Genomic Selection for growth traits in *Eucalyptus benthamii* and *E. pellita* populations using a genome-wide *Eucalyptus* 60K SNPs chip

Bárbara S. F. Müller[1,2,3*], Leandro G. Neves[4], Márcio F. R. Resende Jr[4], Patricio R. Muñoz[5], Matias Kirst[3], Paulo Eduardo T. dos Santos[6], Estefano Paludzyszyn Filho[6], Dario Grattapaglia[1,2]

[1]University of Brasília, Molecular Biology Program, Campus Darcy Ribeiro, 70910-900, Brasília, DF, Brazil (mullerbsf@gmail.com).
[2]EMBRAPA Genetic Resources and Biotechnology, Estação Parque Biológico, 70770-910, Brasília, DF, Brazil (dario.grattapaglia@embrapa.br).
[3]University of Florida, School of Forest Resources and Conservation, Forest Genomics Laboratory, 32611, Gainesville, FL, USA (mkirst@ufl.edu).
[4]RAPiD Genomics LLC, 32601, Gainesville, FL, USA (lneves@rapid-genomics.com and mresende@rapid-genomics.com).
[5]University of Florida, Agronomy Department, 32611, Gainesville, FL, USA (p.munoz@ufl.edu).
[6]EMBRAPA Forestry Center, Estrada da Ribeira, Km 111, 83411-000, Colombo, PR, Brazil (paulo.telles@embrapa.br and estefano.paludzyszyn@embrapa.br).

Background

Genomic Selection (GS) was proposed by Meuwissen *et al.* [1] and has been gaining interest for prospective application in forest tree breeding. This predictive methodology provides an alternative to Quantitative Trait Locus (QTL) mapping and association genetics, to increase genetic gain per unit of time through early selection. GS has good potential in tree species due to their long generation time and late expressing traits as was initially shown for *Eucalyptus* [2]. The application of GS typically requires thousands of markers covering the whole-genome depending on the organism and the extent of Linkage Disequilibrium (LD). Thus, it is expected that most alleles of interest will likely be in LD with at least one or more markers genotyped and, therefore, properly captured in the predictive models. High-throughput and low-cost genotyping platforms constitute an essential tool to apply GS in forest tree breeding. Recently, SNP genotyping has become more accessible and a high-throughput SNP chip was developed for *Eucalyptus* [3]. *Eucalyptus* are the most planted species of hardwood trees worldwide for energy, pulp, paper and solid wood. *E. benthamii* is a species of interest due to its cold tolerance and high quality wood for plantation as pure species or hybrid combinations in subtropical regions of southern Brazil. *E. pellita* on the other hand

is a fast growing tropical species used in hybrid breeding with *E. grandis* providing drought tolerance and resistance to some diseases and pests. In this study we report the development of genomic prediction models for growth traits in two breeding populations of *E. benthamii* and *E. pellita.* The overall objective is to accelerate breeding cycles of these two specialty germplasm resources to provide selected individuals to be used in advanced hybrid breeding programs of *E. grandis.*

Methods

This study was carried out using two progeny trials of *E. benthamii* and *E. pellita* as training populations. These trials were developed as part of the breeding program of EMBRAPA Forestry. For *E. benthamii* the trial involved 40 treatments, being 36 open-pollinated maternal families from wild Australian populations and 4 samples from mixed seed sources. The trial was composed by ~2,000 trees that were phenotyped at age 4 years and 8 months. The experimental design was a randomized complete block with 50 blocks in single-tree plots, with the 40 treatments and 50 repetitions per treatment, planted in May 2007 at Candói, PR, Brazil. The *E. pellita* trial was composed by 24 open-pollinated maternal families from a 1st generation seed orchard in Australia with 960 trees. The experimental design was a randomized complete block composed by the 24 progenies and 40 trees per progeny in single-tree plots planted in Rio Verde, GO, Brazil. In *E. benthamii* growth traits were measured on 506 trees: Diameter at Breast Height (DBH, cm), Total Height (HT, m) and Wood Volume (WV, m^3). For *E. pellita* only DBH was measured for 769 trees. In total 569 trees of *E. benthamii* and 772 trees of *E. pellita* were genotyped at GeneSeek (Neogen Corporation, Lincoln, NE) using the *Eucalyptus* Infinium EucHIP60k.br [3]. The genotypic data was filtered to remove SNPs with call rate ≤ 90% and a Minimum Allele Frequency (MAF) ≤ 0.05. These markers were used to construct a Genomic Relationship Matrix (GRM) following Powell's method [4]. Individual SNPs had their effects estimated adjusting all the allelic effects simultaneously using the genomic BLUP (GBLUP) approach. A 10-fold cross-validation approach was performed where 90% of the total population was used as a training population and 10% as the validation population. The selection gain of GS (GBLUP) was compared with traditional phenotypic selection (Pedigree BLUP, PBLUP) considering a reduced breeding cycle as a result of early selection.

Results and Conclusions

A total of 12,177 SNPs for *E. benthamii*, with a median of 10 individuals per family, were used to build prediction models for GS. The effective population size was estimated from SNP data as N_e ~51. Accuracies of Genomic Estimated Breeding Values (GEBVs) were 0.32 for DBH and 0.33 for WV. For *E. pellita* 18,605 SNPs were used and a median of 32 individuals per family. Accuracy of GEBV was 0.62 for DBH and the estimated N_e ~35. The selection responses of GS, assuming a conservative reduction of 50% in the length of the breeding cycle, were 237% and 239% for DBH and WV in *E. benthamii* respectively highlighting the increase in efficiency per unit of time. For *E. pellita* the traditional selection (PBLUP) could not be estimated due a detected inconsistency between the expected pedigrees based on family information and the realized pedigree based on the GRM built from SNP data. This result indicates that the original seeds allegedly from maternal families in fact were most likely mixtures from several families possibly due to tree labeling or seed collection errors in the original seed orchard. This observation revealed an additional advantage of using SNP data not only for genomic prediction but also to correct pedigree information for conventional breeding. This study demonstrated that GS could dramatically increase genetic gains per unit of time of these *Eucalyptus* breeding programs and aid in the correct estimation of quantitative genetic parameters.

Acknowledgments

References

1. Meuwissen TH, Hayes BJ, Goddard ME: **Prediction of total genetic value using genome-wide dense marker maps.** *Genetics* 2001, **157**:1819–29.

2. Resende MD V, Resende MFR, Sansaloni CP, Petroli CD, Missiaggia A a, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria D a, Pappas GJ, 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 Phytol* 2012, **194**:116–28.

3. Silva-junior OB, Faria DA, Grattapaglia D: **Methods A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species**. 2015.

4. Powell JE, Visscher PM, Goddard ME: **Reconciling the analysis of IBD and IBS in complex trait studies.** *Nat Rev Genet* 2010, **11**:800–805.