Performance of Genomic Prediction using Different Multi-Allelic Genomic Relationship Matrices for Genotyping-by-Sequencing (GBS) Data

Jie Kang[1,3], Phillip Wilcox[1], Michael Black[1], Stephen Byrne[2], Dan Milbourne[2], Marty Faville[3], Joanne Jacobs[3], Ken Dodds[3]

1. Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand

2. Teagasc, Crop Research Centre, Oak Park, Carlow, Ireland

3. AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand

Introduction

- Genomic prediction (GP) has been frequently implemented using GBS data since its introduction almost a decade ago.
- Genomic relationship matrices (GRM) can be derived from GBS data, and used with genomic best linear unbiased prediction (GBLUP) to perform GP.
- Short haplotypes or 'ShortHaps', that is, multiple variants situated in small genomic segments such as those captured within GBS reads.
- ShortHaps are theoretically more advantageous than single nucleotide polymorphism (SNP) because they are more likely to be strongly associated with causal variants. MA-GRMs constructed using shortHaps are expected to be superior.

Aim

Investigate whether predictive ability of genomic prediction can be further improved by using multi-allelic genomic relationship matrices (MA-GRM).

Method

MA-GRM

1. Haplotype-dosage method (Hesselness et al., 2014)

$$MAGRM_1 = HH' / (2p(1-p))$$

where H denotes the matrix of centred haplotype score (dosage), and p denotes the haplotype allele frequencies.

2. Haplotype similarity method (Ferdosi et al., 2016)

$$MAGRM_2 = KTK' / 2$$

where K is the Kronecker product between an n x n identity matrix

(n = number of individuals) and a 1 x 2 unit matrix; T is the genetic relationship matrix.

3. Linkage disequilibrium method (Mathew et al., 2018)

$$MAGRM_3 = HS^{-1}H'$$

where S is a block-diagonal matrix containing the (averaged) multinomial variance-covariance matrix of haplotypes.

MA-GBLUP

$$y = X\beta + Zu + \varepsilon$$

where y is a matrix of phenotypes, X and Z are incidence matrices, β and u are fixed and random effects with Var(u) = $G\sigma_u^2$ and ε is the error term. In particular, G can be any

Simulation

- ApeKI GBS data of 96 ryegrass samples (with a known pedigree, Fig. 1) were simulated using simGBS (Kase et al., 2018), based on the ryegrass reference genome described in Byrne et al. (2015).
- 448,100 out of 1,860,744 virtually digested GBS tags/fragments were retained after fragment size-selection at (65, 135) bp. 306,807 tags contained SNPs (2 SNPs per tag on average) with 676,890 SNPs in total.
- 1000 QTL from a quantitative trait with heritability of 0.1, 0.4, 0.7, 0.9 were simulated (1000 replicates of each scenario), with the QTL effects and residuals drawn from standard normal distributions.

Figure 1: Pedigree used in simulation

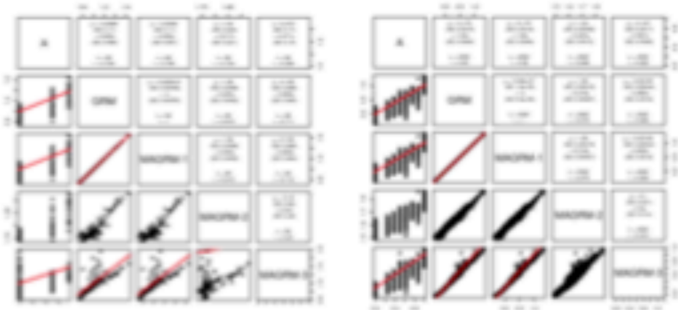

Figure 2: Pairwise comparisons of diagonal (left) and off-diagonal (right) between the relationship matrices

Results & Discussion

- The pedigree, genotype and three haplotype-based relationship matrices returned consistent results (Fig. 2), where off-diagonals and diagonals of GRMs are highly correlated (apart from the diagonals of MAGRM-2).
- Performance of GBLUP using different GRMs are almost identical (Fig. 3). However, the results may be constrained by the population size and make-up (including SNP density and linkage disequilibrium between SNPs).
- Current simulation assumed GBS data has infinite depth, the impacts of low sequencing depth and other GBS-specific sources of variation needs to be considered. Future experiments will compare the performance of biallelic with MA-GRM approaches in these situations.

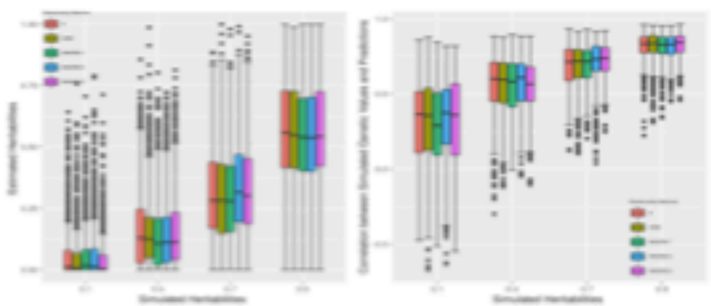

Figure 3: Performance of GBLUP using five relationship matrices 1. estimated heritabilities (left) 2. correlation between simulated genetic values and predictions (right)

Conclusion

Despite a strong correlation between MA-GRMs and other relationship matrices, the use of MA-GRMs did not significantly improve predictive accuracy of GP under current settings.

References

Hesselness et al. (2014). Genetics Selection Evolution, 46, 1-8.
Ferdosi et al. (2016). Genetics Selection Evolution, 48, 75.
Mathew et al. (2018). Heredity, 120, 356-368.
Kase et al. (2018). Proc 11th World Congr Genet Appl to Livest Prod, 388
Byrne et al. (2015). Theoretical and Applied Genetics, 124, 709-720.