Genomic prediction in a multi-generation breeding population: maritime pine as a case study

Jérôme Bartholomé, Christophe C. Boury, Joost van Heerwaarden, Fikret Isik, Christophe Plomion, Laurent Bouffier

Genomic prediction in a multi-generation breeding population: maritime pine as a case study

Jérôme Bartholomé[1,2], Christophe Boury[1,2], Marjorie Vidal[1,2], Joost van Heerwaarden[3], Fikret Isik[4], Christophe Plomion[1,2], Laurent Bouffier[1,2]

1/ INRA, UMR1202, BIOGECO, Cestas F-33610, France, 2/ Univ. Bordeaux, BIOGECO UMR1202, Talence F-33170, 3 France, 3/ Biometris, Wageningen University and Research Centre, Wageningen NL-6700 AC, Netherlands, 4/ Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA

Context

➤ **Maritime pine breeding program**
- Base population selected in the Southwest of France during the 60's(~700 G0 trees)
- From this base population, recurrent breeding scheme:
  - double-pair matings and forward selection to create the next breeding generation
  - polycross matings and backward selection to select the best genotypes for seed orchards
- Currently reaches its 3rd generation (breeding population effective size ~ 135)
- Three main selection criteria:
  - circumference (h² ~ 0.17)
  - height (h² ~ 0.32)
  - stem straightness (h² ~ 0.26)

➤ **Breeding values evaluation**
- Data base with > 400,000 trees over three generations (G0, G1, G2) connected with a pedigree matrix
- BLUP evaluation = multi-traits, multi-sites, multi-generations (Treeplan evaluation system)

Objectives

➤ **Genomic selection proof-of-concept for maritime pine**

Low number of markers available (~3 markers / cM) ➔ population of limited Ne

➤ **Comparison of genomic predictions vs. pedigree predictions**

Materials and Methods

➤ **Reference population**: **G2** trees + their ancestors (**G0** and **G1**)

46 G0
62 G1
710 G2

G2 selected for high growth and high stem straightness with a diversity constraint (maximum contribution of a single founder limited to 0.15)
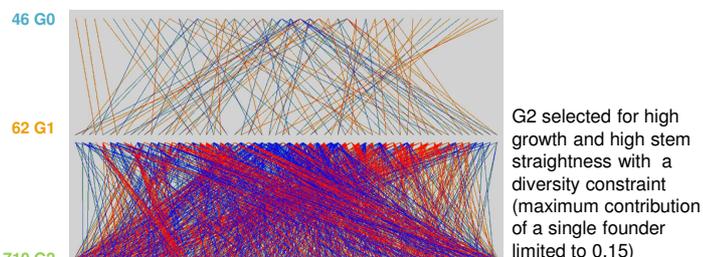
➔ **818 trees (Ne=24)**
(Pedigree checked with parentage recovery analyses based on 80 SNPs)

➤ **Pseudo phenotypes**: estimated breeding values (EBV) from BLUP evaluation for circumference, height and stem straightness

➤ **Genotyping**: 4,330 SNPs (only 0.05% of missing data)

➤ **Genomic predictions**:
- 4 validation models ➡
- 3 genomic prediction methods (GBLUP, Bayesian Ridge Regression, Bayesian LASSO)

Model	Training	Validation
Model_1	G0	G2
Model_2	G1	G2
Model_3	G0+G1	G2
Model_4	G2 (80%)	G2 (20%)

Results

➤ **High accuracy for genomic prediction between and within generations**
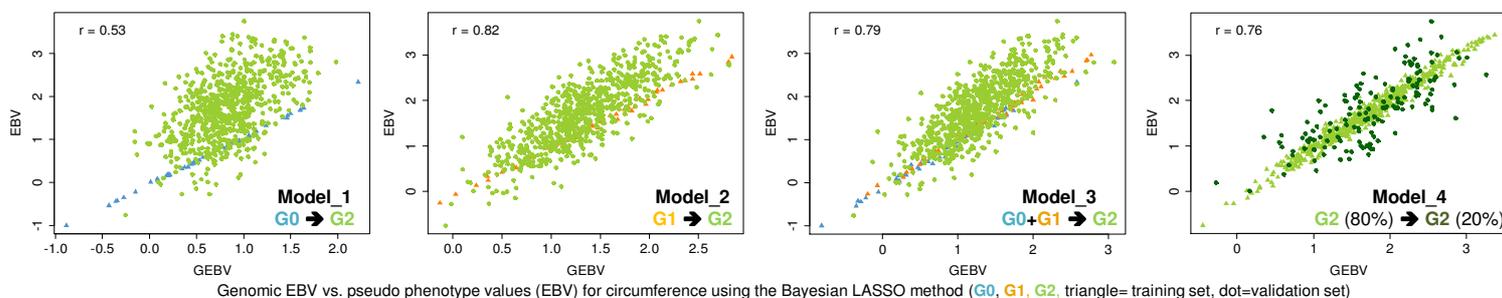
Genomic EBV vs. pseudo phenotype values (EBV) for circumference using the Bayesian LASSO method (G0, G1, G2, triangle= training set, dot=validation set)

➤ **Coefficient of numerator vs. realized relationship matrix in the reference population (n=818)**

➤ **Similar accuracy between pedigree and genomic predictions**

EBV: estimated breeding value (pseudo phenotype)

MPBV: mid-parent breeding value

GBLUP: best linear unbiased prediction based on realized relationship matrix (Model_3 considered)
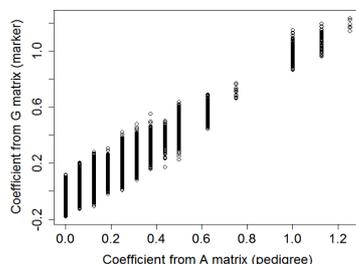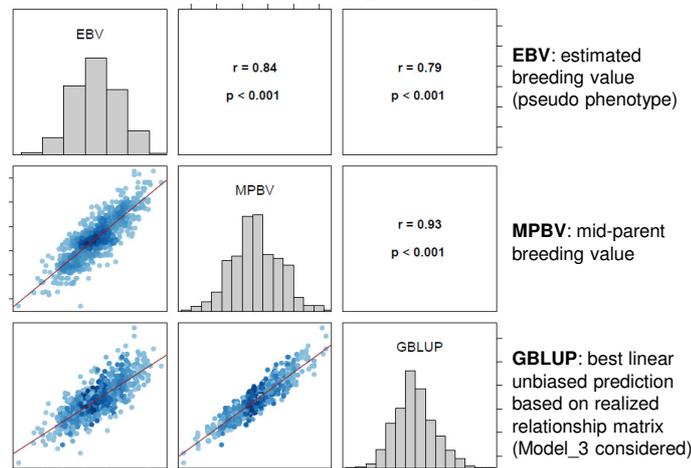
Conclusions

➤ **High accuracy for genomic predictions**

➤ **Inclusion of previous generations does not improve GS models**

➤ **Similar accuracy between pedigree and genomic predictions**
- Full pedigree relationships known through 3 generations
- Poor estimates for G2 pseudo phenotypes (no clonal or progeny data)

INRA SCIENCE & IMPACT — NC STATE UNIVERSITY