

Correction for measurement error from genotyping-by-sequencing in genomic variance and genomic prediction models

Bilal H. Ashraf, Luc L. Janss, Just Jensen, Torben Asp and Christian S. Jensen

Centre for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University Denmark
DLF-Trifolium, Store Heddinge, Denmark

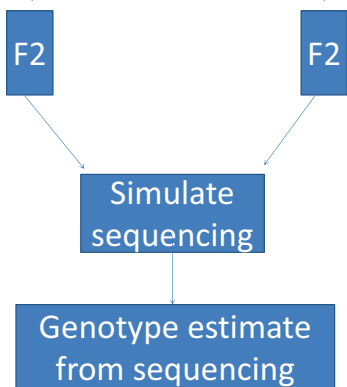
Aim

Prediction from family pools, where genotypes in a Pool are determined by sequencing

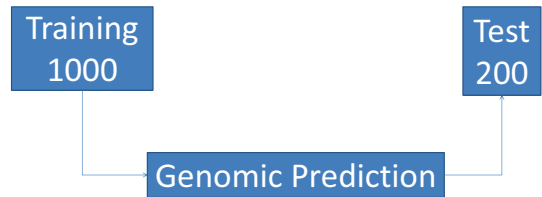
Methods

Generate F2 pools from 4 inbred wheat genomes

(1279 SNPs, dataset: R Package BLR, CIMMYT Mexico)



Statistical Models



Training: $Y_{Train} = \mu + \sum_i f_i b_i + e$

Test: $\hat{g} = \sum_i f_i \hat{b}_i$ (genomic prediction)

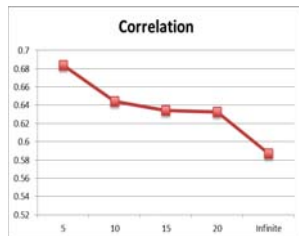
f_i = frequency for genotypes from sequences

Comparison: $\text{cor}(Y_{pred}, \hat{g})$

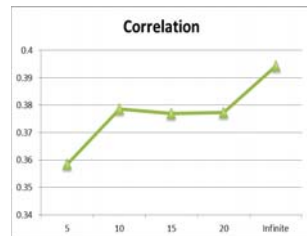
Regression $Y_{pred} \sim \hat{g}$

Results

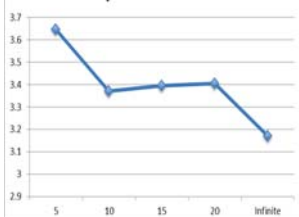
Training Data



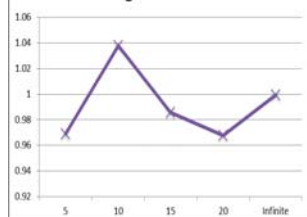
Testing Data



Explained Variance



Regression bias



Conclusions

- We can predict phenotype of a pool
- Low sequencing depth causes overfit and lower prediction
- Low sequencing depth causes underestimation of allele effect, this underestimation can be determined by:

$$\text{Bias} = \frac{1}{1 + \frac{3}{S_T}}$$

- Genomic prediction offer tools to develop new breeding design