P1186

Predictive Ability of Genomic Estimated Family Values (GEFV)

Esteban Rios[1], Marcio Resende[2], Matias Kirst[1], Marcos D. V. Resende[3], Janeo E. de Almeida Filho[4] and Patricio Munoz[1]

[1]University of Florida, [2]RAPiD Genomics LLC, [3] EMBRAPA-Brazil, [4]Federal University of Viçosa - Brazil

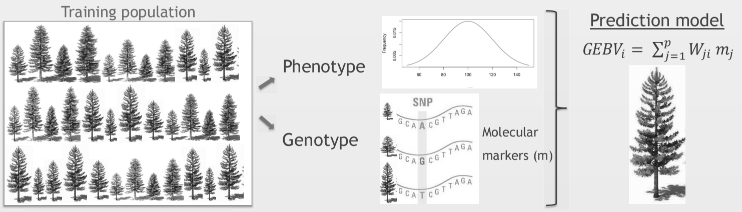
Abstract

Genomic selection (GS) has been used to compute genomic estimated breeding values (GEBV) of individuals; however, it has only been applied to animal and major plant crops due to high costs. Besides, breeding and selection is performed at the family level in some crops. We aimed to study the implementation of genome-wide family selection (GWFS) in two loblolly pine (*Pinus taeda* L.) populations: i) the breeding population CCLONES composed of 63 families (5-20 individuals per family), phenotyped for four traits (stem diameter, stem rust susceptibility, tree stiffness and lignin content) and genotyped using an Illumina Infinium assay with 4740 polymorphic SNPs, and ii) a simulated population that reproduced the same pedigree as CCLONES, 5000 polymorphic loci and two traits (oligogenic and polygenic). In both populations, phenotypic and genotypic data was pooled at the family level *in silico*. Phenotypes were averaged across replicates for all the individuals and allele frequency was computed for each SNP. Marker effects were estimated at the individual (GEBV) and family (GEFV) levels with Bayes-B using the package BGLR in R and models were validated using 10-fold cross validations. Predicted ability, computed by correlating phenotypes with GEBV and GEFV, was always higher for GEFV in both populations, even after standardizing GEFV predictions to be comparable to GEBV. Results revealed great potential for using GWFS in breeding programs that select families, such as most outbreeding forage species. A significant drop in genotyping costs as one sample per family is needed would allow the application of GWFS in minor crops.

Introduction

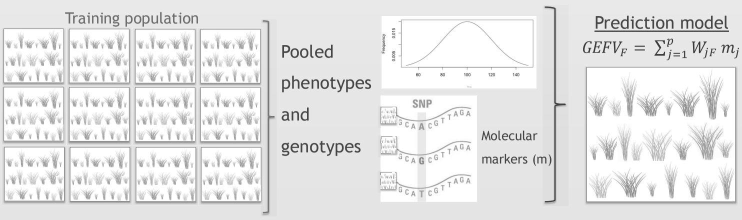
GS is used to compute GEBV of individuals (Figure 1), but it has only been applied to breeding animals and major crops that can invest in high costs.

Figure 1. GS scheme to predict breeding values of individuals



In some plant crops such as most outbreeding forages, breeding and selection is performed at the family level and GWFS could be implemented in these crops (Figure 2).

Figure 2. GWFS scheme to predict breeding values of families



We aimed to study the implementation of GWFS in two loblolly pine populations: the experimental breeding population CCLONES (Resende et al., 2012) and a simulated population (de Almeida Filho et al., 2016).

Materials and Methods

Predictive ability for individual and family GS models was computed using two populations:

1. The loblolly pine experimental breeding population CCLONES (Comparing Clonal Lines On Experimental Sites), composed of 63 full-sib families (5 to 20 individuals per family). Four traits (stem diameter, stem rust susceptibility, tree stiffness and lignin content) were phenotyped and genotypic data was obtained using an Illumina Infinium assay, resulting in 4740 polymorphic SNPs.
2. A simulated population that had similar genetic properties as CCLONES. Initially, 1000 diploid individuals were created (G0) by randomly sampling 2000 haplotypes from a population with effective size (N_e) of 10000 and mutation rate of 2.5×10^{-8}. Secondly, the 10% highest phenotypic values from G0 were selected and randomly mated to generate the first breeding generation (G1). From G1, 42 individuals were selected and used in a mating design that reproduced the same pedigree as CCLONES (G2), which was used in this study. The simulated genome had 5000 polymorphic loci and two traits were simulated with different genetic architectures: i) oligogenic: 30 QTL were sampled from a gamma distribution with rate 1.66 and shape 0.4, with positive or negative QTL effects, and ii) polygenic: 1000 QTL were used and their additive effects were sampled from a standard normal distribution.

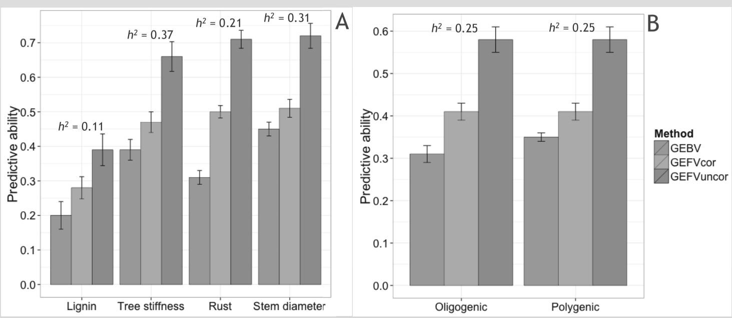
In both populations, phenotypic and genotypic data was pooled at the family level *in silico*. Phenotypes were averaged across replicates for all the individuals per family and allele frequency was computed for each SNP. Marker effects were estimated at the individual (GEBV) and family (GEFV) levels with Bayes-B using the package BGLR 1.0.4 (Pérez and de los Campos, 2014) in R and models were validated using 10-fold cross validations. In order to compare the predictive ability computed with individual and family models, a correction factor ($\surd 0.5$, the parentage between full sibs) was imposed to GEFV to get the individual accuracy via family selection.

Results

Predicted ability was obtained through simple correlation analysis between the observed phenotypes and the GEBV and GEFV, using the correction factor ($\surd 0.5$) (GEFVcor) and without correcting (GEFVuncor) to get the individual accuracy via family selection.

Predicted ability was always higher for GEFV in both populations and all traits, even after standardizing GEFV to be comparable to GEBV (Figure 3 A and B).

Figure 3. Predictive ability of four traits with different heritability in the loblolly pine breeding population CCLONES (A) and of two traits with different genetic architecture in a simulated population (B).



Conclusion

GWFS models revealed great potential for using GWFS in breeding programs that select families to develop cultivars, as is the case of most outbreeding forage species such as alfalfa (*Medicago sativa* L.), and annual (*Lolium multiflorum* Lam.) and (*L. perenne* L.) perennial ryegrasses.

The significant drop in genotyping costs in GWFS, as one sample per family is needed instead of n samples per family would enable the application of GWFS in minor crops, especially in those that phenotyping is performed at the plot level.

References

de Almeida Filho, J., J. F. Guimarães Rodrigues, F. Fonseca e Silva, M. D. V. Resende, P. Muñoz, M. Kirst and M.F. R. Resende Jr. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. Submitted.

Pérez, P., G. de los Campos. 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics 198: 483–495.

Resende, M. F. R. Jr., P. Muñoz, M. D. V. Resende, D. J. Garrick, R. L. Fernando, J.M. Davis, E. J. Jokela, T. A. Martin, G. F. Peter and M. Kirst. 2012. Accuracy of genomic selection methods in a standard dataset of Loblolly pine (*Pinus taeda* L.). Genetics 190: 1503–1510.

International Plant & Animal Genome XXIV / January 9-13, 2016 - San Diego, CA, USA