Effect of preselection of whole-genome sequence variants for kinship estimation

Eduard Molinero, Joan Estany, Roger Ros-Freixedes

Departament de Ciència Animal, Universitat de Lleida – Agrotecnio, 25198 Lleida, Spain. Presenting author: eduard.molinero@udl.cat

Introduction

- Population structure can be accounted for by fitting a genomic kinship matrix in the mixed model.
- The kinship matrix is typically calculated from the set of genotypes available from marker arrays.
- As whole-genome sequence data becomes more affordable, the number of available genotypes increases.
- The computational requirements do not scale well and therefore there is a need to preselect variants from sequence data to calculate kinship.

Materials and Methods

- 146 pigs sequenced at 8.0x (SD 2.2x).
- Genotypes from 250k variants with high genotype certainty.
- Subsets of variants were preselected based on:
  - linkage disequilibrium (to retain 10k, 25k, 50k or 100k variants)
  - minor allele frequency (all, MAF>0.05, or MAF>0.25).
- Kinship matrices calculated with standardized genotypes with each subset.
- Kinship estimates of related animals were compared to the matrix calculated with all markers and the pedigree-based matrix.

Results

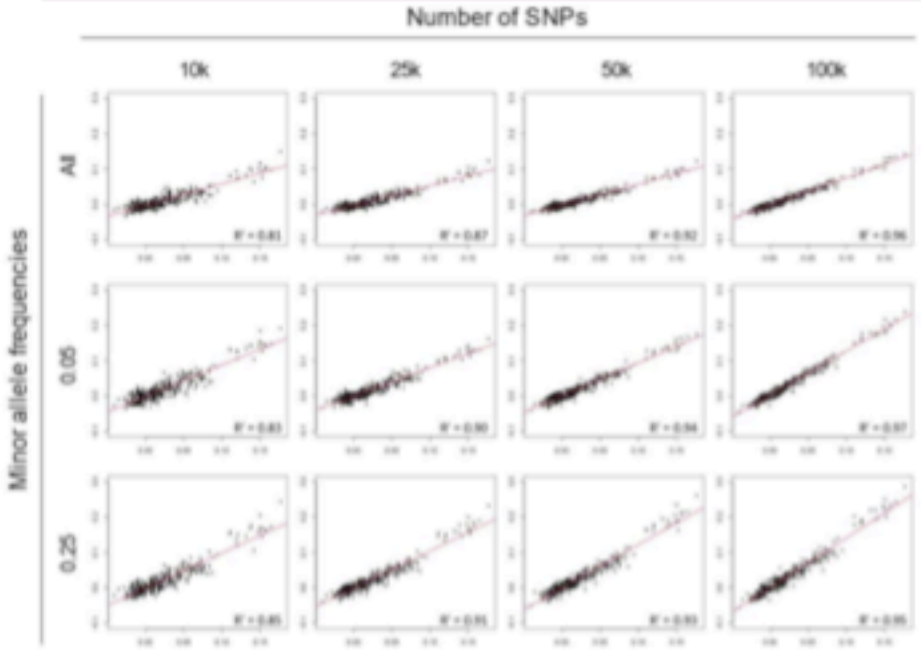

Figure 1. Linear regression between kinship estimates calculated with all variants (x) and with a subset of variants preselected based on linkage disequilibrium and minor allele frequencies (y)

Figure 2. Comparison between kinship estimates calculated with a pedigree (x) and with a subset of variants preselected based on linkage disequilibrium and minor allele frequencies (y)

Conclusion

- As long as genotypes with high certainty are used, genomic kinship estimation is quite robust to the preselection of whole-genome sequence variants.

Acknowledgments

E Molinero: PhD scholarship by University of Lleida & Banco Santander.

Project grant **RTI2018-101346-B-I00**