The Effects of Population Parameters and Genetic Models on Power and Accuracy of GWAS

Zhi Loh[1*], Sam Clark[1], Julius H. J. van der Werf[1]

[1] School of Environmental and Rural Science, University of New England, Armidale NSW, Australia
* zloh@myune.edu.au

Introduction

- Genome Wide Association Studies (GWAS) are popular to identify causal variants for a trait.
- Besides the effect of sample size and markers density, relatively few studies have looked at the effects of population parameters and genetic models on the power and false positive rate of GWAS.
- Population parameters includes selection strength and effective population size, while genetic models are characterized by number and sizes of QTLs, heritability and non-additive genetic components.

Aim

To investigate the effect of population parameters and genetic models on power, proportion of false positives and accuracy of estimated QTL effect sizes in GWAS

Genetic Model

(Based on Duenk et al, 2020)

$$y = G_{a_k}\alpha_k + G_{d_k}\delta_k|\alpha_k| + G_{\gamma_{kl}}\gamma_{kl}\sqrt{|\alpha_k\alpha_l|} + e$$

y	Phenotype
G_{a_k}	Additive Genotype Value; set at (0, 1, 2)
α_k	Additive Genetic Effect Size for locus k
G_{d_k}	Dominance Genotype Value; set at (0, 1, 0)
δ_k	Dominance Coefficient
$G_{\gamma_{kl}}$	Epistatic Genotype Value
γ_{kl}	Epistatic Coefficient for loci k and l
e	Residual Value

Distribution Used:
$\alpha_k \sim N(0,1)$
$\delta_k \sim N(0,\delta)$
$\gamma_{kl} \sim N(0,\gamma)$

- α_k was then adjusted with narrow sense heritability
- Only $\alpha_k \geq 0.1$ were included in result calculation.

Parameter Tested

Parameter	Default Value	Alternative Value
Number of Generations	15	35
Proportion of Selected Males	0.5	0.2
Effective Population Size	1200	200
Narrow Sense Heritability	0.3	0.15
Number of QTL	100	10000
Sample Size	2000	800
SD of Dominance Coefficient (δ)	0	1.0
SD of Epistatic Coefficient (γ)	0	0.7
Allele Frequency	Continuous Variable	

Calus, M. P. L., Meuwissen, T. H. E., de Roos, A. P. W., and Veerkamp, R. F. (2008) Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. Genetics, 178(1), 553-561. doi:10.1534/genetics.107.080838

Duenk, P., Bijma, P., Calus, M. P. L., Wientjes, Y. C. J., and van der Werf, J. H. J. (2020) The Impact of Non-additive Effects on the Genetic Correlation Between Populations. G3: Genes|Genomes|Genetics, 10(2), 783-795. doi:10.1534/g3.119.400663

Layout of Algorithm

(Based on Calus et al., 2008)

Ancestral Population

Phase I
- Number of Generations: 20000
- Effective Population Size: 2000
- Number of Markers: 20000
- Species: *Equus caballus* (domestic horse)
- Recombination and Mutation
- No selection

Founder Population

Phase II
- Number of Generations: 15 or 35
- Effective Population Size: 1200
- Selection
- Final Generation 2400 animals

Final Population

GWAS
- Sample Size: 2000 or 800
- Single SNP Regression
- Bonferroni Correction
- Threshold (alpha): 0.10

Result

Result Recorded	Definition
Power	$\frac{TP}{TP+FN}$
True Positive Rate	$\frac{TP}{TP+FP}$
Mean Squared Error (MSE) of Estimated Effect Sizes	$\frac{\sum(\alpha_k - \widehat{\alpha_k})}{nqtl}$

TP: True positives; FP: false positives; FN: False negatives; α_k: True effect sizes; $\widehat{\alpha_k}$: Estimated effect size; nqtl: Number of QTLs

Results

Parameters	Power		True Positive Rate		MSE of Estimated Effect Sizes	
	Default Value	Alternative Value	Default Value	Alternative Value	Default Value	Alternative Value
Number of Generations	0.039	0.035*	0.708	0.763	0.043	0.030**
Proportion of Selected Males	0.039	0.039	0.692	0.478***	0.042	0.044
Effective Population Size	0.039	0.042*	0.706	0.437***	0.044	0.043
Narrow Sense Heritability	0.039	0.002***	0.694	0.610***	0.037	0.166***
Number of QTL	0.038	0.000***	0.683	0.514**	0.042	0.602***
Sample Size	0.040	0.008***	0.673	0.760*	0.037	0.107***
SD of Dominance Coefficient	0.039	0.041***	0.681	0.658	0.042	1.302***
SD of Epistatic Coefficient	0.040	0.112***	0.686	0.374***	0.044	2.768***

Significance Remarks:
* Weakly significant ($1e-3 \geq p > 1e-4$); **Significant ($1e-4 \geq p > 1e-10$); ***Strongly Significant ($p \leq 1e-10$)

Effect of Allele Frequency

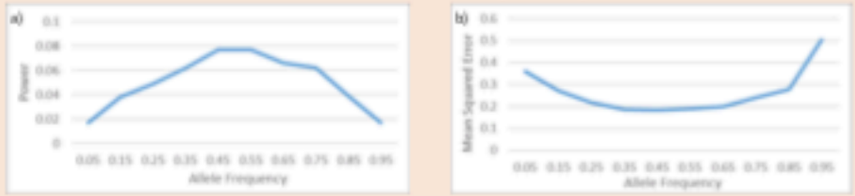

Figure 1: The effect of allele frequency on (a) power and (b) mean squared error of estimated QTL effect size. Allele frequency do not have significant effect on True Positive Rate of GWAS.

Conclusion

- Low selected male proportion and effective population size decrease true positive rate of GWAS.
- Large number of QTL and low heritability decrease the power and true positive rate of GWAS, while increase the MSE of estimated QTL effect sizes.
- With exception of dominance on true positive rate of GWAS, non-additive genetic components such as epistasis and dominance increase the power, but decrease the true positive rate and accuracy of estimated QTL effect sizes