 0.25mm).

2.2.1.2.2 Wood physical traits

Increment wood cores were collected for all 1000 trees of the selected population. Increment cores (12 mm) were collected at breast height, in a north-south direction (figure 2.1c). The northern half of the core was used for density and microfibril angle (MFA) analysis and the southern half for fiber length, width and coarseness.

Wood density:The northern half of the wood cores were precision cut in 1.67mm-thick sections, using a custom-built twin-blade pneumatic saw (PORTH et al., 2013), and acetone extracted in Soxhlet apparatus for 12 hours. The wood sections were acclimated to 7% moisture content and then scanned by X-ray densitometry (QTRS-01X; Quintek Measurement Systems Inc., Knoxville, TN, USA), from pith to bark. The measurements across the section were averaged to determine the sample density. With the purpose of establishing a regression for the X-ray densitometry, ten samples were randomly selected and had precisely recorded their weight over volume and then scanned in the equipment, to estimate the other samples of the phenotyping calibration set.

Microfibril angle (MFA): Unlike subtropical and temperate tree species, tropical eucalypts have a continuous growth pattern, and annual growth rings are unclear. Thus, MFA was measured 1.0 cm adjacent to the bark, instead of choosing a certain year, following procedures described earlier (PORTH et al., 2013). Briefly, precision-cut samples, used for wood density determination, were also used for MFA in a Bruker D8 Discover (Bruker AXS Inc., Madison, WI, USA) X-ray diffraction instrument equipped with an area detector (GADDS) to collect diffraction patterns, which contain reflection information of the microfibril orientation in the wood sample.

Fiber length, width and coarseness:The southern part of the increment cores were used for fiber analysis. Samples were incubated in Franklin solution (30% hydrogen peroxide and glacial acid; 1:1 ratio) at 70°C for 48 hours and macerated. Afterward, samples were washed in deionized water until the samples had been neutralized. The samples were filtered and dried at 105°C in an oven. A part of the sample had weight recorded and resuspended in distilled, deionized water and analyzed on a Fiber Quality Analyzer (FQA; Optest Equipment Inc., Hawkesbury, Ontario, Canada). Fiber length and width was taken as the average of all fibers measured and coarseness measured by the dry fiber mass per unit length ($mg.100m^{-1}$).

2.2.1.3 NIRS phenotype prediction

The calibration population was randomly divided in two subsets, being one for estimation and the other one for validation of the predictive models. The estimation set was formed by 250 samples and the validation set by 100 samples (figure 2.3). The Unscrambler® software (v.9.0; CAMO A/S, Oslo, Norway) was used to estimate the model parameters for each of the wood quality traits using Partial Least Squares (PLS) analysis for phenotype prediction, based on each sample spectra. For each trait, different transformations for the spectra were tested, in order to obtain higher accuracies. Root mean squared error of prediction (RMSEP), bias (the average value of the difference between predicted and measured values) and correlation (R) were calculated for the validation set and used to compare model estimates. The selected model and transformation for each trait was used to predict the phenotype of the remaining 650 samples of the full set of 1,000 trees that were not analytically phenotyped.

2.2.1.4 Genotypic data

Genomic DNA was extracted from xylem (figure 2.1a), scratching part of the sapwood, using a previously described CTAB method (GRATTAPAGLIA; SEDEROFF, 1994), and quantified in a nanodrop equipment to concentrations between 20-40ng.uL^{-1}. DNA samples were genotyped at Geneseek (Lincoln, NE) using an Infinium technology (Illumina) custom made chip for *Eucalyptus* (EucHIP60k.br)(SILVA-JUNIOR et al., 2013). Only SNPs with call rate ≥ 0.90 and MAF (minor allele frequency) ≥ 0.01 were used in further analyses.

2.2.1.5 Estimation of phenotypic and genetic parameters

Phenotypic data were adjusted for experimental design effects using the R package 'lme4' and the fitted phenotypes were used for the subsequent phenotypic and genetic analysis. Family-mean heritability (H^2) and accuracy (r) (FALCONER; MACKAY, 1996) are given by

$$H^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_e^2}{R}}$$

$$r = \sqrt{H^2}$$

whereσ_p^2 is the variance across progenies, σ_e^2 is the residual variance and R is the number of repetitions. Accuracy is the square-root of the family-mean heritability.

The genotypic parameters analysis was carried out using the package rrBLUP(ENDELMAN, 2011) , that solves the mixed model

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zg} + \mathbf{e}$$

$$\mathbf{g} \sim N(0, \mathbf{K}\sigma_g^2)$$

where$\boldsymbol{\mu}$ is the intercept, **g** is the vector for random effects (breeding values), **Z** is the incidence identity matrix for random effects and **e** is the random residual effect. The **K** matrix is the relationship matrix, which can be a realized (based on SNP data) or expected (pedigree-based numerator relationship matrix).The realized relationship matrix was calculated using the

function ‘A.mat’, in the packagerrBLUP, that uses the formula proposed by (VANRADEN, 2008)

$$K = \frac{\mathbf{WW'}}{2\sum p_i(1-p_i)}$$

where **W** is the $(n \times p)$ SNP matrix, centered by the population mean. The numerator relationship matrix was calculated in R using the package ‘pedigreemm’.

The function ‘mixed.solve’ returns the BLUP solutions for EBVs (estimated breeding values) and variance components. Narrow-sense heritabilitieswere calculated based on the variance components estimated via BLUP as

$$h^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2}$$

where$\hat{\sigma}_g^2$ is the estimated additive variance and $\hat{\sigma}_e^2$ is the estimated residual variance. Narrow sense heritabilities were estimated using variance components estimates from pedigree ($h^2_{g_{PED}}$) and SNP ($h^2_{g_{SNP}}$) information.

Pearson’s correlations between pairs of traits were calculated using the phenotype, excluding experimental design effects, and breeding values estimated using pedigree (pedigree-based numerator relationship matrix) and SNP information (SNP-based realized relationship matrix), using the R package ‘Hmisc’. Traits were grouped into three distinct classes: growth, wood chemical and wood physical traits. Genotypic correlations based on SNP-based realized relationship matrix ($r_{g_{SNP}}$), genotypic correlations based on pedigree-based numerator relationship matrix ($r_{g_{PED}}$) and phenotypic correlations (r_p) are presented with their respective significance levels (table 2.5).

Figure 2.1 – Wood sampling in the *Eucalyptus* population. For DNA extraction (a,b) pieces of wood were sampled. Wood dust samples (c,d) were collected for NIRS and chemical analyses. Wood cores samples (e,f) of 1.2mm were collected for physical analyses

Figure 2.2 – Preparation of wood dust samples. Wood samples were stored in paper envelopes and dried at room temperature (a, b). A Willey mill was used to grind the samples (c). Wood dust classification in 60/40 mesh sieves (d)

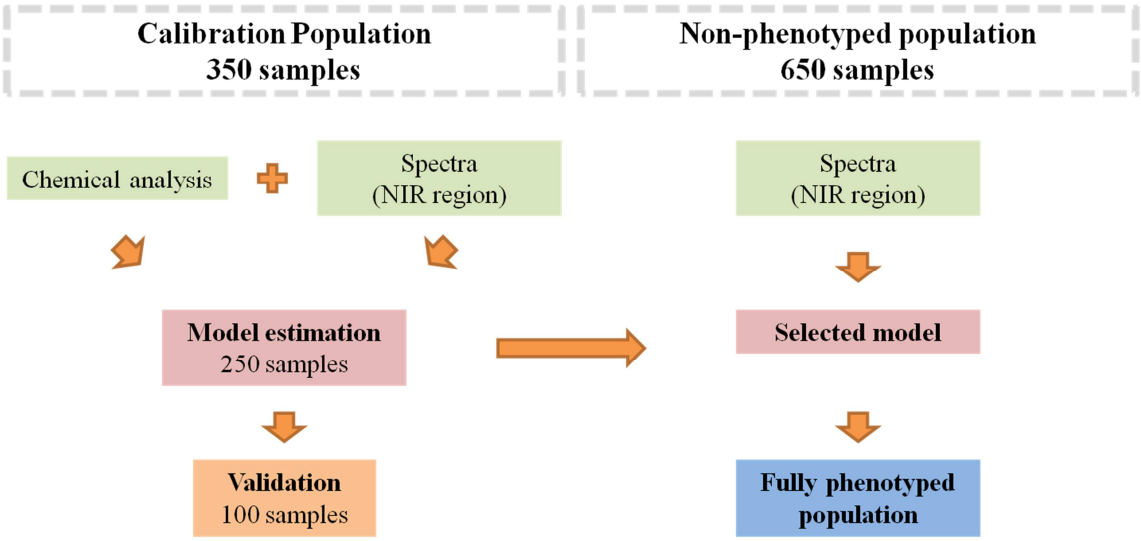

Figure 2.3 – Phenotyping scheme for wood chemistry and physical traits, based on NIRS (near-infrared spectroscopy) data. A subset of 350 samples was selected based on maximizing spectra variation and phenotyped to serve as a calibration population. Out of the 350 samples, 250 samples were used for parameters estimation and 100 samples were used for model validation

2.2.2 Results

2.2.2.1 SNP data

SNP genotype data with a call rate>90% were obtained for a total of 47,903 SNPs. After applying a filter for minimum allele fequency (MAF) greater than 0.01, a total of 29,090 SNPs were retained and used in the subsequent analyses. Since the *Eucalyptus* genome was used to design the chip, all the SNPs positions are known, in each one of the 11 chromosomes. SNPs genotyping coverage was high with an average of 2,645 SNPs per chromosome and an averageintermarker distance of ~20 kbp (table 2.2).

Table 2.2 –Distribution of SNPs markers retained with a call rate ≥ 0.90 and MAF ≥ 0.01 across the 11 chromosomes of the *Eucalyptus* reference genome

Chromosome	**Number of SNPs**	**Chromosome length (kb)**	**Average intermarker distance (kb)**
01	2,123	40,282	19.0
02	3,178	64,172	20.2
03	2,832	79,787	28.2
04	2,049	41,861	20.4
05	2,617	74,615	28.5
06	3,321	53,832	16.2
07	2,291	52,364	22.9
08	3,643	74,284	20.4
09	2,121	38,967	18.4
10	2,332	39,217	16.8
11	2,583	45,357	17.6
Total	29,090	604,738	-
Average	2,645	54,976	20.8

2.2.2.2 Genetic relationships

SNP data was used to estimate the realized relationship matrix, where the inbreeding coefficient (f) equals $\langle A_{ii} \rangle - 1$, where A_{ii} is the IBS (identity by state) relationship between an individual and itself and the angular brackets denote the average (ENDELMAN; JANNINK, 2012). In other words is the average of relationship matrix diagonal minus one. The calculated inbreeding coefficient in the experimental population was equal to -0.18. Such a negative value is not biologically explainable, standing as a mathematical artifact, however revealing that no detectable inbreeding exists in the population.

Off-diagonal elements reveal estimates of genetic covariance between pairs of individuals by the IBS approach. When using an expected relationship between pairs of individuals, most of the relationships are expected to be zero, since the pedigree is incomplete and only the parents are known and considered unrelated. However, using SNP marker data, the realized relationship between individuals can be estimated. In the studied population, the marker-estimated relationships displayed a wide variation, varying from -0.32 to 0.75, with an average equal to zero (figure 2.4).

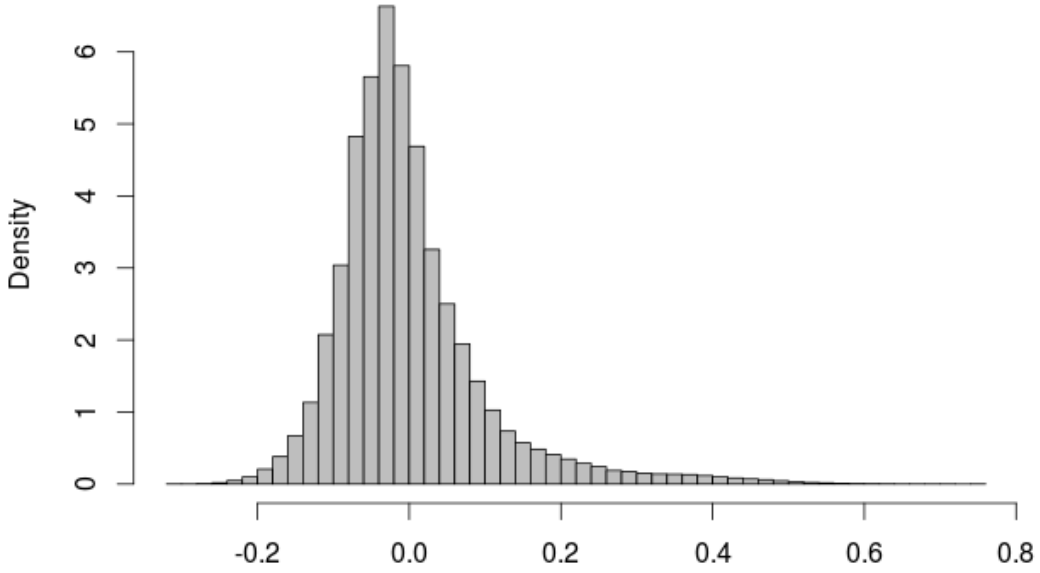

Figure 2.4 – Histogram of additive relationship between individuals of the *Eucalyptus* population (realized relationship matrix, excluding the diagonal) calculated using 29,090 SNPs

The difference between the expected and realized relationship matrices can be visualized using heatmaps (figure 2.5). In the expected relationship matrix only three classes of relationship are possible: full-sibs (0.50), half-sibs (0.25) and unrelated (< 0.01). Half-sibs are observed because some parents are used in more than one cross. Analyzing the realized relationship matrix the same pattern is observed, however the color pattern is progressive. Furthermore, unexpected relatedness between individuals is revealed and, in some cases, it is possible to detect putative pedigree errors, when individuals belonging to one family are more related to individuals in other families. Such unexpected crosses become a problem only when using the expected relationship matrix as the variance and covariance matrix. The SNP marker realized matrix corrects this kind of ocurrences. In the realized relationship matrix some negative values can be found, indicating that pairs of trees are less related than expected.

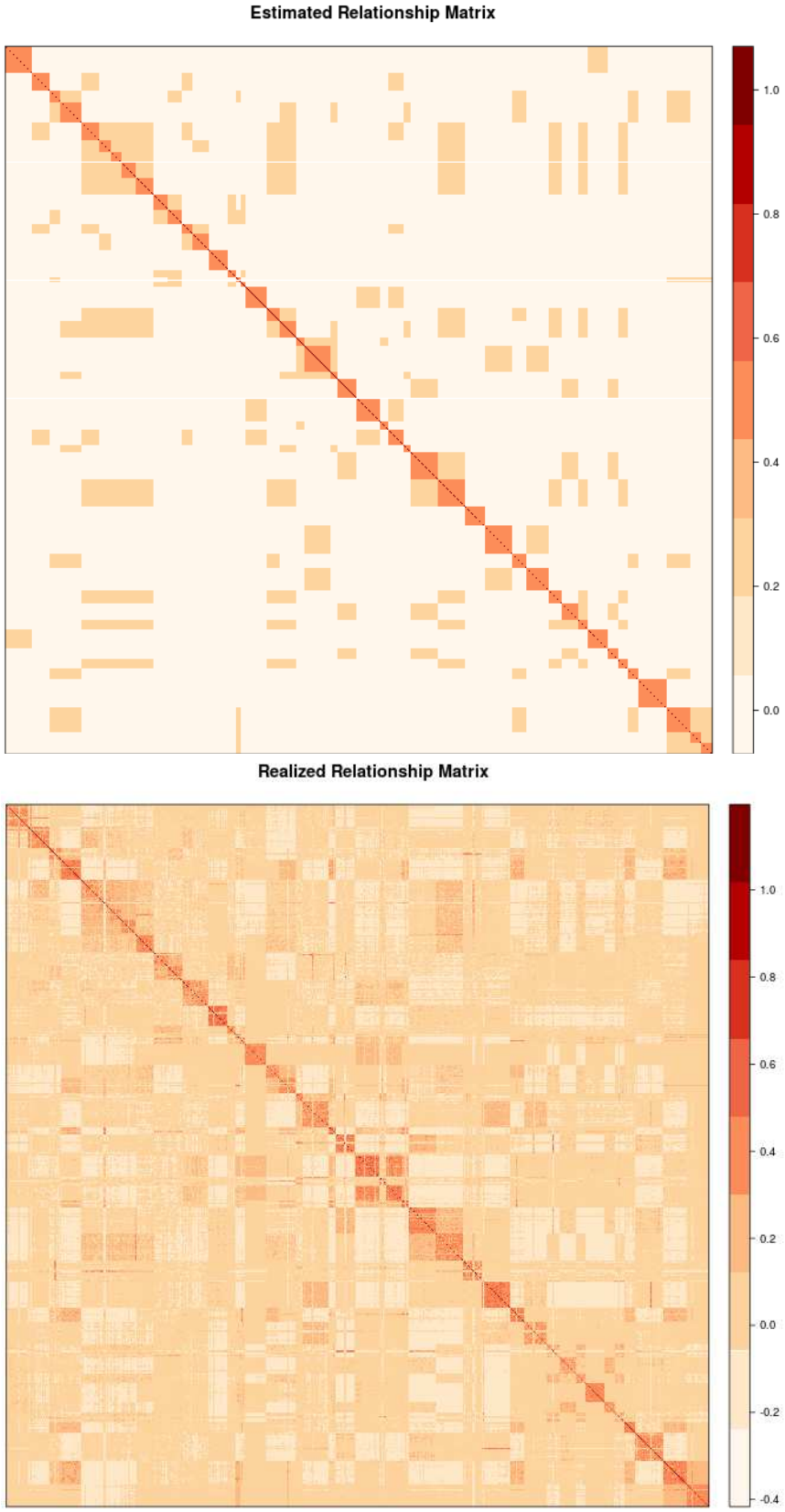

Figure 2.5 – Relationship matrices between pairs of individuals for the *Eucalyptus* population, comparing the expected relationship matrix (top) and the realized relationship matrix (bottom). Individuals are sequentially ordered according to their pedigree-expected families. Heatmap scale of relationship is depicted on the right hand side of the matrices

2.2.2.3 Wood traits prediction

NIRS models parameters were estimated in the phenotyping calibration set for chemical and physical traits. In order to obtain better estimates, spectra data transformation was necessary for some of the traits (2^{nd} derivative of Savitzky-Golay and Norris transformations), carried out using The Unscrambler® software. Wood chemical traits showed higher correlations between the predicted and real data, when compared to physical traits. For S:G ratio, for example, the correlation was 0.93 and for coarseness 0.17. The worst prediction occurred for MFA, where the correlation was negative (-0.13) (table 2.3).

Table 2.3 – Parameters used in the definition of the best NIRS models for prediction of chemical (cellulose, hemicellulose, S:G ratio, insoluble lignin, soluble lignin and total lignin) and physical (density, MFA, fiber length, fiber width and coarseness) traits in the experimental *Eucalyptus* population

Trait	Transformation[1]	Correlation[2]	RMSEP[3]	Bias
Cellulose (%)	-	0.57	1.993	-0.897
Hemicelluloses (%)	Norris	0.60	1.070	-0.198
S:G ratio	-	0.93	0.160	-0.030
Insoluble lignin (%)	Savitzky-Golay (2^{nd} derivative)	0.84	0.956	-0.233
Soluble lignin (%)	-	0.83	0.288	-0.158
Total lignin (%)	Savitzky-Golay (2^{nd} derivative)	0.84	0.922	-0.275
Density ($kg.m^{-3}$)	-	0.74	32.589	1.607
MFA (°)	-	-0.13	1.048	-0.060
Fiber length (mm)	Norris	0.43	0.055	0.024
Fiber width (µm)	-	0.24	1.197	0.810
Coarseness ($g.100m^{-1}$)	-	0.17	0.917	0.466

[1]Spectra data transformation; [2]Correlation in the validation population; [3]RMSEP, root mean square error of prediction

Higher prediction power was expected for chemical traits, since near-infrared spectra capture the chemical composition of wood. Cellulose and hemicellulose predictions power was lower than the other wood chemical traits as result of the calibration population size (200 samples). Poorer predictions were observed for physical traits with the exception of density,

where the correlation was high (0.74). This can be explained by the high correlation typically existing between density and increased levels of chemical compounds such as lignin found in wood (see below). All other physical traits such as MFA, fiber length and width and coarseness could not be predicted accurately and only the actual measured phenotypes in the calibration set were used in further analyses.

2.2.2.4 Trait genetic variation and heritabilities

Growth traits (DBH, plant height, wood volume and MAI) were collected directly on all 1000 standing trees. All chemical traits measurements were either obtained directly on the calibration set of 350 trees, or NIRS predicted for the remaining set of 650 trees, totaling 1000. Howeverbecause one of the samples was not genotyped, only 999 samples were eventually used in further analyses. Cellulose and hemicelluloses were measured in 200 samples due totime limitations as existing protocols had to beadaptated for *Eucalyptus*. For the physical traits only data for the calibration set was used to estimate the parameters. For MFA 348 out of 350 samples were phenotyped, because two wood cores presented knots. For coarseness, one sample could not be phenotyped because of equipment issues. Extensive genetic variation was observed for all measured traits in the breeding population (table 2.4). Noteworthy is the fact that trees with S:G ratio up to 4.2, density up to 645.5 kg/m^3 and cellulose content up to 55.4% were observed, an interesting result showing great potential for selecting high pulp yielding trees in this population. Estimated variation for volume growth was also high in the sampled trees reflecting the mild selection intensity applied for this particular study. Family-mean heritability and accuracy were also estimated. These are quality parameters, revealing experimental and trait characteristics. The highest accuracy was observed for S:G ratio, closely followed by the other lignin traits. MFA and fiber width accuracies were the lowest ones. This pattern was followed by the estimate of narrow-sense heritability, calculated using SNP data.Narrow-sense heritabilitiesestimated using a genomic realized relationship matrix varied from a low 0.11, for fiber width, to a high of 0.85 for S:G ratio, while those based on pedigree data were slightly higher for most traits, exception made for cellulose, hemicellulose, insoluble lignin, total lignin, fiber length and coarseness where the SNP-base estimates were higher (table 2.4).Heritabilities(SNP-based) were medium to high for growth traits (0.34 to 0.44), high for wood chemical traits (0.56 to 0.85) and from 0.11 to 0.63 for wood physical traits.

Table 2.4 – Summary of the analyzed data in the *Eucalyptus* population, for growth (DBH, plant height, wood volume and MAI), chemical (cellulose, hemicellulose, S:G ratio, insoluble lignin, soluble lignin and total lignin) and physical (density, MFA, fiber length, fiber width and coarseness) traits

Traits	n	Min	Mean	Max	H^2	Accuracy	$h^2_{g_{PED}}$	$h^2_{g_{SNP}}$
DBH (cm)	999	12.4	16.6	23.2	0.84	0.92	0.49	0.44
Plant height (m)	999	19.6	24.19	27.6	0.76	0.87	0.35	0.34
Wood Volume (m^3)	999	0.12	0.24	0.50	0.84	0.92	0.48	0.42
MAI ($m^3.ha^{-1}.year$)	999	28.9	58.6	121.6	0.84	0.92	0.48	0.43
Cellulose (%)[1]	999	41.8	48.9	55.4	0.82	0.91	0.36	0.56
Hemicelluloses (%)[1]	999	13.9	17.3	20.2	0.84	0.92	0.35	0.65
S:G ratio[1]	999	1.8	2.9	4.2	0.95	0.98	1.00	0.85
Insoluble lignin (%)[1]	999	20.7	25.2	28.8	0.91	0.96	0.60	0.66
Soluble lignin (%)[1]	999	2.2	3.5	4.9	0.94	0.97	0.92	0.69
Total lignin (%)[1]	999	24.4	28.8	32.1	0.91	0.95	0.60	0.67
Density ($kg.m^{-3}$)[1]	999	407.1	512.5	646.5	0.90	0.95	0.75	0.58
MFA (°)	348	10.5	12.9	17.5	0.36	0.60	0.13	0.13
Fiber length (mm)	350	0.59	0.75	0.93	0.66	0.81	0.46	0.63
Fiber width (µm)	350	17.2	19.8	24.7	0.34	0.58	0.14	0.11
Coarseness ($mg.100m^{-1}$)	349	4.4	7.1	11.0	0.43	0.65	0.21	0.29

[1]350 samples analyzed by laboratory procedures and 650 predicted via NIRS; n, sample size; Min, minimum value; Max, maximum value; H^2, family-mean heritability; Accuracy, given by $\sqrt{H^2}$; $h^2_{g_{PED}}$, narrow-sense heritability calculated using pedigree data; $h^2_{g_{SNP}}$, narrow-sense heritability calculated using SNP data

2.2.2.5 Phenotypic and genotypic correlations

Correlations amongst growth traits were strong and the variation in magnitude between genotypic and phenotypic correlations was not large, with some exceptions where significance also varied, notably for correlations of several traits with MFA likely due to the low heritability of this trait in particular. Correlationsbetween height and the other growth traits were the weakest. The highest correlations were calculated between wood volume and MAI (1.00), as expected, since MAI was obtained from wood volume. Growth traits did not show strong correlation with wood chemical traits. Interestingly, in some cases, the phenotypic correlations were considered non-significant, but important significant genotypic correlations were revealed. The opposite was also observed, in which significant phenotypic correlation was observed amongst volume and insoluble lignin (r_p= 0.10), but the genotypic correlations were non-significant ($r_{g_{SNP}}$= -0.03; $r_{g_{PED}}$=0.02). Between total lignin and DBH, the SNP-based genotypic correlation was not significant, while pedigree-based genotypic and

phenotypic correlations were significant ($r_{g_{PED}}$=0.09;r_p= 0.13). Significant correlations were observed between growth and wood physical traits. The strongest genotypic correlation was observed between density and DBH which were found negatively correlated ($r_{g_{SNP}}$= -0.30; $r_{g_{PED}}$=-0.26). For fiber width and coarseness no significant phenotypic correlations were observed among any growth traits, however genotypically the correlations showed strong correlations. The same behavior was observed for MFA and the other growth traits, except height.

The correlation analysis amongst wood chemical traits showed strong correlations. The negative correlations between cellulose and total lignin was smaller for phenotypic (r_p= -0.59) than for genotypic data ($r_{g_{SNP}}$= -0.67; $r_{g_{PED}}$=-0.66). Between S:G ratio and soluble lignin the SNP-based genotypic correlation was stronger than for pedigree-based and phenotypic correlations (r_p= 0.85;$r_{g_{SNP}}$= 0.92; $r_{g_{PED}}$=0.86). As expected, the higher the value for the ratio, the higher is the syringyl lignin content, which is more soluble than the guaiacyl lignin. Consequently, the correlation observed for S:G ratio and insoluble lignin was negative (r_p= -0.35;$r_{g_{SNP}}$= -0.42; $r_{g_{PED}}$= -0.38).Analyzing the correlation among wood physical traits, the highest phenotypic correlation was observed between fiber width and coarseness (r_p= 0.51), however the highest genotypic correlations were observed amongst density and coarseness ($r_{g_{SNP}}$= 0.56; $r_{g_{PED}}$= 0.45). All the other significant phenotypic correlations amongst wood physical traits were average or lower than 0.25, while genotypic correlations were higher. Analyzing the correlations between wood chemical and physical traits, significant values were observed. Density was genotypically and phenotypically correlated to all wood chemical traits except insoluble lignin, being positive for cellulose and negative for hemicelluloses, S:G ratio, soluble lignin and total lignin. In all cases, in the comparison between density and wood chemical traits, SNP-based genotypic correlations were stronger than pedigree-based correlations. For example, between density and cellulose, the SNP-based genotypic correlation was estimated at 0.31 while the pedigree-based was 0.27. MFA showed non-significant phenotypic correlations with all wood chemical traits, however the genotypic correlations were all significant, especially with insoluble lignin ($r_{g_{SNP}}$= -0.49; $r_{g_{PED}}$= -0.48). Fiber length showed SNP-based significant genotypic correlations to all wood chemical traits, however only with cellulose the correlation was positive ($r_{g_{SNP}}$= 0.34). Coarseness showed strong negative pedigree-based genotypic correlation with hemicelluloses ($r_{g_{PED}}$= -0.52), while the SNP-based genotypic correlation was average ($r_{g_{SNP}}$= -0.34).

Table 2.5 - Correlations among traits in the *Eucalyptus grandis* x *E. urophylla* breeding population. In each cell from top to bottom: $r_{g_{SNP}}$- Genotypic correlation based on SNP-based realized relationship matrix; $r_{g_{PED}}$ - Genotypic correlation based on pedigree-based numerator matrix;r_p - Phenotypic correlation (continue)

Traits	Height	Volume	MAI	Cellulose	Hemi-celluloses	S:G ratio	Insolublel ignin	Solublelig nin	Total lignin	Density	MFA	Fiberlengt h	Fiberwidt h	Coars-eness
DBH	0.61**	0.99**	0.99**	-0.14**	0.12**	0.22**	-0.02	0.19**	0.06	-0.30**	0.10**	-0.22**	0.13**	-0.09**
	0.62**	0.99**	0.99**	-0.17**	0.14**	0.14**	0.02	0.10**	0.09**	-0.26**	0.13*	-0.27**	0.25**	-0.11*
	0.61**	0.98**	0.98**	-0.10**	-0.01	0.06*	0.11**	0.04	0.13**	-0.11**	0.07	-0.15**	0.00	-0.01
Height		0.71**	0.71**	0.03	-0.02	-0.16**	0.04	-0.14**	-0.02	0.16**	-0.06	0.07*	-0.05	0.08*
		0.72**	0.72**	-0.03	0.02	-0.14**	0.10**	-0.12**	0.04	0.07*	0.02	-0.05	0.04	0.02
		0.72**	0.72**	0.02	-0.09**	-0.07*	0.05	-0.02	0.05	0.08*	0.01	0.04	-0.07	-0.01
Volume			1.00**	-0.12**	0.12**	0.17**	-0.03	0.15**	0.03	-0.26**	0.09**	-0.19**	0.11**	-0.09**
			1.00**	-0.15**	0.15**	0.11**	0.02	0.08*	0.07*	-0.24**	0.12*	-0.25**	0.23**	-0.11*
			1.00**	-0.09**	-0.02	0.04	0.10**	0.03	0.12**	-0.09**	0.06	-0.12*	-0.01	-0.02
MAI				-0.11**	0.12**	0.17**	-0.03	0.14**	0.03	-0.26**	0.09**	-0.19**	0.11**	-0.09**
				-0.15**	0.14**	0.11**	0.02	0.08*	0.07*	-0.24**	0.12*	-0.25**	0.23**	-0.11*
				-0.08**	-0.02	0.04	0.10**	0.03	0.12**	-0.09**	0.06	-0.12*	0.00	-0.01
Cellulose					-0.37**	-0.05	-0.59**	-0.08**	-0.67**	0.31**	0.30**	0.34**	0.04	0.21**
					-0.25**	-0.07*	-0.56**	-0.10**	-0.66**	0.27**	0.30**	0.30**	-0.08	0.26**
					-0.36**	0.00	-0.55**	-0.01	-0.59**	0.25**	0.05	0.34**	0.00	0.18**
Hemicelluloses						0.08*	-0.28**	0.02	-0.29**	-0.29**	0.17**	-0.19**	-0.11**	-0.34**
						0.09**	-0.37**	0.01	-0.39**	-0.28**	0.30**	-0.13*	-0.09	-0.52**
						0.02	-0.17**	-0.03	-0.19**	-0.18**	0.04	-0.24**	0.00	-0.17**
S:G ratio							-0.42**	0.92**	-0.06	-0.46**	0.23**	-0.26**	0.28**	-0.18**
							-0.38**	0.86**	-0.05	-0.40**	0.23**	-0.25**	0.24**	-0.17**
							-0.35**	0.85**	-0.05	-0.37**	0.07	-0.18**	0.07	-0.12*
Insolublelignin								-0.37**	0.92**	0.04	-0.49**	-0.16**	-0.23**	-0.05
								-0.34**	0.92**	0.04	-0.48**	-0.15**	-0.21**	-0.08
								-0.34**	0.93**	-0.05	-0.08	-0.23**	-0.14**	-0.16**
Solublelignin									0.02	-0.34**	0.23**	-0.19**	0.32**	-0.04
									0.03	-0.28**	0.18**	-0.10	0.29**	0.04
									0.03	-0.26**	0.06	-0.05	0.12*	0.00
Total lignin										-0.10**	-0.43**	-0.26**	-0.12**	-0.07*
										-0.08**	-0.43**	-0.22**	-0.07	-0.06
										-0.16**	-0.07	-0.26**	-0.11*	-0.17**

Table 2.5 - Correlations among traits in the *Eucalyptus grandis* x *E. urophylla* breeding population. In each cell from top to bottom: $r_{g_{SNP}}$- Genotypic correlation based on SNP-based realized relationship matrix; $r_{g_{PED}}$ - Genotypic correlation based on pedigree-based numerator matrix; r_p - Phenotypic correlation (conclusion)

Traits	Height	Volume	MAI	Cellulose	Hemi-celluloses	S:G ratio	Insolublel ignin	Solublelig nin	Total lignin	Density	MFA	Fiberlengt h	Fiberwidt h	Coars-eness
Density											-0.07*	0.41**	-0.11**	0.56**
											-0.13*	0.36**	-0.25**	0.45**
											-0.03	0.24**	-0.20**	0.21**
MFA												0.02	0.22**	0.08*
												-0.01	0.14**	-0.06
												-0.01	0.05	0.11*
Fiberlength													0.07*	0.35**
													-0.08	0.25**
													0.03	0.23**
Fiberwidth														0.47**
														0.43**
														0.51**

Significance of the correlation indicated as: * p value≤0.05; ** p value ≤0.01

The correlations amongst the phenotypic observations and breeding values were also calculated (table 2.6). The correlation between observed phenotype and SNP-based breeding values varied from 0.59 (MFA) to 0.97 (S:G ratio) and wood chemical traits displayed the highest correlations. Similarly, the correlations between the observed phenotypes and pedigree-based breeding values revealed higher correlations varying from 0.63 to 1.00, for MFA and S:G ratio, respectively. Finally, the correlations amongst SNP-based and pedigree-based breeding values were high, ranging from 0.86 (fiber width) to 0.97 (S:G ratio).

Table 2.6 – Correlations amongst breeding values derived from observed phenotypes (y), with breeding values estimated from pedigree information ($\hat{y}_{PED}$) and SNP data ($\hat{y}_{SNP}$)

Trait	$r_{y\hat{y}_{SNP}}$	$r_{y\hat{y}_{PED}}$	$r_{\hat{y}_{SNP}\hat{y}_{PED}}$
DBH (cm)	0.76	0.88	0.90
Plant height (m)	0.70	0.81	0.88
Wood Volume (m^3)	0.76	0.88	0.90
MAI ($m^3.ha^{-1}.year$)	0.76	0.88	0.90
Cellulose (%)[1]	0.83	0.80	0.87
Hemicelluloses (%)[1]	0.88	0.79	0.88
S:G ratio[1]	0.97	1.00	0.97
Insoluble lignin (%)[1]	0.89	0.93	0.94
Soluble lignin (%)[1]	0.92	1.00	0.93
Total lignin (%)[1]	0.89	0.92	0.94
Density ($kg.m^{-3}$)[1]	0.86	0.97	0.92
MFA (°)	0.59	0.63	0.90
Fiber length (mm)	0.91	0.89	0.94
Fiber width (µm)	0.63	0.67	0.86
Coarseness ($mg.100m^{-1}$)	0.73	0.73	0.90

2.2.3 Discussion

To the best of our knowledge this study constitutes the first comprehensive public report of wood properties traits assessment in a large elite breeding population of the widely planted *Eucalyptus grandis x E. urophylla* hybrid. Besides growth data, phenotype measurements for 11 wood properties traits were taken by wet chemistry analyses on a subset

of 200 to 350 most representative trees based on Euclidean distances of near-infrared reflectance spectra, so as to maximize variation of the calibration population. This subset of samples was used as training set by which near-infrared calibration models were constructed and subsequently used to estimate trait values for the remaining 650 trees. It is also the first study in the genus *Eucalyptus* to employ high-precision genome-wide SNP-based relatedness as a substitute of the conventional pedigree based relationships for genetic parameter estimation. The extensive genome sharing observed between individuals in different families resulted in levels of genetic relatedness originally unexpected from pedigree data alone. Considerably more accurate estimates of trait variance components, heritabilities, genetic and phenotypic correlations for the target traits were therefore obtained that should be valuable to inform improved breeding decisions.

2.2.3.1 Wood properties phenotyping

A key finding of this study is the wide genetic variation in wood chemical and physical composition observed in this hybrid breeding population (table 2.4) Noteworthy is the fact that trees with S:G ratio up to 4.2 were observed, a very interesting result showing great potential for selection of high pulp yielding trees even in this tropical *E. grandis x E. urophylla* background. Wide variation was also observed for total lignin content, as low as 24%, wood density, as high as 646.5 kg/m^3 and growth, which currently constitute the main targets of selection in breeding programs. With this study, the availability of data for a much wider array of traits should promote a better exploitation of the existing variation for the additional traits, therefore improving overall gains by the identification of elite trees that consolidate superior trait combinations. Our study also demonstrates that a better use of the existing phenotypic variation in breeding populations depends heavily on fast and efficient ways to collect data for wood properties traits for large numbers of samples in progeny trials. To that end, however, direct wet chemistry methods are usually not an option. In spite of the development of faster analytical methods for some traits, such procedures are still very laborious and expensive. Until a few years ago in most *Eucalyptus* breeding programs, density was the only wood property trait effectively measured, given its direct impact on final traits such as pulp yield or charcoal calorific power. Additional wood properties traits such as lignin and cellulose content and composition were measured on a very limited set of trees that had passed established selection thresholds for volume growth and density, therefore precluding a wider sampling of the available variation for such traits. Near-infrared reflectance

spectroscopy (NIRS) applied to wood properties in *Eucalyptus* has radically changed this scenario (SCHIMLECK et al., 2000). NIRS allows the prediction of the wood traits by non-destructive sampling of very large numbers of samples. In our study, better correlations were observed between actual and predicted phenotypes for wood chemical traits in comparison to wood physical traits, and in particular for lignin related traits (table 2.3). As reflectance data results from the vibration of chemical bonds present in the wood, such better predictive ability was in fact expected. NIRS-predictions for cellulose and hemicelluloses, were less precise compared to the other wood chemical traits. This result might be partly due to the smaller calibration set used for these traits, 200 instead of the 350 samples, although other issues such as not yet optimized wet chemistry protocolsfor the tropical *Eucalyptus* wood cannot be ruled out. For the temperate species *E. globulus*, NIRS correlations for cellulose content varied between 0.82 and 0.94 in different populations (RAYMOND; SCHIMLECK, 2002), considerably higher than the 0.57 obtained in our study. For the other wood chemical traits related to lignin content, calibrations showed good correlations, varying between 0.83 and 0.93 (table 2.3). A relatively high correlation was observed for wood density (r = 0.74), a physical trait generally considered a difficult one to predict using NIRS, although Hein et al. (2009) reported coefficients of determination between 0.74 and 0.86 for *E. grandis* and *E. urophylla* at age 6.5 years. This result can possibly be explained by the positive and significant correlation observed between density and wood chemical traits such as cellulose, hemicelluloses, S:G ratio, soluble lignin and total lignin. Calibrations for the other wood physical traits had low correlations and the phenotypes could not be predictedaccurately for the remainingindividuals in the population. Correlations between the other physical traits (MFA, fiber length, fiber width and coarseness) and wood chemical traits were low (table 2.5). Our results are generally in line with those reported by Denis et al. (2013) that also showed generally low to moderate correlation values amongst chemical and physical traits, with the highest genetic correlations seen for traits that shared the same constituents such as S and G lignin and S:G ratio.

2.2.3.2 Genetic parameters estimation using a genome-wide SNP-based realized relationship matrix

Although marker data in breeding practice is usually envisaged in the context of QTL mapping, marker assisted selection and, more recently, genomic selection (GRATTAPAGLIA; RESENDE, 2011), markers can be very useful to provide more accurate

estimates of genetic parameters, as shown so far for animal breeding (HAYES; VISSCHER; GODDARD, 2009; LEE et al., 2010). When phenotypes are collected on a sample of individuals whose relatedness is partially or totally unknown, markers allow inferring relatedness between pairs of individuals, because related individuals tend to share more marker alleles than unrelated individuals. The inferred relatedness can then be correlated with phenotypic similarity, and quantitative genetic parameters estimated. This method has been applied in evolutionary studies to estimate heritability for quantitative traits when phenotypes and DNA samples are available but pedigree information is not. A recent study summarized studies comparing estimates of quantitative genetic parameters using pedigree-record-free estimates. The vast majority was carriedout in animals, while te few ones in plants were done using a very limited set of 54 to 20 microsatellites (GAY; SIOL; RONFORT, 2013). In poplar, however, a recent study used a dataset of ~9,000 SNPs to estimate heritabilities for wood properties traits based on a genomic relationship matrix amongst 334 individuals sampled across natural populations spanning a wide geographical distribution (PORTH et al., 2013).

To the best of our knowledgeour study is the firstone inplants in general to compare estimates of genetic parameters obtained by using a high density SNP based genomic-based realized relationship matrix,with those generated by a standard pedigree relationship matrix, showingthe additional benefit that genomic data can bring to a breeding program. The proportion of the genome that is shared identical by descent (IBD) varies around its expectation for pairs of relatives because of the random nature of segregation and recombination. Segregation causes variation in the actual number of alleles shared IBD between relatives. For example, for full sib pairs at a single locus, 25% of all the sib pairs share no alleles IBD, 25% share two alleles IBD and the remaining 50% share one allele IBD. SNP markers genotyped at high density therefore allowed capturing all this variation among all types of relatives existing in our target population, including ancestral relatedness that the pedigree data cannot account for. It is possible, for example, that some of the parents used to generate the full sib progenies in the trial are in fact half-sibs or even full-sibs from original provenance/progeny introductions with seed sources that came from Australia. While our pedigree data is agnostic of such possibilities, SNP marker data actually detects these facts and embeds them in the estimates of variance and co-variancesamong relatives, therefore providing more accurate estimates of genetic parameters.

Estimated heritabilities based on a genomic realized relationship matrix varied from a low 0.11, for fiber width, to a high of 0.85 for S:G ratio, while estimates based on pedigree data were slightly higher for most traits, varying from 0.13 to 1.00. Chemical composition traits were the most heritable ones, irrespective of being it a pedigree-based or SNP-based estimate (table 2.4). High pedigree-based heritabilities for lignin related traits have also been reported for *E. urophylla*(DENIS et al., 2013) and *E. globulus*(STACKPOLE et al., 2011), possibly explained by a relatively simpler genetic control involving a relatively smaller number of loci when compared to growth traits. In fact, five QTL were mapped in *E. globulus* for S:G ratio, and they were mostly stable across families and sites (FREEMAN et al., 2013), adding evidence to a putative oligogenic control. Generally, low to moderate heritabilities have been observed for growth and wood physical traits, consistent with a more complex trait architecture.Fiber length and wood density showed the highest heritabilities, (0.46 and 0.63) and (0.75 and 0.58) respectively for pedigree-based and genomic-based estimates. Our estimates for wood density are consistent with previous studies in*E. globulus*(0.51)(STACKPOLE et al., 2011) and *E. urophylla* (0.53) (DENIS et al., 2013). In our study, growth traits displayed genotypic and phenotypic correlations of the same magnitude and sign, revealing that the environment likely affects these traits in the same way. In addition, the strong correlations between DBH and wood volume (0.99) and a relatively weaker correlation between plant height and wood volume (0.72) indicates that DBH should be chosen as abetter proxy over height for total tree volume assessment. A similar situation happened between S:G ratio and soluble lignin, that have strong positive correlation, coherent with other studies (DENIS et al., 2013). Unsurprisingly, strong negative correlations were observed between cellulose and total lignin ($r_{g_{SNP}}$= -0.67; $r_{g_{PED}}$= -0.66). This corroborates the fact that by selecting trees with higher cellulose content, positive selection for a lower lignin content is also achieved, a particularly useful situation of programs breeding treesfor pulp and paper. Similar but even stronger negative correlations between cellulose and lignin content were estimated in *E. globulus*, with pedigree-based genotypic correlation reaching 0.90 (STACKPOLE et al., 2011).Correlations between wood physical traits were stronger for genotypic than phenotypic data, revealing that the breeder should be in a good position to simultaneously select for these traits in any desired direction, with the advantage of using SNP data to better asses response to selection. Growth and wood chemical trait correlations were generally not strong, although some genetic correlations stand out. The weak correlations suggest that these groups of traits are likely controlled by different genetic and physiological mechanisms. In fact no significant genotypic correlations were found between

volume growth and lignin content, although a weak positive correlation of 0.12 was found, consistent with a 0.10 phenotypic correlation also found in *E. globulus*(STACKPOLE et al., 2011). These result contrast with earlier findings of strong negative correlations between these two traits in eucalypts and poplars and the complex biochemical modelling proposed thereof (NOVAES et al., 2010). This difference could be readily explained by the fact that those results were obtained in a single biparental family (i.e. a very limited genetic representation of the existing variation in these traits) and in very young developmental stages of wood formation (20 month old for *Eucalyptus* and 5 week old for poplars).Growth traits were also not strongly correlated to wood physical traits. Exception made for wood density, which had the highest heritability among the physical traits, and showed average genotypic correlations with all growth traits. S:G ratio was also genotypically correlated to all wood physical traits, revealing the impact of this trait on wood formation. Wood density and S:G ratio showed a significant and relatively high negative correlation ($r_{g_{SNP}}$= -0.46; $r_{g_{PED}}$= -0.40,) consistent with previous reports in *E. globulus* with an additive genetic correlation at the population level of -0.28 (STACKPOLE et al., 2011),-0.22 in a biparental segregating family of *E. grandis* and *E. urophylla*(GION et al., 2011) and -0.1 in a*E. urophylla*breeding population (DENIS et al., 2013). This is a noteworthy result for breeding eucalypts for pulp and paper, indicating that selecting for elite clones with an average wood density between 500 and 550 kg/m3, that results in high cellulose content, combined with high S:G ratio for easier pulping, is a feasible target. MFA had no significant phenotypic correlations with any wood chemical traits, but showed significant genotypic correlationswiththem.

2.2.3.3 Implications for *Eucalyptus* breeding

One of the main contributions of this study is the demonstration that substantial genetic variation in wood chemical and physical composition exists in a typical *E. grandis*x *E. urophylla* hybrid breeding population (table 2.4). Wide ranges of variation in essentially all traits were seen,with high values of volume growth, S:G ratio, cellulose content and wood density, key traits when breeding for pulp, fiber and energy. These results should foster an improved utilization of the existing variation in this and similarly formatted populations. The results of our study should therefore be directly useful to all *Eucalyptus* breeding programs that exploit this widely planted hybrid combination. We applied NIRS based phenotype prediction tools together with the power of high-density SNP marker data to estimate genetic

parameters for a number of traits that are widely relevant to the forest based industry. We used a large set of individuals from several families, representative of a standard breeding population. The need of large data sets with a representative number of crosses and parents for genetic parameters estimations is a key aspect to avoid unwanted variation in the estimates (BOUVET; SAYA; VIGNERON, 2009).

Our study expands the understanding on the genetic architecture of wood chemical and physical properties in *Eucalyptus*, beyond the relatively few existing studies that so far have mostly been carried out with temperate*E. globulus* and *E. nitens* populations (RAYMOND, 2002; APIOLAZA; RAYMOND; YEO, 2005; VOLKER; POTTS; BORRALHO, 2008; HAMILTON et al., 2009). The genetic variation observed for all traits studied in this hybrid population clearly show that tropical*Eucalyptus* breeding programs are far from a plateau in gains. Considerable gains are still achievable and wood properties offer extensive opportunities for improvement. Estimated heritabilities were average to high for most traits suggesting that continued gains should be possible for several generations ahead before heritability might decay due a decrease in available genetic variance (FALCONER; MACKAY, 1996).Most heritabilities estimated based on pedigree information were slightly higher than the SNP-based heritabilities, exception made for cellulose. This indicatesthat the use of a SNP-based relationship matrix can better separate genetic and non-genetic causes of variation and thus avoid overestimation of heritabilities, a result that has a direct impact on breeding practice in eucalypts in particular and likely for trees in general. This was the case in a study in mice where the use of a 10,000 SNP-based realized relationship matrix provided more accurate and less biased estimation of variance components(LEE et al., 2010) than pedigree-based heritabilities. Moreover, the use of SNP data revealed important correlations and non-correlations between pairs of traits that pedigree alone would not have shown, or in some cases had them overestimated or underestimated. The noise of non-genetic information is excluded when SNP markers are used to determine relatedness which becomes independent from historical records. Low but significant and negative correlations (-0.26 to -0.30) were observed between growth traits and wood density, opposite to previous reports that showed positive correlation of 0.15 between circumference and density in *E. urophylla*(DENIS et al., 2013) or an absence of correlation in *E. globulus*(STACKPOLE et al., 2011). Strong negative correlations were observed between cellulose and total lignin, confirming the balance metabolism of the traits and in accordance with previous studies (STACKPOLE et al., 2011; DENIS et al., 2013).

The phenotypic and genetic correlations estimated in this study constitute important guidelines to understanding the impact of indirect selection and correlated response to selection. More accurate estimates of all genetic parameters were obtained by using a realized relationship matrix, showing that SNP data can be wisely used as an aid to quantitative genetics beyond its commonly envisaged applications in genomics, association genetics and genomic prediction.

2.3 Conclusions

Convergence of quantitative genetics and genomics is rapidly becoming the way that fundamental genetics and applied breeding will be carried out in the next decades. The results of this study provide an example of how to interface the quantitative genetics of complex growth and wood properties traits with SNP based data towards a modern approach to eucalypt breeding. *Eucalyptus* breeding practice can be enhanced by combining high-throughput phenotyping, genome-wide marker data and more accurate genetic parameters estimation using SNP information. The variability observed in the hybrid population studied denotes how far *Eucalyptus* breeding programs are from inertia. Great gains are still possible, for growth and wood quality traits and genotypic correlations indicate that options for selection in different directions are vast. NIRS prediction of wood chemical traits was shown to be a definite advantage to reduce time and costs of phenotyping. The use of SNP data for genetic parameters estimations is clearly a good alternative and should become standard practice with the continuous dropping in costs of genotyping technologies. Modern tree breeding will move into using DNA data routinely for genetic management of populations and phenotype prediction of complex traits, significantly reducing the length of breeding cycles and, consequently, optimizing resource allocation and sustainability of the overall breeding program.

References

APIOLAZA, L.A.; RAYMOND, C.A.; YEO, B.J. Genetic variation of physical and chemical wood properties of *Eucalyptus globulus*. **Silvae Genetica**, Frankfurt, v. 54, n. 4/5, p. 160-166, 2005.

ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS.**Statistical yearbook 2013**: base year 2012. Brasília, 2013. 148 p.

BOUVET, J.M.; SAYA, A.; VIGNERON, P. Trends in additive, dominance and environmental effects with age for growth traits in Eucalyptus hybrid populations. **Euphytica**, Dordrecht, v. 165, n. 1, p. 35-54, 2009.

BUSH, D.; KAIN, D.; MATHESON, C.; KANOWSKI, P. Marker-based adjustment of the additive relationship matrix for estimation of genetic parameters-an example using *Eucalyptus cladocalyx*. **Tree Genetics & Genomes**, Heidelberg, v. 7, n. 1, p. 23-35, 2011.

COSTA E SILVA, J.; BORRALHO, N.; ARAÚJO, J.; VAILLANCOURT, R.; POTTS, B. Genetic parameters for growth, wood density and pulp yield in *Eucalyptus globulus*. **Tree Genetics & Genomes**, Heidelberg, v. 5, n. 2, p. 291-305, 2009.

DENIS, M.; FAVREAU, B.; UENO, S.; CAMUS-KULANDAIVELU, L.; CHAIX, G.; GION, J.M.; NOURRISIER-MOUNTOU, S.; POLIDORI, J.; BOUVET, J.M. Genetic variation of wood chemical traits and association with underlying genes in *Eucalyptus urophylla*. **Tree Genetics & Genomes**, Heidelberg, v. 9, n. 4, p. 927-942, 2013.

ENDELMAN, J.B. Ridge regression and other kernels for genomic selection with R Package rrBLUP. **Plant Genome**, Madison, v. 4, n. 3, p. 250-255, 2011.

ENDELMAN, J.B.; JANNINK, J.L. Shrinkage estimation of the realized relationship matrix.**G3: Genes, Genomes, Genetics**, Bethesda, v. 2, n. 11, p. 1405-1413, 2012.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th ed. Englewwod Cliffs: Pearson Prentice Hall, 1996. 464 p.

FREEMAN, J.S.; POTTS, B.M.; DOWNES, G.M.; PILBEAM, D.; THAVAMANIKUMAR, S.; VAILLANCOURT, R.E. Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in *Eucalyptus globulus.* **New Phytologist**, Hoboken, v. 198, n. 4, p. 1121-1134,2013.

GAY, L.; SIOL, M.; RONFORT, J. Pedigree-free estimates of heritability in the wild: promising prospects for selfing populations. **PLoS One,** San Francisco, v. 8, n. 6, p. e66983, 2013.

GION, J.M.; CAROUCHE, A.; DEWEER, S.; BEDON, F.; PICHAVANT, F.; CHARPENTIER, J.P.; BAILLERES, H.; ROZENBERG, P.; CAROCHA, V.; OGNOUABI, N.; VERHAEGEN, D.; GRIMA-PETTENATI, J.; VIGNERON, P.; PLOMION, C. Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: Eucalyptus. **BMC Genomics**, London, v. 12, n. 301, p. 1-19, 2011.

GRATTAPAGLIA, D.; RESENDE, M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, Heidelberg, v. 7, n. 2, p. 241-255, 2011.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic-linkage maps of*Eucalyptus-grandis* and *Eucalyptus-urophylla*using a pseudo-testcross - mapping strategy and Rapd markers. **Genetics**, Bethesda, v. 137, n. 4, p. 1121-1137, 1994.

GRATTAPAGLIA, D.; SILVA, O.B.; KIRST, M.; DE LIMA, B.M.; FARIA, D.A.; PAPPAS, G. J. High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. **Bmc Plant Biology**, London, v. 11, n. 65, p. 2-18,2011.

GRATTAPAGLIA, D.; VAILLANCOURT, R. E.; SHEPHERD, M.; THUMMA, B.R.; FOLEY, W.; KULHEIM, C.; POTTS, B.M.; MYBURG, A.A. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. **Tree Genetics & Genomes**, Heidelberg, v. 8, n. 3, p. 463-508, 2012.

HAMILTON, M.G.; RAYMOND, C.A.; HARWOOD, C.E.; POTTS, B.M. Genetic variation in *Eucalyptus nitens* pulpwood and wood shrinkage traits. **Tree Genetics & Genomes**, Heidelberg, v. 5, n. 2, p. 307-316, 2009.

HAYES, B.J.; VISSCHER, P.M.; GODDARD, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, New York, v. 91, n. 1, p. 47-60, 2009.

HEIN, P.R.G.; CAMPOS, A.C.M.; TRUGILHO, P.F.; LIMA, J.T.; CHAIX, G. Near infrared spectroscopy for estimating wood basic density in *Eucalyptus urophylla* and *Eucalyptus grandis*. **Cerne**, Lavras, v. 15, n. 2, p. 133-141, 2009.

HEIN, P.R.G.; BOUVET, J.M.; MANDROU, E.; VIGNERON, P.; CLAIR, B.; CHAIX, G. Age trends of microfibril angle inheritance and their genetic and environmental correlations with growth, density and chemical properties in *Eucalyptus urophylla* ST Blake wood. **Annals of Forest Science**, Paris, v. 69, n. 6, p. 681-691, 2012.

HENDERSON, C.R. Estimation of variances and covariances under multiple trait models. **Journal of Dairy Science**, New York, v. 67, n. 7, p. 1581-1589, 1984.

HUNTLEY, S.K.; ELLIS, D.; GILBERT, M.; CHAPPLE, C.; MANSFIELD, S. D. Significant increases in pulping efficiency in C4H-F5H-transformed poplars: Improved chemical savings and reduced environmental toxins. **Journal of Agricultural and Food Chemistry**, Washington, v. 51, n. 21, p. 6178-6183, 2003.

KENNARD, R.W.; STONE, L.A. Computer aided design of experiments. **Technometrics**, Alexandria, v. 11, n. 1, p. 137-148, 1969.

KORECKÝ, J.; KLÁPŠTĚ, J.; LSTIBŮREK, M.; KOBLIHA, J.; NELSON, C.D.; EL-KASSABY, Y. Comparison of genetic parameters from marker-based relationship, sibship, and combined models in Scots pine multi-site open-pollinated tests. **Tree Genetics & Genomes**, Heidelberg, v. 9, n. 5, p. 1227-1235, 2013.

LEE, S.H.; GODDARD, M.E.; VISSCHER, P.M.; VAN DER WERF, J.H.J. Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. **Genetics Selection Evolution**, London, v. 42, n. 22, p. 1-14,2010.

MARANAN, M.C.; LABORIE, M.P.G. Analysis of energy traits of *Populus spp.* clones near-infrared spectroscopy. **Journal of Biobased Materials and Bioenergy**, Valencia, v. 1, n. 1, p. 155-162, 2007.

NOVAES, E.; KIRST, M.; CHIANG, V.; WINTER-SEDEROFF, H.; SEDEROFF, R. Lignin and biomass: a negative correlation for wood formation and lignin content in trees. **Plant Physiology**, Rockville, v. 154, n. 2, p. 555-561, 2010.
POKE, F.S.; POTTS, B.M.; VAILLANCOURT, R.E.; RAYMOND, C.A. Genetic parameters for lignin, extractives and decay in *Eucalyptus globulus.***Annals of Forest Science**, Paris, v. 63, n. 8, p. 813-821, 2006.

PORTH, I.; KLAPSTE, J.; SKYBA, O.; LAI, B.S.K.; GERALDES, A.; MUCHERO, W.; TUSKAN, G.A.; DOUGLAS, C.J.; EL-KASSABY, Y.A.; MANSFIELD, S.D. *Populus trichocarpa* cell wall chemistry and ultrastructure trait variation, genetic control and genetic correlations. **New Phytologist**, Hoboken, v. 197, n. 3, p. 777-790, 2013.

RAYMOND, C.A. Genetics of *Eucalyptus* wood properties. **Annals of Forest Science**, Paris, v. 59, n. 5/6, p. 525-531, 2002.

RAYMOND, C.A.; SCHIMLECK, L.R. Development of near infrared reflectance analysis calibrations for estimating genetic parameters for cellulose content in *Eucalyptus globulus*. **Canadian Journal of Forest Research/Revue Canadienne De Recherche Forestiere**, Ottawa, v. 32, n. 1, p. 170-176, 2002.

RAYMOND, C.R.; APIOLAZA, L.A. Incorporating wood quality and deployment traits in *Eucalyptus globulus* and *Eucalyptus nitens*. In: CARSON, M. (Ed.). **Plantation forest biotechnology for the 21st century**. Rotorua: Forest Research New Zealand, 2004. p. 87-99.

RESENDE, M.D.V.; OLIVEIRA, E.B. Sistema Selegen:seleção genética computadorizada para o melhoramento de espécies perenes. **Pesquisa Agropecuaria Brasileira**, Brasília, v. 32, n. 9, p. 931-939, 1997.

ROBINSON, A.R.; MANSFIELD, S.D. Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. **Plant Journal**, Hoboken, v. 58, n. 4, p. 706-714, 2009.

SANSALONI, C.P.; PETROLI, C.D.; CARLING, J.; HUDSON, C.J.; STEANE, D.A.; MYBURG, A.A.; GRATTAPAGLIA, D.; VAILLANCOURT, R.E.; KILIAN, A. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. **Plant Methods**, London, v. 6, p. 16, 2010.

SCHIMLECK, L.R.; SUSSENBACH, E.; LEAF, G.; JONES, P.D.; HUANG, C.L. Microfibril angle prediction of *Pinus taeda* wood samples based on tangential face NIR spectra. **Iawa Journal**, Leiden, v. 28, n. 1, p. 1-12, 2007.

SCHIMLECK, L.R.; RAYMOND, C.A.; BEADLE, C.L.; DOWNES, G.M.; KUBE, P.D.; FRENCH, J. Applications of NIR spectroscopy to forest research. **Journal of the Technical Association of the Australian and New Zealand Pulp and Paper Industry**, Bundoora,v. 53, p. 458-464, 2000.

SILVA-JUNIOR, O.B.; FARIA, D.A.; TOGAWA, R.C.; GRATTAPAGLIA, D. *Eucalyptus* genotyping taken to the next level: development of the " EucHIP60k.br" based on large scale multi-species SNP discovery and ascertainment. In: IUFRO TREE BIOTECHNOLOGY CONFERENCE 2013, 2013, Asheville.**Proceedings…**2013.

STACKPOLE, D.J.; VAILLANCOURT, R.E.; ALVES, A.; RODRIGUES, J.; POTTS, B.M. Genetic variation in the chemical components of *Eucalyptus globulus*wood. **G3: Genes, Genomes, Genetics**, Bethesda, v. 1, n. 2, p. 151-159, 2011.

STIRLING, R.; TRUNG, T.; BREUIL, C.; BICHO, P. Predicting wood decay and density using NIR spectroscopy. **Wood and Fiber Science**, Madison, v. 39, n. 3, p. 414-423, 2007.

VANRADEN, P.M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, New York, v. 91, n. 11, p. 4414-4423, 2008.

VOLKER, P.; POTTS, B.; BORRALHO, N. Genetic parameters of intra- and inter-specific hybrids of *Eucalyptus globulus* and *E. nitens*. **Tree Genetics & Genomes**, Heidelberg, v. 4, n. 3, p. 445-460, 2008.

3 GENOME-WIDE PREDICTION OF GROWTH, CHEMICAL AND PHYSICAL WOOD PROPERTIES IN *Eucalyptus* USING A HIGH-DENSITY 60K SNP CHIP

Abstract

Notwithstanding their rapid growth rates, breeding cycles of tropical *Eucalyptus* generally take 12 to 18 years to deliver clonally tested elite genotypes. While growth traits are measured in all individuals of a progeny trial, the assessment of wood chemical and physical properties is usually carried out in a considerably smaller number of trees in the late stages of the selection process. This procedure precludes taking advantage of the full range of existing genetic variation in wood properties known to have a significant impact on the downstream uses of the wood. Genomic selection (GS) have been proposed to predict genomic breeding values for complex traits based on molecular marker data. This approach aimsat shortening breeding cycles, accelerating selections and increasing gain per unit time. In our study, we built genomic predictive models for fifteen growth and wood chemical and physical properties of*Eucalyptus* using a training population of 999 individuals sampled in a progeny trial of an elite breeding population. Individuals were genotyped with a high-density genotyping chip containing 60,639 SNPs. A total of 29,090 SNPs were retained for prediction analyses following allele frequency and call rate filters, providing one SNP every 22 kbp in the *Eucalyptus* reference genome, a significantly higher marker density than any previous study in forest trees. Two analytical approaches that differ regarding the assumed distribution of marker effects were used for genomic predictions, which were compared to those obtained by standard BLUP phenotypic selection. Equivalent narrow-sense heritabilities were obtained with the two models, varying from 0.22 to 0.93 by RR-BLUP and from 0.15 to 0.93 by BLASSO. Wood chemical traits displayed the highest heritabilities and wood physical traits the lowest. Predictive abilities were strongly correlated to the heritabilities and reached very similar estimates with the two prediction models, varying from a low of 0.10, for MFA, to 0.42 for volume growth, and up to 0.83, for S:G ratio using RR-BLUP. Spearman's correlations were close to unity between the genomic prediction models, and between 0.771 and 0.929 between them and phenotypic BLUP prediction. Accurate genomic-enabled predictions were obtained for wood chemical traits particularly related to lignin content and composition, wood density and growth, although generally 15 to 25% lower than those achieved by phenotypic BLUP prediction. Both genomic prediction models yielded a coincidence>70% for the top 30 trees ranked by phenotypic selection for volume growth, wood density and S:G ratio, and >60% the top 10 trees. When tandem multi-trait selection was applied to these three traits simultaneously, 15 out of the top 25 trees (60%) selected by BLUP phenotypic selection were also selected by genomic selection. These results support earlier propositions by which genomic selection could significantly reduce the length of a breeding cycle in *Eucalyptus* by applying ultra-early selection of genomically top ranked seedlings therefore precluding the progeny trial stage.

Keywords: Marker-assisted selection; Molecular breeding; Genome-wide selection; Plant breeding

3.1 Introduction

Fast growing *Eucalyptus* plantations cover more than 5 million hectares in Brazil, a country with worldwide unparalleled forest potential, prone to become one of the main suppliers of high quality renewable wood products in the next twenty years, given the rapid expansion of planted forest in underutilized or abandoned pasture land (ABRAF, 2013). Native to Australia and adjacent islands (LADIGES; UDOVICIC; NELSON, 2003), several eucalypt species (*Eucalyptus* spp.) found in Brazil optimal conditions for their development, achieving some of the highest forest productivities in the world (ELDRIDGE et al., 1993).However, despite their fast growth in the tropics, *Eucalyptus* breeding cycles generally take 12 to 18 years to deliver clonally tested elite genotypes. While growth traits are measured in all trees of a progeny trial, the assessment of detailed wood chemical and physical properties is usually carried out in a considerably smaller number of trees during clonal trials in the late stages of the breeding cycle. Such a procedure precludes taking advantage of the full range of existing genetic variation in wood properties that have a significant impact on the industrial uses of the forest.

The advent of molecular markers, generated great promises for tree breeding in general and eucalypts in particular, especially for accelerating selection for complex quantitative traits This expectation fostered a large number of QTL studies in *Eucalyptus* and some attempts to use this information for breeding practice (GRATTAPAGLIA et al., 2012). Still, despite the discovery of numerous QTLs and some genes for several growth and wood properties, the information has not effectively been used in applied tree breeding. QTL mapping studies when envisaging actual use of the resulting information for tree breeding, suffer from a number of drawbacks (GRATTAPAGLIA; RESENDE, 2011). Only a small proportion of QTLs underlying the target trait are detected given the low power of biparental mapping populations,therefore explaining limited portions of the observed variation. Estimates of the variance explained by the QTLs detected are usually overestimated given the Beavis effect (BEAVIS, 1998). The limited genetic representativeness of the single or even the few biparental populations used, makes the detected QTLs highly susceptible to erratic behaviors, also known as QTL by genetic background interaction. Association genetics studies in more representative populations have been also reported in eucalypts, an attempt to provide population-wide markers based on candidate genes(THUMMA et al., 2005; MANDROU et al., 2012; DENIS et al., 2013), and more recently on a low-density GWAS approach using 7,680 DArT markers (CAPPA et al., 2013). Although a few marker-trait associations have been found, no independent validation has yet been reported and, more

importantly, the magnitude of effects estimated are usually less than 5% leaving a very large proportion of the variation unaccounted for, i.e. the ‘missing heritability’ dilemma (MAKOWSKY et al., 2011).

Genome-wide selection (GWS) or genomic selection (GS) (MEUWISSEN; HAYES; GODDARD, 2001), on the other hand, has been proposed and applied as a breeding approach to predict phenotypic performance based only on a genome-wide panel of markers whose effects on the phenotype are estimated simultaneously in a large and representative population of individuals without applying rigorous significance tests. In GS a marker effect does not need to exceed a stringent significance threshold to be used in the subsequent breeding phase, and the effects of the marker alleles are estimated in a larger population rather than within one or a few mapping populations. GS therefore works on the principle that linkage disequilibrium (LD), provided by dense genotyping, is sufficient to track all relevant QTL effects for the target trait which are expected to be in LD with at least some of the queried markers (GRATTAPAGLIA, 2014). Successfully implemented in domestic animals(HAYES et al., 2009; GODDARD; HAYES; MEUWISSEN, 2010), the technique has finally realized the promise of MAS mainly in dairy cattle by reducing time and cost by early selection of elite sires based on their estimated breeding values. Based on a training population of individuals, for which one has phenotypes and genotypes, a statistical model is "trained" and estimates of genome-wide marker effects are used to predict phenotypes for individuals for which only genotypes are available (GODDARD; HAYES, 2007). For example, in countries like the USA, Canada and Ireland, genomically tested bulls for which daughters were not yet milking, make up over 40%, of the market-share of elite sires, with gains between 28 and 108%, when compared to the conventional progeny testing breeding route (PRYCE; DAETWYLER, 2012).

The success of GS in animal breeding stimulated the consideration of this breeding approach in plants (BERNARDO, 2008; GRATTAPAGLIA et al., 2009; HEFFNER; SORRELLS; JANNINK, 2009;LORENZ et al., 2011). The main objective in plants has been the reduction of breeding cycles length, therefore increasing gain per unit time and reducing the overall cost of the program (HEFFNER et al., 2010). Besides a number of encouraging simulation studies of the potential of genomic selection in crop plants, reports of experimental results of GS have recently been published for maize and wheat (POLAND et al., 2012; CROSSA et al., 2013; MASSMAN et al., 2013; MASSMAN; JUNG; BERNARDO, 2013; CROSSA et al., 2014) and sugarcane (GOUY et al., 2013). In forest trees, the optimistic prospects of GS to accelerate breeding cycles were initially detailed in a simulation study

looking at the impact of linkage disequilibrium (modeled by variable effective population size and inter-marker distance), the size of the training set, trait heritability, and the number of QTL on the predicted accuracy of GS (GRATTAPAGLIA; RESENDE, 2011; RESENDE, M. D. V. et al., 2012). Experimental studies in loblolly pine and *Eucalyptus* soon followed corroborating the positive outlook of this molecular breeding approach to accelerate breeding cycles. Results in a loblolly pine (*Pinustaeda* L.) population, indicated that good predictive accuracies could be achieved for growth using ~4,800 SNPs, as long as models are used at the relevant selection age and within the breeding zone in which they were estimated.(RESENDE, M. F., JR. et al., 2012). In a follow-up study in the same loblolly pine population, different statistical methods provided only marginally different accuracies for most method/trait combinations(RESENDE, M. F. R. et al., 2012).

In *Eucalyptus*, GS predictive models built for growth, wood density and pulp yield using a set of ~3000 dominant DArT markers, reached accuracies as good as those provided by conventional phenotypic selection. GS was modeled in two different populations with contrasting effective population sizes, showing higher accuracies in the population with smaller N_e, as expected from previous simulations. GS predictions were poor when tested across populations, likely resulting from variable patterns of linkage disequilibrium, inconsistent allelic effects and genotype by environment interaction, suggesting that with such a modest marker density population-specific models will, in principle, be necessary (GRATTAPAGLIA; RESENDE, 2011; RESENDE, M. D. V. et al., 2012). In this study, we expand our understanding of the prospects of genomic selection for *Eucalyptus* breeding by developing predictive models in yet another operational breeding population, this time for a considerably wider range of traits. Wood chemical and physical traits were measured on 1000 trees of a progeny trial following the development of near infrared reflectance spectroscopy calibration models (see chapter 2). Genotyping was carried out using a high-density, high-throughput 60,639 SNP genotyping chip developed for *Eucalyptus* species (SILVA-JUNIOR et al., 2013)and two different models were usedfor genomic prediction. Besides assessing the predictive abilities and selection accuracies of the two models, a comparative assessment of the performance of genomic selection against standard BLUP phenotypic selection in ranking top individual trees was carried out.

3.2 Development

3.2.1 Material and methods

3.2.1.1 *Eucalyptus* population and growth data

The study was carried out with an operational elite population belonging to the breeding programof International Paper Brazil. The population was composed of a full-sib progeny trial involving second and third-generation hybrids mostly between *Eucalyptus grandis* and *E. urophylla,* and to a much smaller extent with *E. camaldulensis*. The trial planted in randomized complete blocks with five trees per plot comprised 58 full-sib families, totaling 2,784 trees planted in 2006, in Brotas, São Paulo State, Brazil (22°S; 48ºW). Trees were measured for diameter at breast height (DBH), plant height and wood volume estimated thereof by an equation that takes into account a tree form factor. Standard quantitative analyses and ranking were carried out using SELEGEN-REML/BLUP (RESENDE; OLIVEIRA, 1997) and a subset of 1,000 trees out of the 2,784 in the trial was selected based on growth (wood volume) to compose a training population, therefore avoiding off-type trees for form and growth. Selection was carried according to the mixed model:

$$\mathbf{y = Xr + Zg + Wp + e}$$

where y is the vector of phenotypic data (wood volume), **r** is the vector of fixed effects (i.e. mean and experimental effects), **g** is the vector of random additive genetic effects of individuals, **p** is the vector of fixed plot effects and **e** is the random residual effect. **X**, **Z** and **W** are incidence matrices for each of the effects. This selected *Eucalyptus* training population of 1,000 trees included 45 full-sib families and 46 different parents, in which 25 families were *E. grandis* x *E. urophylla* hybrids, totaling 610 trees (61%). *E. camaldulensis* was involved in only one cross with *E. grandis*, with 24 full-sibs (2%). Second generation hybrid families involving *E. grandis* and *E. urophylla* comprised the remaining trees (table 2.1).

3.2.1.2 Wood properties phenotyping

Wood sampling was performed on 5-year-old trees, close to the planned harvest age, which usually happens at age 6. Wood properties assessment was carried out according to the

methods presented earlier (see chapter 2), by sampling wood dust and cores, and establishing Near-Infrared Reflectance Spectroscopy (NIRS) prediction models to allow predicting wood traits for the whole population based on wet chemistry and physical measurements on a calibration set of reduced size (~350 trees). The wood chemical traits analyzed were cellulose content (%), hemicelluloses (%), soluble lignin (%), insoluble lignin (%) and total lignin (%); a composite trait by the sum of soluble and insoluble lignin). The wood physical traits analyzed were wood density ($kg.m^{-3}$), microfibril angle (°;MFA), fiber length (mm), fiber width (µm) and coarseness ($g.100m^{-1}$). Wood chemical traits were predicted by NIRS predictions while for wood physical traits only density had NIRS predictions. For the remaining wood physical traits NIRS did not provide usable calibration models, therefore GS models were based on the 350 trees directly phenotyped.

3.2.1.3 SNP genotyping

Genomic DNA was extracted from xylem (figure 2.1a), scratching part of the sapwood, using a previously described CTAB method (GRATTAPAGLIA; SEDEROFF, 1994), quantified by gel electophoresis and spectophotometry in a nanodrop 2000 instrument and normalized to concentrations between 20-40$ng.uL^{-1}$. DNA samples were genotyped at Geneseek using a custom made IlluminaInfinium chip developed in our laboratory that contains 60,639 assayable SNPs (the *EucHIP60k.br*) (SILVA-JUNIOR et al., 2013) Only SNPs that were genotyped with a call rate $\geq$ 0.90 and displayed a MAF (minor allele frequency) $\geq$ 0.01 were used in further analyses.

3.2.1.4 Standard phenotypic Best Linear Unbiased Prediction (BLUP)

To provide a standard breeding benchmark against which we compared the prediction accuracies obtained by Genomic Selection, Best Linear Unbiased Predictor (BLUP) analysis was carried out using the R package rrBLUP. The phenotypic BLUP is the

traditional model in plant breeding for genotype selection. The model estimates genotypic values, assuming fixed and random effects, as follows:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Zg} + \mathbf{e}$$

$$\mathbf{g} \sim \mathrm{N}(0, \mathbf{K}\sigma_{\mathrm{g}}^2)$$

where$\boldsymbol{\mu}$ is the intercept, **g** is the vector for random effects (breeding values), **Z** is the incidence identity matrix for random effects and **e** is the random residual effect. The **K** matrix is the numerator relationship matrix. The numerator relationship matrix was calculated in R using the package 'pedigreemm'. Phenotypic data were adjusted for experimental design effects using the R package 'lme4' and the deregressed phenotypes (**y**) were used for the subsequent genomic predictions.

3.2.1.5 Ridge-Regression BLUP (RR-BLUP)

Genomic predictions were developed using RR-BLUP (WHITTAKER; THOMPSON; DENHAM, 2000) implemented in the R package rrBLUP(ENDELMAN, 2011). The model is analogous to the Genomic-BLUP (MEUWISSEN; HAYES; GODDARD, 2001) with marker effects estimation. The model assumes that all markers explain *a priori* equal amounts of the genetic variation, with average shrunk to zero. The method solves the following mixed linear model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon}$$

$$\mathbf{u} \sim \mathrm{N}(0, \mathrm{K}\sigma_{\mathrm{u}}^2)$$

where y is the vector of phenotypic data, **β** is the vector of fixed effects, **u** is the vector of random effects (SNP markers) and ε is the vector of random error effects. **X**and**Z** are incidence matrices for fixed and random effects, respectively. The vector **u** has variance σ_{u}^2. The SNP matrix is the Z matrix, coded as -1, 0 and 1, in which -1 and 1 are codes for homozygous and 0 for the heterozygous genotype. Genomic estimated breeding values (GEBV; $\hat{\mathrm{y}}_{\mathrm{j}}$) is given by $\mathrm{GEBV} = \hat{\mathrm{y}}_{\mathrm{j}} = \sum_{\mathrm{i}} \mathrm{Z_i u_i}$, where $\hat{y}_j$ is the GEBV of the jth individual.

3.2.1.6 Bayesian LASSO (BLASSO)

The BLASSO model was also used to obtain genomic predictions of phenotypes based on Bayesian assumptions (PARK; CASELLA, 2008). The model assumes *a priori* the double-exponential distribution of marker effects, with average shrunk to zero. The model implemented in the R-package BLR (PEREZ et al., 2010) is represented by the same linear mixed model presented for RR-BLUP.

The SNP matrix was also coded as 1 and -1 for homozygous loci and 0 for heterozygous. Priors assumptions for the model were calculated according to the author's descriptions (PEREZ et al., 2010).

3.2.1.7 Cross-validation of genomic predictions

Validation followed a 10-fold cross-validation scheme using random subsampling replication. This procedure warrants independence between the predicted and the reference phenotypes (WRAY et al., 2013). Training population was randomly divided into ten subsets. Nine subsets were used for genomic selection parameters estimation, while phenotypes were predicted for the tenth independent subset.The correlation between reference and predicted phenotype was then estimated for this subset. This procedure was repeated 10 times, until all the subsets were used for parameter estimation and validation. Correlations for all subsets were averaged at the end. The procedure was run on R for RR-BLUP and BLASSO predictions. The estimated accuracy ($r_{g\hat{g}}$) for each trait is given by

$$r_{g\hat{g}} = \frac{r_{y\hat{y}}}{\sqrt{h^2}}$$

where$r_{y\hat{y}}$ is the predictive ability, i.e. the correlation between the predicted and observed phenotype and $h^2$is the narrow-sense heritability.

3.2.1.8 Individual tree ranking using Genomic Selection and phenotypic BLUP selection

To assess the effective accuracy of the two genomic selection approaches in identifying top trees for growth and wood properties traits, a preliminary breeding oriented analysis was carried out by comparing the coincidence of individual tree ranking between the two genomic methods (RR-BLUP and BLASSO) and between them and BLUP phenotypic

selection. GEBVs for all 999 individual trees in the population were obtained by cross validation as described above, and trees were ranked for three key traits for the pulp and paper industry: volume growth, density and S:G ratio. The coincidence between the top 30 and top 10 individual trees identified by the two genomic selection models and by BLUP phenotypic selection was estimated. Additionally a multi-trait sequential tandem selection was applied by selecting initially the top 100 trees (10% selection intensity) for volume growth, followed by the top 50 trees for wood density within the 100 selected ones for volume, and finally the top 25 for S:G ratio within the 50 selected trees for volume and wood density. This procedure was applied separately to the rankings obtained by RR-BLUP, BLASSO and phenotypic BLUP and the results compared.

3.2.2 Results

Genome-wide predictions were performed for all 15 traits, including growth, wood chemical and wood physical properties, using two prediction models: RR-BLUP and BLASSO. Out of the 60,441 assayable SNPs on the multi-species IlluminaInfinium EucHIP60k.br, genotype data with a call rate> 90% were obtained for a total of 47,903 SNPs in the population. After applying a filter for minimum allele frequency (MAF) ≥ 0.01, a total of 29,090 SNPs were retained and used in the subsequent analyses. Genomic predictions were compared with a standard phenotype-based BLUP analysis to provide a benchmark against which one may compare the results and expectations of future GS efforts. Heritabilities were estimated for each trait and under each prediction model given their different assumptions regarding the distribution of marker effects. A final adjustment was applied to the estimated accuracies by factoring in the correlations observed between observed and predicted phenotypes measured using the NIRS calibration models developed earlier (see chapter 2), An adjusted accuracy was therefore the product of the NIRS correlation by the genome-wide prediction accuracy. Wood physical traits, with exception of density, had only part of the population phenotyped and were not NIRS predicted, because of the poor correlations provided by the NIRS calibrations. For all other traits, a training population of 999 trees was used for genomic predictions, since one of the trees eventually was not genotyped and therefore excluded from the analyses.

3.2.2.1 Trait heritabilities

Narrow-sense heritabilities estimated with the two analytical methods were essentially equivalent (tables 3.1 and 3.2). Using RR-BLUP, heritabilities varied from 0.22 for fiber width, to 0.93, for S:G ratio while using BLASSO they varied from 0.15 to 0.92 (table 3.4) The highest heritabilities were observed for wood chemical traits, with a range between 0.74 for cellulose content to 0.93 for S:G ratio in the case of RR-BLUP estimates.. Wood physical traits, on the other hand, displayed variable heritabilities depending on the trait, ranging from a low of 0.15 for fiber width to a high of 0.75 or wood density and 0.80 for fiber length that were highly heritable as estimated by RR-BLUP. The only distinctive exception was the heritability for MFA, estimated at 0.26 by RR-BLUP and 0.46 by BLASSO, Growth traits showed relatively high heritabilities, with diameter growth slightly more heritable than height growth and an overall average of 0.61.

Table 3.1 – Heritabilities, predictive abilities and genome-wide selection accuracies estimated using RR-BLUP, and adjusted accuracies penalized by factoring the NIRS calibration correlations (when applicable), for growth, wood chemical and wood physical traits, under a 10-fold cross-validation scheme

Trait	Population size	NIRS correlation [1]	Heritability (h^2)	Predictive ability[2]	Accuracy	Adjusted accuracy [3]
DBH (cm)	999	-	0.64	0.44	0.55	0.55
Plant Height (m)	999	-	0.54	0.34	0.46	0.46
Wood Volume (m^3)	999	-	0.63	0.42	0.53	0.53
MAI ($m^3.ha^{-1}.year^{-1}$)	999	-	0.63	0.42	0.52	0.52
Cellulose (%)	999	0.57	0.74	0.52	0.60	0.34
Hemicelluloses (%)	999	0.60	0.81	0.55	0.62	0.37
S:G ratio	999	0.93	0.93	0.83	0.86	0.80
Insoluble lignin (%)	999	0.84	0.82	0.64	0.70	0.59
Soluble lignin (%)	999	0.83	0.84	0.72	0.79	0.66
Total lignin (%)	999	0.84	0.82	0.63	0.70	0.59
Density ($kg.m^{-3}$)	999	0.74	0.76	0.63	0.72	0.53
MFA (°)	348	-	0.26	0.17	0.34	0.34
Fiber length (mm)	350	-	0.80	0.49	0.54	0.54
Fiber width (µm)	350	-	0.22	0.14	0.29	0.29
Coarseness ($g.100m^{-1}$)	349	-	0.49	0.33	0.48	0.48

[1]NIRS based correlation between the predicted and observed phenotypes; [2]Predictive ability given by the correlation between genomic-estimated breeding value (GEBV) and observed deregressed phenotype; [3]Accuracy adjusted by multiplying the genome-wide selection accuracy by the NIRS correlation

3.2.2.2 Genomic predictive abilities

Predictive abilities estimated using the two analytical models were very similar (tables 3.1 and 3.2). For both methods the predictive abilities were strongly and equally correlated with trait heritabilities (R^2 = 0.90). Thus, for simplicity we will refer only to the estimates derived from the RR-BLUP model, unless otherwise stated. The highest predictive abilities were obtained for wood chemical properties related to lignin content and composition followed by cellulose related traits. Wood density and fiber length were slightly more predictable than growth traits, while the remaining physical traits were the least predictable. Fiber width displayed a low heritability at 0.15 with a prediction ability of 0.10 when BLASSO was used (table 3.2). Noteworthy is the fact that the size of the training population used to estimate effects for the physical traits (N= 350) was considerably smaller than the one

used for the other traits which certainly impacted negatively on the predictive power of the models.

Table 3.2 - Heritabilities, predictive abilities and genome-wide selection accuracies estimated using BLASSO, and adjusted accuracies penalized by factoring the NIRS calibration correlations (when applicable), for growth, wood chemical and wood physical traits, under a 10-fold cross-validation scheme

Trait	Population size	NIRS correlation [1]	Heritability (h^2)	Predictive ability[2]	Accuracy	Adjusted accuracy [3]
DBH (cm)	999	-	0.62	0.44	0.55	0.55
Plant Height (m)	999	-	0.50	0.35	0.50	0.50
Wood Volume (m^3)	999	-	0.64	0.40	0.50	0.50
MAI ($m^3.ha^{-1}.year^{-1}$)	999	-	0.61	0.43	0.56	0.56
Cellulose (%)	999	0.57	0.72	0.52	0.61	0.35
Hemicelluloses (%)	999	0.60	0.81	0.57	0.63	0.38
S:G ratio	999	0.93	0.92	0.83	0.86	0.80
Insoluble lignin (%)	999	0.84	0.79	0.64	0.72	0.60
Soluble lignin (%)	999	0.83	0.83	0.73	0.80	0.66
Total lignin (%)	999	0.84	0.82	0.63	0.70	0.59
Density ($kg.m^{-3}$)	999	0.74	0.75	0.62	0.71	0.53
MFA (°)	348	-	0.46	0.20	0.29	0.29
Fiber length (mm)	350	-	0.72	0.49	0.57	0.57
Fiber width (µm)	350	-	0.15	0.10	0.26	0.26
Coarseness ($g.100m^{-1}$)	349	-	0.43	0.33	0.50	0.50

[1]NIRS based correlation between the predicted and observed phenotypes; [2]Predictive ability given by the correlation between genomic-estimated breeding value (GEBV) and observed deregressed phenotype; [3]Accuracy adjusted by multiplying the genome-wide selection accuracy by the NIRS correlation

Scatter plots between the genomic predicted breeding values (GEBVs) and the observed deregressed phenotypes, both for RR-BLUP and BLASSO, visually illustrate the genomic prediction ability achieved for the different traits (figures 3.1 and 3.2). Calculated slopes for these regressions were not significantly different from one indicating that the predictions resulting from both models do not contain any appreciable bias.

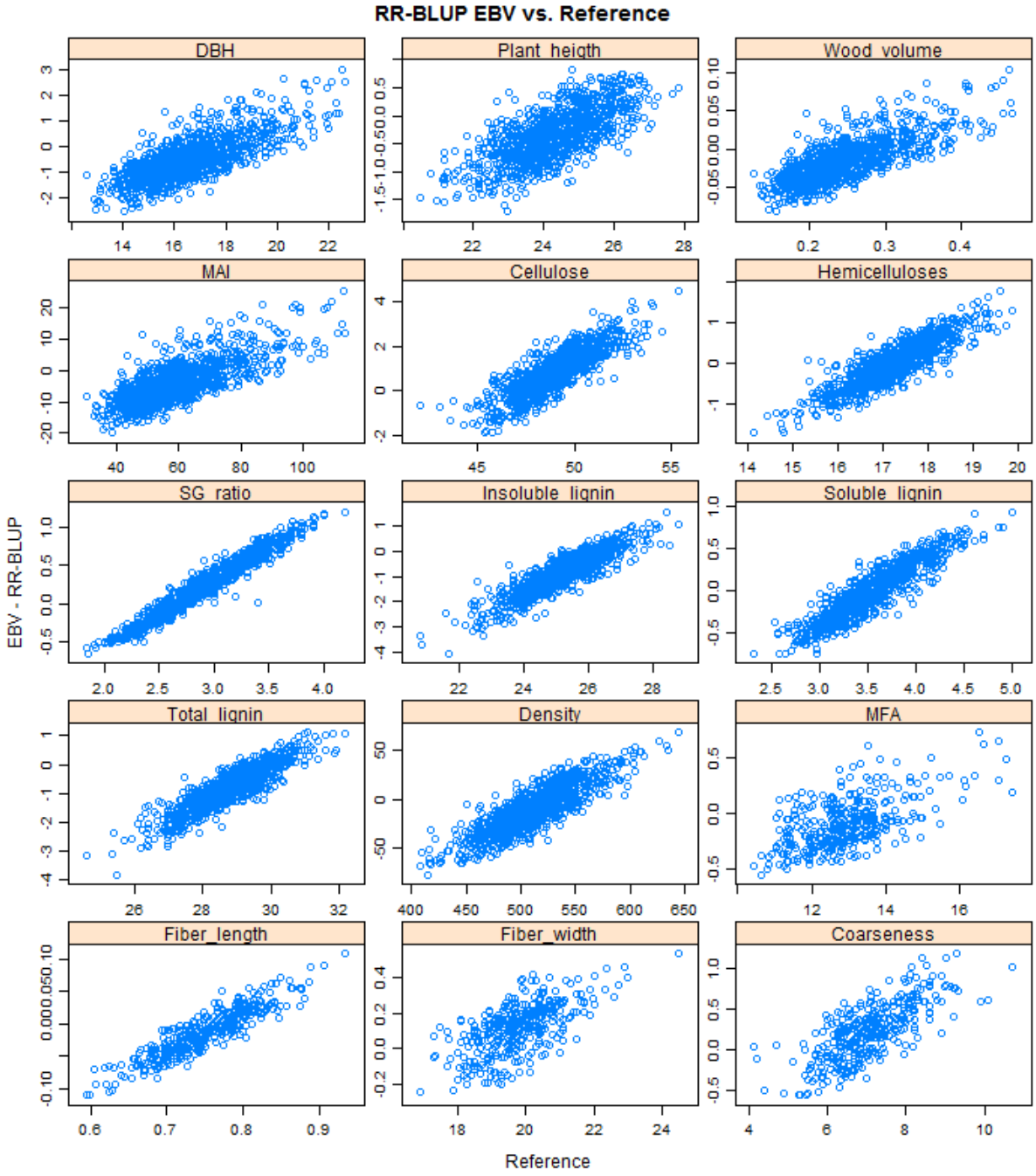

Figure 3.1 – Scatter plots of RR-BLUP genomic estimated breeding values (GEBVs) and observed deregressed phenotypes for fifteen traits in the *Eucalyptus* training population

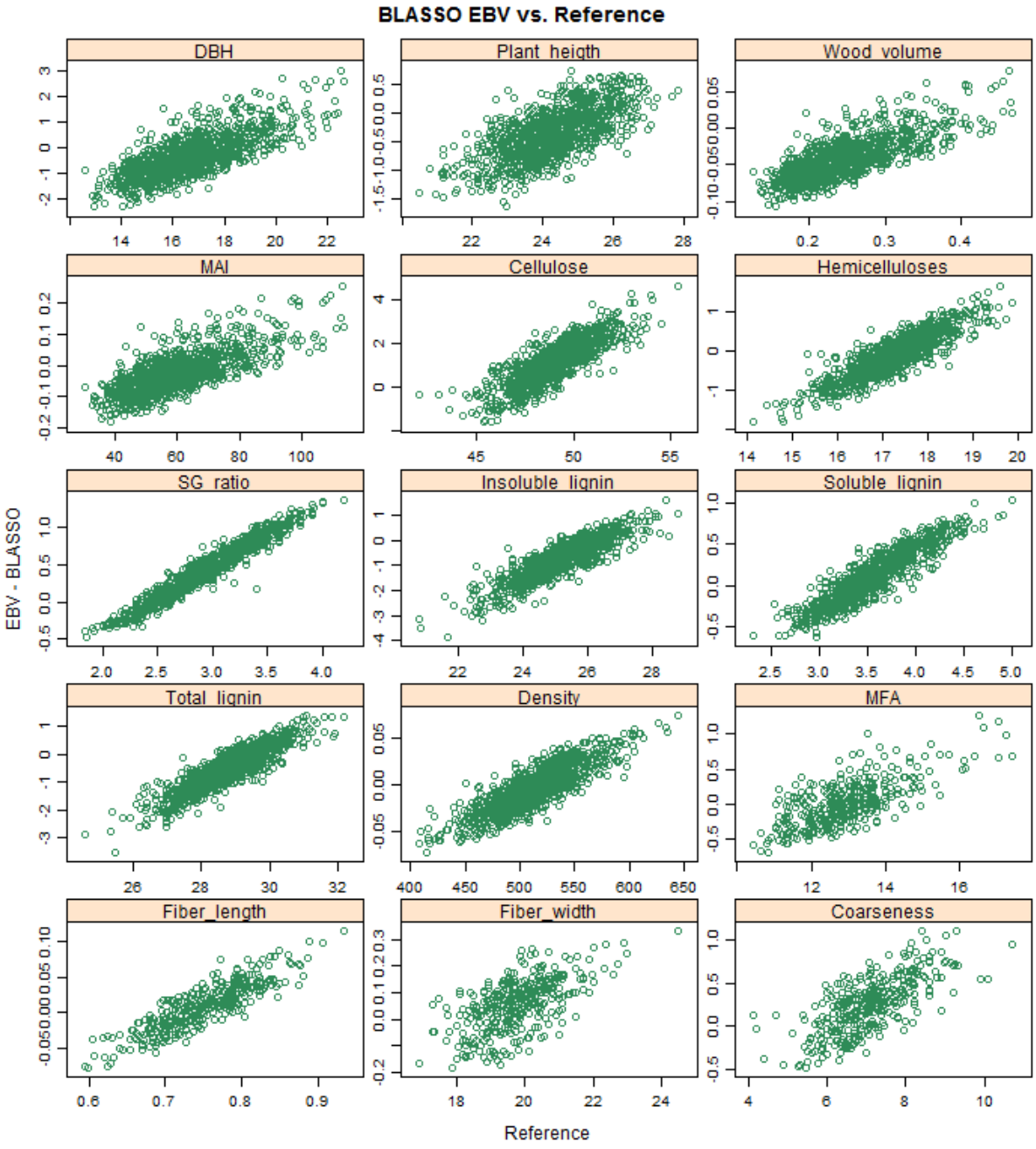

Figure 3.2 - Scatter plots of BLASSO genomic estimated breeding values (GEBVs) and observed deregressed phenotypes for fifteen traits in the *Eucalyptus* training population

To examine the expected consistency of individual tree ranking we estimated Spearman's correlations between the GEBV estimated by the two models and by each one versus standard BLUP estimated phenotypes for all 15 traits. Such correlations were taken as an approximate way to assess how differently would each one of the three methods rank individuals in a selection stage during a breeding program (table 3.3). Correlations between RR-BLUP and BLASSO were very high, as expected, close to unity in most cases. The

weaker correlation was observed for MFA (0.981), corroborating previous results of a general equivalence of the two analytical approaches to prediction. Correlations between the genomic predicted breeding values and observed BLUP phenotypes, were generally lower, but still considerably high, varying from 0.771 for MFA to 0.929 for total lignin content under the RR-BLUP model, suggesting that genomic selection would deliver very similar individual tree rankings to the ones yielded by BLUP phenotypic selection (see below).

Table 3.3 – Spearman's correlations of the individual GEBV estimated by the two models and by each one versus standard BLUP estimated phenotypes for all individuals in the training populations, for all 15 traits

Traits	Spearman'sCorrelations		
	RR-BLUP/ BLASSO	Phenotypic BLUP/ RR-BLUP	Phenotypic BLUP/ BLASSO
DBH (cm)	1.000	0.863	0.859
PlantHeight (m)	0.999	0.843	0.838
Wood Volume (m3)	1.000	0.857	0.858
MAI (m3.ha-1.year-1)	1.000	0.857	0.853
Cellulose (%)	1.000	0.882	0.881
Hemicelluloses (%)	1.000	0.902	0.903
S:G ratio	1.000	0.928	0.926
Insolublelignin (%)	0.999	0.922	0.912
Solublelignin (%)	1.000	0.899	0.895
Total lignin (%)	1.000	0.929	0.927
Density (%)	1.000	0.910	0.908
MFA (°)	0.981	0.771	0.818
Fiberlength (mm)	0.997	0.942	0.932
Fiberwidth (µm)	0.994	0.824	0.803
Coarseness (g.100m-1)	0.998	0.872	0.862

3.2.2.3 Estimation of SNP marker effects

The only evident difference between the predictions obtained with RR-BLUP and BLASSO concerns the estimation of the vector of marker effects, which results from the different assumptions of the two models. While with RR-BLUP the marker effects are expected to be normally distributed, in BLASSO marker effects are expected to follow a double-exponential distribution. With a double-exponential distribution, marker effects are shrunk to values close to zero and fewer markers display larger effects. Looking at the Manhattan plots of SNP marker effects plotted against their position in the 11 chromosomes of the *Eucalyptus* reference genome, this pattern becomes clear (figures 3.4 and 3.5). Variation in the magnitude of SNP effects estimated by RR-BLUP is broader than by

BLASSO and it becomes clearer for traits with higher heritabilities. For S:G ratio, for example, with RR-BLUP several markers with relatively large effects are seenin a genomic region on chromosome 10. Using BLASSO, this same region also is detected even more distinctively due to the shrinkage effect of the other markers. For traits like DBH, cellulose content and wood density, no such regions with a large concentration of markers with larger effects is observed, although the differential shrinkage provided by the two models can be similarly observed.

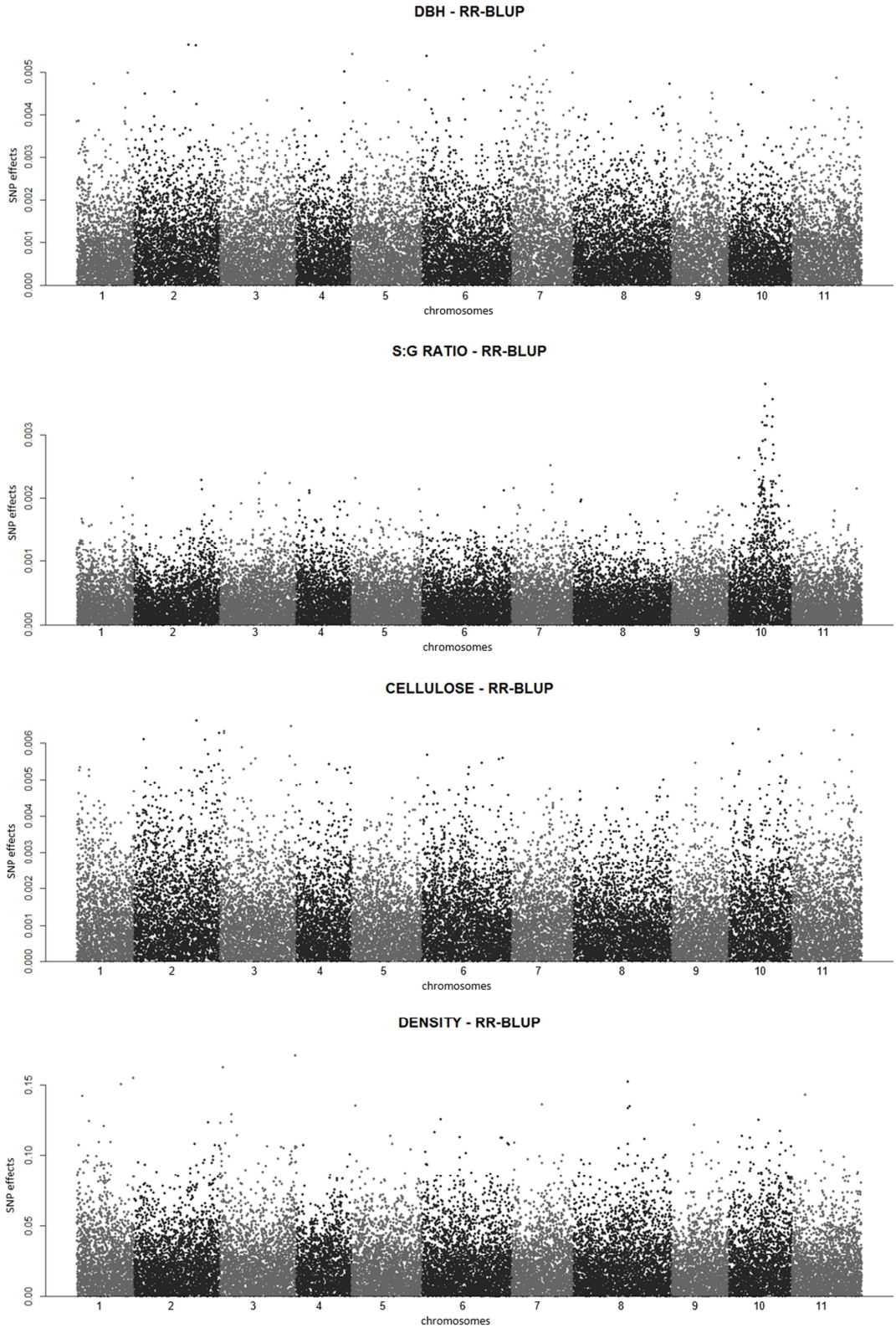

Figure 3.4 – Distribution of the magnitude SNP effects in additive genetic standard deviations estimated using RR-BLUP along the eleven chromosomes of the *Eucalyptus* reference genome

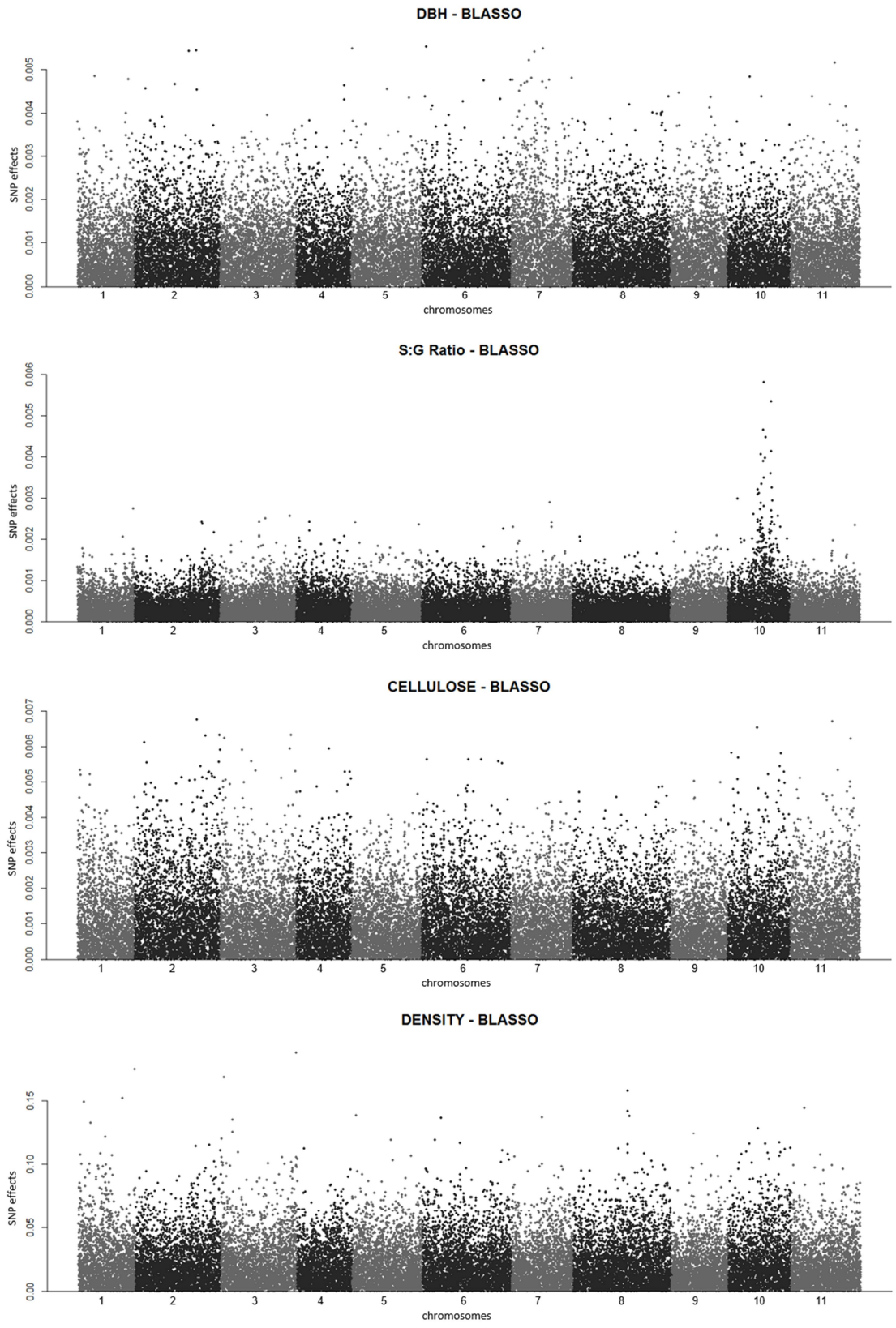

Figure 3.5 – Distribution of the magnitude SNP effects in additive genetic standard deviations, estimated using BLASSO along the eleven chromosomes of the *Eucalyptus* reference genome

3.2.2.4 Genome-wide selection accuracies

Genome-wide selection accuracies generally followed the trend observed for predictive abilities (tables 3.1 and 3.2). For those traits for which phenotypes were predicted using NIRS calibration models and such models had lower correlations, accuracies were reduced. Still, the highest adjusted accuracy was observed for S:G ratio at 0.80, reflecting the high quality of the NIRS calibration model previously built, particularly for lignin related traits. Adjustment of accuracies was particularly severe for cellulose and hemicellulose content and wood density. For example the accuracy was reduced by 43% going from 0.60 to 0.34 for hemicellulose content and by 26% for wood density, going from 0.72 to 0.53 (table 3.1). A comparison of the estimated selection accuracies by standard BLUP based phenotypic selection and the two analytical prediction approaches suggests, in principle, that genomic selection will be less accurate than standard phenotypic selection for all traits (figure 3.3). The largest differences in accuracies between phenotypic BLUP based selection and genomic selection were observed for growth traits where standard selection would reach accuracies above 0.70 while GS would provide accuracies around 0.50. For wood properties traits, differences were relatively smaller, exception made for MFA and fiber width, consistent with the low heritabilities estimated for these traits.

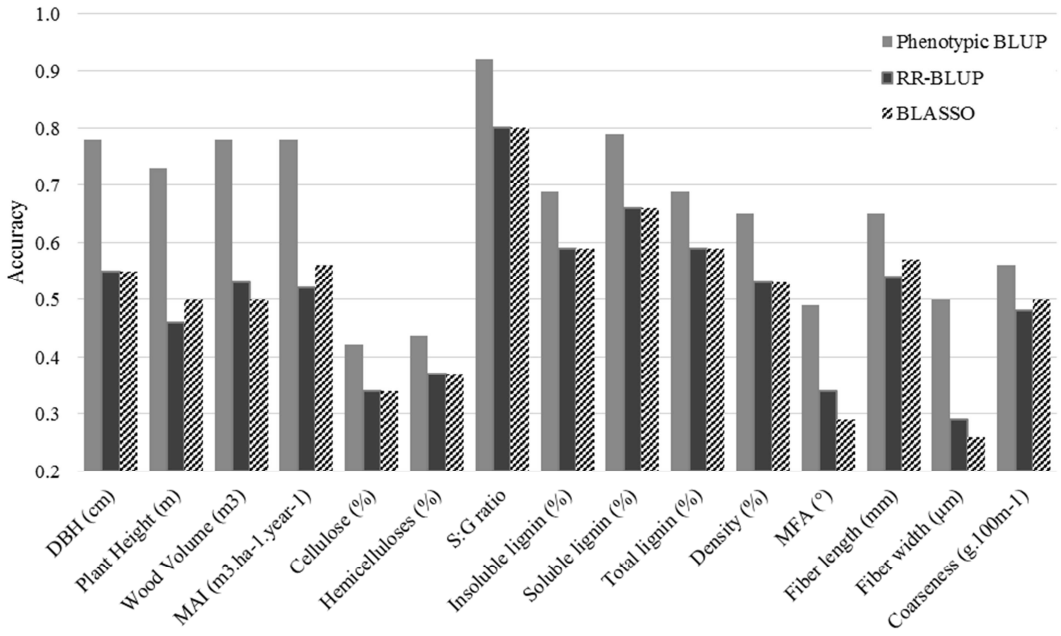

Figure 3.3 – Comparison of adjusted selective accuracies expected from phenotypic BLUP, and RR-BLUP and BLASSO models in the *Eucalyptus* population for all 15 traits

3.2.2.5 Comparative individual tree ranking using GS and phenotypic BLUP

The individual tree level assessment of selective accuracy provided by the genomic predictions indicated that the two models, RR-BLUP and BLASSO, would produce essentially the same rankings for the top 30 trees for all three traits examined, with a few minor rank changes, always within one or two positions at most (table 3.4). When compared to the benchmark ranking resulting from phenotypic BLUP based selection, a larger rank change variation at the individual level was observed. Nevertheless, the overall coincidence of the top 30 trees ranked by genomic selection with the top 30 ranked by BLUP phenotypic selection was high: 83% for S:G ratio, 77% for volume growth and 70% for wood density. Examining how would the genomic selection models do at an even higher selection intensity, i.e. the top 10 phenotypic BLUP ranked trees for each trait, the results show that GS would identify 8 of them for volume, 7 for wood density and 6 for S:G ratio (table 3.4). In other words, an undisputable majority of the top ranked trees by phenotypic selection would also be successfully identified by both genomic selection models. When a multi-trait tandem selection was applied, starting with a 10% selection intensity (top 100 trees) for volume growth, followed by selecting therein the 50 top trees for wood density and then the top 25 for S:G ratio, a total of 15 trees out of the 25 (60%) that were identified by phenotypic selection were also identified by genomic selection, again showing that genomic predictions were successful even under a more stringent and realistic multi-trait selective procedure (table 3.5).

Table 3.4 - Top 30 individual trees (numbers follow their ID in the trial) for volume growth (VOL), wood density (DENS) and S:G ratio based on Phenotypic BLUP, RR-BLUP and BLASSO. Coincident individuals in the three rankings for each trait are shaded; coincident individuals in the top 10 ranked by phenotypic BLUP are indicated in bold letters; the proportion of coincident individuals with phenotypic BLUP (%COI) is indicated at the bottom

RANK	VOL PBLUP	VOL RRBLUP	VOL BLASSO	DENS PBLUP	DENS RRBLUP	DENS BLASSO	S:G PBLUP	S:G RRBLUP	S:G BLASSO
1	**710**	**710**	**710**	**738**	**892**	**892**	**607**	**607**	**607**
2	**950**	**950**	**950**	**658**	**835**	**835**	**12**	**12**	**12**
3	**382**	607	607	**892**	**658**	**658**	**5**	**5**	**5**
4	**49**	**380**	**380**	**835**	239	239	**237**	622	622
5	585	**46**	**46**	**752**	**738**	**738**	48	**237**	**237**
6	**46**	**382**	**382**	**366**	114	114	924	326	326
7	**444**	**606**	**606**	759	**752**	**752**	328	387	47
8	859	**49**	**49**	182	**366**	**366**	622	47	387
9	**380**	**444**	**444**	**112**	503	503	**425**	816	816
10	**606**	881	881	359	**112**	**112**	169	**425**	**425**
11	604	711	711	114	304	304	816	490	490
12	226	713	713	501	359	359	476	329	169
13	879	206	206	503	759	751	47	169	329
14	607	948	948	115	751	571	843	328	328
15	713	308	585	155	571	759	819	843	843
16	47	585	308	157	113	113	387	48	77
17	951	312	312	239	61	61	580	77	327
18	176	879	879	304	853	853	333	327	48
19	711	381	381	897	805	805	383	924	817
20	443	586	586	9	367	754	534	817	924
21	881	947	65	900	155	367	976	383	383
22	882	65	947	657	754	155	490	333	333
23	947	176	176	333	156	156	206	476	476
24	586	604	951	805	165	165	329	28	28
25	48	951	604	961	157	157	327	882	882
26	308	859	859	209	657	657	963	819	819
27	379	687	224	571	501	739	326	402	742
28	948	224	687	567	739	501	882	742	402
29	582	379	379	609	209	609	605	206	206
30	574	205	205	871	609	209	885	580	580
% COI.	-	77	77	-	70	70	-	83	83

Table 3.5 - Top 25 trees (numbers follow their ID in the trial) selected by sequential tandem selection for volume growth, wood density and S:G ratio based on Phenotypic BLUP, RR-BLUP and BLASSO (see text for details); coincident individuals in the rankings are shaded; the proportion of coincident individuals with phenotypic BLUP (%COI.) is indicated

RANK	PBLUP	RRBLUP	BLASSO
1	881	218	218
2	412	793	793
3	711	951	951
4	793	894	894
5	955	832	832
6	587	582	582
7	894	1	1
8	951	65	65
9	65	311	732
10	582	732	311
11	863	863	863
12	311	729	729
13	226	310	584
14	300	584	310
15	1	224	969
16	584	969	224
17	585	233	233
18	445	404	404
19	732	445	445
20	233	226	226
21	88	585	585
22	63	869	869
23	709	958	958
24	825	104	104
25	550	550	550
% COINCIDENCE	-	60	60

3.2.3 Discussion

This study expands our current understanding and experimental results of the positive prospects of applying genomic selection to *Eucalyptus* breeding and tree breeding in general (GRATTAPAGLIA, 2014). Genomic prediction models were developed for a wider range of traits using a training population of up to 999 individuals genotyped with a high-throughput SNP genotyping platform that provided an effective polymorphic marker density around 1 SNP/21 kbp, significantly higher than any previous study in any forest tree. The accuracy of predictions for growth traits and eleven chemical and physical wood properties were estimated. Two analytical approaches that differ regarding the assumed distribution of marker effects were used. Predictive abilities observed with RR-BLUP and Bayesian LASSO were similar, although the distribution of marker effects varied and considerably more so depending on the trait under study. Extensive coincidence of ranking at the individual tree level was observed between genome based selection and standard BLUP phenotypic selection when the majority of the top 10 and top 30 trees selected by BLUP phenotypic selection also successfully identified by genomic selection.

3.2.3.1 Phenotyping quality directly impacts heritability and the accuracy of GS

High quality phenotyping and its resulting impact on trait heritability and predictive ability of the model is a critical aspect for the successful implementation of GS. We used previously developed NIRS calibration models to measure wood chemical and physical traits in the training population. This tool has been extensively used for assessing wood properties mainly in temperate *Eucalyptus globulus*(DOWNES et al., 1997; COSTA E SILVA et al., 2009; STACKPOLE et al., 2011)and more recently in tropical *E. urophylla*(HEIN; LIMA; CHAIX, 2009; DENIS et al., 2013). In our study and specific breeding population such NIRS models were accurate for chemical traits related to lignin content and composition, less so for cellulose and physical traits related to fibers characteristics. For these latter ones the NIRS models were in fact unusable, a fact that resulted in the reduction of the training population size to only 350 trees for which these traits were actually directly measured. The correlation of the NIRS calibration model was incorporated in the final adjusted accuracy of both selection methods, standard BLUP phenotypic and genomic-based. This conservative correction resulted in a considerable reduction of the accuracies depending on the trait, with higher penalization for cellulose and wood density. Accuracies went from ~0.7 down to 0.3

and 0.5 for these two traits respectively, due to this correction. In our previous assessments of the accuracy of GS in *Eucalyptus* for wood density and pulp yield, we also used indirect measurement methods. Pilodyn penetration was used for wood specific gravity and NIRS for pulp yield. However since we had no information on the actual correlation of the models such correction was not carried out(RESENDE, M. D. V. et al., 2012). We believe, however, that it is sensible to use such stringent corrections given the imprecision introduced by the NIRS predictions that will evidently reflect in the final selection accuracy. This provides a more realistic expectation of selection accuracies when eventually implementing GS. We recommend that this correction should be adopted in all studies dealing with prediction of traits which are actually measured by any indirectmean in the training population. The lack of correction most likely explains the difference in GS accuracies obtained in our study for wood density (0.53) with those seen in that previous study (0.65), notwithstanding some important differences in genotyping density, marker data use and cross-validation carried out (see below).

3.2.3.2 RR-BLUP and Bayesian modelling provided similar accuracies

Predictive abilities for all examined traits were strongly correlated with trait heritabilities, corroborating the fact that phenotypes for traits with lower heritabilities contain a larger environmental noise and therefore will be less predictable by genomic data as well(VISSCHER, 2008). This was the case for the wood fiber traits, aggravated by the smaller training population, well below the suggested minimum size of 1,000 individuals, from which point on our earlier simulation studies for a forest tree breeding context have shown that accuracies tend to plateau (GRATTAPAGLIA; RESENDE, 2011). Despite the different assumptions of the two prediction models regarding the distribution of SNP marker effects, predictive abilities estimated were essentially the same for all traits examined (tables 3.1 and 3.2), and the correlations of the GEBV between RR-BLUP and BLASSO were equal to one or close for all traits (table 3.3). Similar consistency of predictive abilities between ridge regression and Bayesian methods has been reported in genomic selection studies in apple (KUMAR et al., 2012), loblolly pine (RESENDE, M. F. R. et al., 2012), sugarcane (GOUY et al., 2013) and dairy cattle (HAYES et al., 2009). These results suggest that the genetic architecture for all growth and wood properties traits examined in this study should fit a quasi-infinitesimal model involving a large number of QTLs of small and possibly equal effect. Our results, generated from data for a wide representation of genetic variation

contributed by 45 full-sib families, deeply contrasts from several QTL mapping results in eucalypts, where putatively major effect QTLs explaining up to 30% of the phenotypic variation have been reported (reviewed by GRATTAPAGLIA et al. 2012). As expected, all those reported QTLs suffer from overestimation due to the well-known Beavis effect and would hardly be usable for any breeding application

3.2.3.3 High-density SNP genotyping allows tracking the distribution of QTLs across the genome and a more durable marker-QTL LD

To the best of our knowledge, this is the first study in forest trees to use a high-density and genome-wide distributed set of SNP markers across all the sequenced reference genome to assess the prospects of Genomic Selection. Our previous genomic selection studies in eucalypt used a ~3,500 DArT marker genotyping platform enriched for probes targeting low copy gene rich space, with good coverage but limited density (PETROLI et al., 2012). A loblolly pine study used 4,825 SNPs that had been specifically selected to target candidate genes and no information on genomic distribution was available (ECKERT et al., 2010). Association genetics studies in poplar used a higher-density chip with 29,233 SNPs, but again SNPs were specifically selected to cover a set of 3,543 candidate genes (GERALDES et al., 2013). The 29,090 polymorphic SNPs that we effectively used in our genomic selection modelling covered the whole genome at an average density of 1/21kbp. This allowed us to estimate the genome-wide distribution of SNP effects for all 15 traits across the whole sequenced genome. With this marker density which corresponds to an average of 1 marker every 0.04 cM, and considering an effective population size around N_e= 50 for the studied population, it is reasonable to assume that the SNP effects are well distributed within the relatively long distance range of linkage disequilibrium, therefore capturing a good proportion of the underlying QTL effects. Accordingly, the distribution of estimated SNP effects should closely follow the distribution of underlying QTL effects. Irrespective of the analytical model used and trait analyzed, the distribution of SNP effects covered the whole genome. Plots for only four traits are presented which closely represent the remaining traits (figures 3.4 and 3.5). No clustering of larger effects is seen for growth, wood density and cellulose, a pattern consistent with a truly polygenic distribution of effects. Some SNPs show slightly larger effects (> 0.005 additive genetic standard deviations), which in practice are still very small for any consideration as being individually relevant for further association or gene discovery studies. For S:G ratio, however, a cluster of effects stands out on chromosome 10.

Major effect QTLs have been mapped for S:G and lignin content on the corresponding linkage group 10 (THUMMA et al., 2010; GION et al., 2011; FREEMAN et al., 2013) and a recent association study in *Eucalyptus globulus* using DArT markers, found a significant effect in the same region on chromosome 10 at less than 1Mbp to its proposed candidate gene ferulate 5-hydoxylase (CAPPA et al., 2013). No attempt was made to carry out association tests in our study, although the data and models could be used for that purpose, provided that adequate correction of the existing structure is carried out to avoid spurious associations. The observed distribution of SNP effects along the genome suggests, however, that if such an analysis were to be performed, a large number of putative marker-trait associations should surface, consistent with the polygenic nature of the traits studied.

The high-density genotyping employed in this study will have an important impact on the continued accuracy of the GS models across the next generations. The extent of marker–QTL LD, modeled by the effective population size and genotyping density showed the largest impact on the prospects of GS in forest tree breeding (GRATTAPAGLIA; RESENDE, 2011). For effective population sizes up to N_e = 100, we had seen that a marker density of 0.05 cM would be necessary for keeping high accuracies of GS. Such a target genotyping density corresponds to some 24,000 SNP markers for a 1,200 cM recombining *Eucalyptus* genome. Additionally, a still largely overlooked aspect in the discussion of genomic selection in plants is that the impact of genotyping density will become even more important as generations of selection advance(GRATTAPAGLIA, 2014). In the absence of selection, higher marker density is beneficial to the persistence of GEBV prediction accuracy due to a slower decay of LD among markers and trait loci. However with directional selection simulations studies have indicated that high-density genotyping will be essential to sustain accuracy and keep selection effective for more generations, especially when a finite number of QTL loci is assumed, rather than an infinitesimal model (MUIR, 2007; LONG et al., 2011). In such cases, selection, together with recombination, may change the pattern of LD between markers and QTLs. The new LD generated by selection can be unfavorable for GEBV prediction, which was based on the original marker-QTL LD structure in the training population. Models developed in our study for growth and wood properties traits are based on 29,090 SNPs, an average marker density of 0.04 cM, therefore surpassing that theoretical requirement. Moreover, our models used all the genotyped markers, irrespective of their effects, and not just a subset of the more significant or highest estimated effect markers, a procedure that is known to upwardly bias estimates of accuracies (WRAY et al., 2013), and

make the models far more susceptible to a decay of accuracy as the LD structure changes over generations.

3.2.3.4 Benchmarking the accuracies of genomic selection models against standard BLUP phenotypic selection

A comparison of the estimated accuracies of selection indicated that genomic selection would be slightly less accurate than conventional phenotypic selection for all traits. This result is consistent with our earlier study with a comparable breeding population with N_e=51. GS accuracies were 15% lower for DBH and 18% for wood density in this study (figure 3.3) whereas accuracies were approximately 10% lower than those obtained by phenotypic BLUP selection for DBH and 15% for wood density in the earlier study(RESENDE, M. D. et al., 2012). Such differences in estimated accuracies between BLUP phenotypic selection and genomic selection could be real, or be simply due to random variation, or due to bias that goes into the estimation process. A source of bias in the estimate of accuracy of standard BLUP phenotypic selection could be the use of a numerator relationship matrix. Cryptic unaccounted relatedness may bias heritability estimates downward, which in turn would result in overestimated accuracy of selection when compared to the heritability and accuracy estimated for genomic selection using a realized relationship matrix. The use of genome-wide SNP relationships information has been shown to efficiently disentangle confounding factors when estimating genetic variances and heritability so that variance components were more accurate and less biased, compared to those based on pedigree information only (HAYES; VISSCHER; GODDARD, 2009; LEE et al., 2010).

3.2.3.5 Ranking trees by Genomic Selection

The number of experimental reports of genomic selection in plants in general is still relatively limited (JONAS; DE KONING, 2013) and much so for forest trees (GRATTAPAGLIA, 2014). Additionally for the few studies published to date, estimations of accuracy are based on cross validation within the same generation, which is still not yet a bona fide demonstration of the value of Genomic Selection. Recently results of the first actual genome-wide experiment across generations were published in maize. Success of GWS compared to marker-assisted selection using only significant markers was seen for grain yield and stover-quality traits in testcrosses of recombinant inbreds(MASSMAN; JUNG;

BERNARDO, 2013). However GWS marker effects estimated from maize single crosses are not advantageous compared to phenotypic selection for predicting single-cross performance and have erratic usefulness for predicting testcross performance within a biparental cross(MASSMAN et al., 2013). Our present study also offers only estimates derived from cross validation within the same generation. Experiments are underway, however, to use these models to selecttop juvenile seedlings derived from a much larger number of individuals of the same progenies used for training (i.e. same generation), and, in the near future, in the offspring of the recombination of the top ranked individuals out if this training population (next generation). To this end, we wanted to go beyond the conventional and "cold" overall estimate of GS accuracy and took a first look at the accuracy of the GS models for selecting at the individual tree level, i.e. in identifying the top ranked trees that BLUP phenotypic selection would have identified. We recognize that such assessment is still not ideal since the ranking of individuals by their GEBV were obtained in the same generation with the same data set, despite using cross validation to estimate the GEBVs. Nevertheless it was remarkable to see that both GS models yielded a coincidence above 70% in selecting the top 30 trees rankedby phenotypic selection, and above 60% for the top 10 (table 3.4). When tandem multi-trait selection was applied to growth wood density and S:G ratio, 15 out of the top 25 trees identified by BLUP phenotypic selection were also identified by GS (table 3.5). These result supports our earlier proposition by which GS be efficient for ultra-early selection of genomically top ranked seedlings, significantly reducing the length of a breeding cycle in *Eucalyptus* by precluding the progeny trial step(GRATTAPAGLIA; RESENDE, 2011; RESENDE, M. D. et al., 2012). Results of our comparative rankings analyses show that precise ranking positions would likely vary between phenotype and genome based selection (see table 3.4). However, within a group of top ranked individuals at high selection intensity, GS would likely identify the majority of the highly ranked trees by standard phenotypic selection. For clonal selection these seedlings would be propagated and tested in a verification trial in the field. If trees were to be used as parents for the next generation the selected juveniles could be induced to flower either by top-grafting or by chemical treatment and recombined to ultimately conclude a cycle of genetic improvement several years earlier(GRATTAPAGLIA, 2014). In this latter case, the realized relationship matrix would find an additional important use by providing the breeder a precise estimate of relatedness amongst the selected individuals, so that he could better decide about what crosses to take ahead to minimize inbreeding. In both selection routes, whether for cloes or for parents, although some top ranked individuals could be missed by genomic selection, the gain

achieved by reducing breeding time would fully warrant the operational application of this revolutionary breeding method.

3.3 Conclusions

We have reported an additional assessment of Genomic Selection in a *Eucalyptus* breeding population using a much higher-density SNP genotyping platform than any other study in forest trees. Results positively point to the potential value of GS as a breeding tool to accelerate selection for complex traits in forest trees. Accurate genomic-enabled predictions were obtained for wood chemical traits, particularly related to lignin content and composition, wood density and growth. These traits largely impact the use of wood for a number of industrial applications for the production of pulp, paper and energy from biomass. Good accuracies were also obtained for some physical fiber properties such as fiber length and coarseness. Cellulose content, MFA and fiber width still require improvements of phenotyping methods to allow amassing large numbers of phenotypes with high precision. No difference was seen between the two modelling approaches to genomic prediction , RR-BLUP and Bayesian LASSO, indicating that a model involving a large number of QTLs of small and equal effect fits the data well for all traits examined. Ranking of individual tree GEBVs largely overlapped with ranking derived from standard phenotypic selection both at the individual trait level and by tandem multi-trait selection, although still evaluated in the same generation. Soon data will be available to effectively asses the practice of GS across generations which should be greatly enhanced by the higher-density SNP chip used in this study. By that time, the incorporation of GS into a forest tree breeding program will require a detailed cost/benefit analysis. We contend, however, that the increased breeding efficiency achieved by replacing progeny testing, streamlining clonal testing and simultaneously selecting trees for several traits at very young age, will make genomic selection an increasingly common tree breeding tool, especially as genotyping costs continue to decline.

References

ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS.**Statistical yearbook 2013**: base year 2012. Brasília, 2013.148 p.

BEAVIS, W.D. QTL analyses: power, precision, and accuracy. In: PATTERSON, A.H. (Ed.). **Molecular dissection of complex traits**. Boca Raton: CRC Publ., 1998. p. 145-162.

BERNARDO, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science**, Madison, v. 48, n. 5, p. 1649-1664, 2008.

CAPPA, E.P.; EL-KASSABY, Y.A.; GARCIA, M.N.; ACUNA, C.; BORRALHO, N.M.; GRATTAPAGLIA, D.; MARCUCCI POLTRI, S.N. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. **PLoS One**, San Francisco, v. 8, n. 11, p. e81267, 2013.

COSTA E SILVA, J.; BORRALHO, N.; ARAÚJO, J.; VAILLANCOURT, R.; POTTS, B. Genetic parameters for growth, wood density and pulp yield in *Eucalyptus globulus*. **Tree Genetics & Genomes**, Heidelberg,v. 5, n. 2, p. 291-305, 2009.

CROSSA, J.; PEREZ, P.; HICKEY, J.; BURGUENO, J.; ORNELLA, L.; CERON-ROJAS, J.; ZHANG, X.; DREISIGACKER, S.; BABU, R.; LI, Y.; BONNETT, D.; MATHEWS, K. Genomic prediction in CIMMYT maize and wheat breeding programs. **Heredity**, London, v. 112, n. 1, p. 48-60, 2014.

CROSSA, J.; BEYENE, Y.; KASSA, S.; PEREZ, P.; HICKEY, J.M.; CHEN, C.; DE LOS CAMPOS, G.; BURGUENO, J.; WINDHAUSEN, V.S.; BUCKLER, E.; JANNINK, J.L.; CRUZ, M.A.L.; BABU, R. Genomic prediction in maize breeding populations with genotyping-by-sequencing. **G3-Genes Genomes Genetics**, Bethesda, v. 3, n. 11, p. 1903-1926, 2013.

DENIS, M.; FAVREAU, B.; UENO, S.; CAMUS-KULANDAIVELU, L.; CHAIX, G.; GION, J.M.; NOURRISIER-MOUNTOU, S.; POLIDORI, J.; BOUVET, J.M. Genetic variation of wood chemical traits and association with underlying genes in *Eucalyptus urophylla*.**Tree Genetics & Genomes**, Heidelberg, v. 9, n. 4, p. 927-942, 2013.

DOWNES, G.M.; HUDSON, I.L.; RAYMOND, C.A.; DEAN, G.H.; MICHELL, A.J.; SCHIMLECK, L.R.; EVANS, R.; MUNERI, A. **Sampling plantation *Eucalypts* for wood and fibre properties**. Melbourne: CSIRO Australia, 1997. 132 p.

ECKERT, A.J.; VAN HEERWAARDEN, J.; WEGRZYN, J.L.; NELSON, C.D.; ROSS-IBARRA, J.; GONZALEZ-MARTINEZ, S.C.; NEALE, D.B. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* l., *Pinaceae*). **Genetics**, Bethesda, v. 185, n. 3, p. 969-982, 2010.

ELDRIDGE, K.; DAVIDSON, J.; HARWOOD, C.; VANWYK, G. **Eucalypt domestication and breeding**. Oxford: Clarendon Press, 1993. 288 p.

ENDELMAN, J.B. Ridge regression and other kernels for genomic selection with r package rrBLUP. **Plant Genome**, Madison, v. 4, n. 3, p. 250-255, 2011.

FREEMAN, J.S.; POTTS, B.M.; DOWNES, G.M.; PILBEAM, D.; THAVAMANIKUMAR, S.; VAILLANCOURT, R.E. Stability of quantitative trait loci for growth and wood properties across multiple pedigrees and environments in Eucalyptus globulus. **New Phytologist**, Hoboken, v. 198, n. 4, p. 1121-1134, 2013.

GERALDES, A.; DIFAZIO, S.P.; SLAVOV, G.T.; RANJAN, P.; MUCHERO, W.; HANNEMANN, J.; GUNTER, L.E.; WYMORE, A.M.; GRASSA, C.J.; FARZANEH, N.; PORTH, I.; MCKOWN, A.D.; SKYBA, O.; LI, E.; FUJITA, M.; KLAPSTE, J.; MARTIN, J.; SCHACKWITZ, W.; PENNACCHIO, C.; ROKHSAR, D.; FRIEDMANN, M.C.;

WASTENEYS, G.O.; GUY, R.D.; EL-KASSABY, Y.A.; MANSFIELD, S.D.; CRONK, Q.C.B.; EHLTING, J.; DOUGLAS, C.J.; TUSKAN, G.A. A 34K SNP genotyping array for *Populus trichocarpa*: design, application to the study of natural populations and transferability to other *Populus* species. **Molecular Ecology Resources**, Hoboken, v. 13, n. 2, p. 306-323, 2013.

GION, J.M.; CAROUCHE, A.; DEWEER, S.; BEDON, F.; PICHAVANT, F.; CHARPENTIER, J.P.; BAILLERES, H.; ROZENBERG, P.; CAROCHA, V.; OGNOUABI, N.; VERHAEGEN, D.; GRIMA-PETTENATI, J.; VIGNERON, P.; PLOMION, C. Comprehensive genetic dissection of wood properties in a widely-grown tropical tree: *Eucalyptus*. **BMC Genomics**, London, v. 12, n. 301, p. 1-19, 2011.

GODDARD, M.E.; HAYES, B.J. Genomic selection. **Journal of Animal Breeding and Genetics**, Hoboken, v. 124, n. 6, p. 323-330, 2007.

GODDARD, M.E.; HAYES, B.J.; MEUWISSEN, T.H.E. Genomic selection in livestock populations. **Genetics Research**, New York, v. 92, n. 5/6, p. 413-421, 2010.

GOUY, M.; ROUSSELLE, Y.; BASTIANELLI, D.; LECOMTE, P.; BONNAL, L.; ROQUES, D.; EFILE, J.C.; ROCHER, S.; DAUGROIS, J.; TOUBI, L.; NABENEZA, S.; HERVOUET, C.; TELISMART, H.; DENIS, M.; THONG-CHANE, A.; GLASZMANN, J.C.; HOARAU, J.Y.; NIBOUCHE, S.; COSTET, L. Experimental assessment of the accuracy of genomic selection in sugarcane. **Theoretical and Applied Genetics**, New York, v. 126, n. 10, p. 2575-2586, 2013.

GRATTAPAGLIA, D. Breeding forest trees by genomic selection: current progress and the way forward. In: TUBEROSA, R.; GRANER, A.; FRISON, E. (Ed.). **Genomics of plant genetic resources**. Dordrecht: Springer, 2014. chap.26, p. 651-682.

GRATTAPAGLIA, D.; RESENDE, M.D.V. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, Heidelberg, v. 7, n. 2, p. 241-255, 2011.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic-linkage maps of *Eucalyptus-grandis* and *Eucalyptus-urophylla* using a pseudo-testcross -mapping strategy and Rapd markers. **Genetics**, Bethesda, v. 137, n. 4, p. 1121-1137, 1994.

GRATTAPAGLIA, D.; PLOMION, C.; KIRST, M.; SEDEROFF, R.R. Genomics of growth traits in forest trees. **Current Opinion in Plant Biology**, London, v. 12, n. 2, p. 148-156, 2009.

GRATTAPAGLIA, D.; VAILLANCOURT, R.E.; SHEPHERD, M.; THUMMA, B.R.; FOLEY, W.; KULHEIM, C.; POTTS, B.M.; MYBURG, A.A. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. **Tree Genetics & Genomes**, Heidelberg, v. 8, n. 3, p. 463-508, 2012.

HAYES, B.J.; VISSCHER, P.M.; GODDARD, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. **Genetics Research**, New York, v. 91, n. 1, p. 47-60, 2009.

HAYES, B.J.; BOWMAN, P.J.; CHAMBERLAIN, A.J.; GODDARD, M.E. Invited review: genomic selection in dairy cattle: progress and challenges. **Journal of Dairy Science**, New York, v. 92, n. 2, p. 433-443, 2009.

HEFFNER, E.L.; LORENZ, A.J.; JANNINK, J.L.; SORRELLS, M.E. Plant breeding with genomic selection: gain per unit time and cost. **Crop Science**, Madison, v. 50, n. 5, p. 1681-1690, 2010.

HEFFNER, E.L.; SORRELLS, M.E.; JANNINK, J.L. Genomic selection for crop improvement. **Crop Science**, Madison, v. 49, n. 1, p. 1-12, 2009.

HEIN, P.R.G.; LIMA, J.T.; CHAIX, G. Robustness of models based on near infrared spectra to predict the basic density in *Eucalyptus urophylla* wood. **Journal of near Infrared Spectroscopy**, West Sussex, v. 17, n. 3, p. 141-150, 2009.

JONAS, E.; DE KONING, D.J. Does genomic selection have a future in plant breeding? **Trends in Biotechnology**, London, v. 31, n. 9, p. 497-504, 2013.

KUMAR, S.; CHAGNE, D.; BINK, M.C.A.M.; VOLZ, R.K.; WHITWORTH, C.; CARLISLE, C. Genomic selection for fruit quality traits in apple (*Malus domestica* Borkh.). **PLoS One**, San Francisco, v. 7, n. 5, p. e36674, 2012.

LADIGES, P.Y.; UDOVICIC, F.; NELSON, G. Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. **Journal of Biogeography**, Hoboken, v. 30, n. 7, p. 989-998, 2003.

LEE, S.H.; GODDARD, M.E.; VISSCHER, P.M.; VAN DER WERF, J.H.J. Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. **Genetics Selection Evolution**, London, v. 42, n. 22, p. 1-14,2010.

LONG, N.; GIANOLA, D.; ROSA, G.J.M.; WEIGEL, K.A. Long-term impacts of genome-enabled selection. **Journal of Applied Genetics**, Heidelberg, v. 52, n. 4, p. 467-480, 2011.

LORENZ, A.J.; CHAO, S.M.; ASORO, F.G.; HEFFNER, E.L.; HAYASHI, T.; IWATA, H.; SMITH, K.P.; SORRELLS, M.E.; JANNINK, J.L. Genomic selection in plant breeding: knowledge and prospects.**Advances in Agronomy**, San Diego, v. 110, p. 77-123, 2011.

MAKOWSKY, R.; PAJEWSKI, N.M.; KLIMENTIDIS, Y.C.; VAZQUEZ, A.I.; DUARTE, C.W.; ALLISON, D.B.; DE LOS CAMPOS, G. Beyond missing heritability: prediction of complex traits. **Plos Genetics**, San Francisco, v. 7, n. 4, p. 1-9,2011.

MANDROU, E.; HEIN, P.R.G.; VILLAR, E.; VIGNERON, P.; PLOMION, C.; GION, J.M. A candidate gene for lignin composition in *Eucalyptus*: cinnamoyl-CoA reductase (CCR). **Tree Genetics & Genomes**, Heidelberg, v. 8, n. 2, p. 353-364, 2012.

MASSMAN, J.M.; GORDILLO, A.; LORENZANA, R.E.; BERNARDO, R. Genomewide predictions from maize single-cross data. **Theoretical and Applied Genetics**, New York, v. 126, n. 1, p. 13-22, 2013.

MASSMAN, J.M.; JUNG, H.J.G.; BERNARDO, R. Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize.**Crop Science**, Madison, v. 53, n. 1, p. 58-66, 2013.

MEUWISSEN, T.H.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-29, 2001.

MUIR, W.M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. **Journal of Animal Breeding and Genetics**, Hoboken, v. 124, n. 6, p. 342-355, 2007.

PARK, T.; CASELLA, G. The Bayesian Lasso. **Journal of the American Statistical Association**, Alexandria, v. 103, n. 482, p. 681-686, 2008.

PEREZ, P.; DE LOS CAMPOS, G.; CROSSA, J.; GIANOLA, D. Genomic-enabled prediction based on molecular markers and pedigree using the bayesian linear regression package in R. **Plant Genome**, Madison, v. 3, n. 2, p. 106-116, 2010.

PETROLI, C.D.; SANSALONI, C.P.; CARLING, J.; STEANE, D.A.; VAILLANCOURT, R.E.; MYBURG, A.A.; DA SILVA, O.B.; PAPPAS, G.J.; KILIAN, A.; GRATTAPAGLIA, D. Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the *Eucalyptus*genome. **Plos One**, San Francisco, v. 7, n. 9, p. 1-14,2012.

POLAND, J.; ENDELMAN, J.; DAWSON, J.; RUTKOSKI, J.; WU, S.Y.; MANES, Y.; DREISIGACKER, S.; CROSSA, J.; SANCHEZ-VILLEDA, H.; SORRELLS, M.; JANNINK, J.L. Genomic selection in wheat breeding using genotyping-by-sequencing. **Plant Genome**, Madison, v. 5, n. 3, p. 103-113, 2012.

PRYCE, J.E.; DAETWYLER, H.D. Designing dairy cattle breeding schemes under genomic selection: a review of international research. **Animal Production Science**, Collingwood,v. 52, n. 2/3, p. 107-114, 2012.

RESENDE, M.D.; RESENDE, M.F. JR.; 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, G.J. JR.; KILIAN, A.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in *Eucalyptus:* capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, Hoboken, v. 194, n. 1, p. 116-128, 2012.

RESENDE, M.D.V.; OLIVEIRA, E.B. Sistema Selegen:seleção genética computadorizada para o melhoramento de espécies perenes. **Pesquisa Agropecuaria Brasileira**, Brasília, v. 32, n. 9, p. 931-939, 1997.

RESENDE, M.D.V.; RESENDE, 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, G.J.; KILIAN, A.; GRATTAPAGLIA, D. Genomic selection for growth and wood quality in *Eucalyptus:* capturing the missing heritability and accelerating breeding for complex traits in forest trees. **New Phytologist**, Hoboken, v. 194, n. 1, p. 116-128, 2012.

RESENDE, M.F. JR.; MUNOZ, P.; ACOSTA, J.J.; PETER, G.F.; DAVIS, J.M.; GRATTAPAGLIA, D.; RESENDE, M.D.; KIRST, M. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. **New Phytologist**, Hoboken, v. 193, n. 3, p. 617-24, 2012.

RESENDE, M.F.R.; MUNOZ, P.; RESENDE, M.D.V.; GARRICK, D.J.; FERNANDO, R.L.; DAVIS, J.M.; JOKELA, E.J.; MARTIN, T.A.; PETER, G.F.; KIRST, M. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). **Genetics**, Bethesda, v. 190, n. 4, p. 1503-1510, 2012.

SILVA-JUNIOR, O.B.; FARIA, D.A.; TOGAWA, R.C.; GRATTAPAGLIA, D. *Eucalyptus* genotyping taken to the next level: development of the " EucHIP60k.br" based on large scale multi-species SNP discovery and ascertainment. In: IUFRO TREE BIOTECHNOLOGY CONFERENCE 2013, 2013, Asheville.**Proceedings…,**2013.

STACKPOLE, D.J.; VAILLANCOURT, R.E.; ALVES, A.; RODRIGUES, J.; POTTS, B.M. Genetic variation in the chemical components of *Eucalyptus globulus* wood. **G3: Genes, Genomes, Genetics**, Bethesda, v. 1, n. 2, p. 151-159, 2011.

THUMMA, B.R.; NOLAN, M.R.; EVANS, R.; MORAN, G.F. Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus spp.* **Genetics**, Bethesda, v. 171, n. 3, p. 1257-1265, 2005.

THUMMA, B.R.; SOUTHERTON, S.G.; BELL, J.C.; OWEN, J.V.; HENERY, M.L.; MORAN, G.F. Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens.* **Tree Genetics & Genomes**, Heidelberg, v. 6, n. 2, p. 305-317, 2010.

VISSCHER, P.M. Sizing up human height variation. **Nature Genetics**, New York, v. 40, n. 5, p. 489-490, 2008.

WHITTAKER, J.C.; THOMPSON, R.; DENHAM, M.C. Marker-assisted selection using ridge regression. **Genetics Research**, New York, v. 75, n. 02, p. 249-252, 2000.

WRAY, N.R.; YANG, J.; HAYES, B.J.; PRICE, A.L.; GODDARD, M.E.; VISSCHER, P.M. Pitfalls of predicting complex traits from SNPs. **Nature Reviews Genetics**, London, v. 14, n. 7, p. 507-515, 2013.