

USE OF GENOTYPE PROBABILITIES AND SELECTIVE GENOTYPING FOR ESTIMATION OF MARKER EFFECTS

M.L.A.N.R. Deepani and B.P. Kinghorn

The Institute for Genetics and Bioinformatics, University of New England, Armidale, NSW

SUMMARY

Cost of genotyping associated with marker assisted selection is reduced by partial genotyping. Genotype probabilities (GPs) and selective genotyping can both be used to alleviate the reduction in power that partial genotyping can bring. This paper tests the effect of the combined strategies on the estimation of a single marker effect. The results suggest that with selective genotyping the marker estimates become less biased when genotype probabilities are used and when animal relationships are included in the selection function.

INTRODUCTION

Marker assisted selection (MAS) aims to give higher response to selection by increasing accuracy of genetic evaluations (Hayes *et al.* 2007). The cost of genotyping associated with MAS is reduced with partial genotyping and use of genotype probabilities for non-genotyped animals. Baruch *et al.* (2008) used the segregation analysis method of Kerr and Kinghorn (1996) to help estimate the effects of two QTL for a population in which only the bulls had been genotyped. Selecting animals for genotyping on the basis of their phenotypes (“Selective Genotyping”, Lander *et al.* (1989); Darvasi and Soller (1992)) also reduces genotyping cost while managing the loss of precision in detecting a QTL effect. This study combines these two strategies, and extends this by considering animal relationships as well as phenotypes when selecting animals to genotype.

MATERIALS AND METHODS

A population was simulated starting with 30 males and 150 females in different age classes. For the foundation animals a bi-allelic SNP marker was simulated with allele frequency 0.5 and allele substitution effect of 6. Alleles were sampled from a uniform distribution, and then propagated in a Mendelian pattern. The mean, phenotypic standard deviation and polygenic heritability for the single trait of interest were assumed to be 100, 30 and 0.25 respectively. At the end of ten years of random selection and mating, the marker effect was estimated using BLUP in which the marker genotype or its probabilities were fitted as a covariable.

There were two major experiments for treatments. Experiment 1 was partial genotyping with random selection of animals for genotyping (Rnd). Under Experiment 1 there were 5 treatments. Treatment 1 used 100% genotype information and this was used as a control. Treatments 2 to 5 used 50%, 35%, 20% and 5% genotyping with random selection of animals for genotyping, and genotype probabilities from Kerr and Kinghorn (1996) for the ungenotyped animals. Treatments 2 to 5 all included sub-treatments with different thresholds of genotype probability index (GPI, Kinghorn, 1997) as minimum levels for accepting animals into the BLUP analysis. A threshold of 100% means that only animals with 100% GPI were included. Low GPI animals tend to be poorly connected to genotyped individuals and have poor quality genotype probabilities.

Experiment 2 involved selective genotyping, using the method introduced by Kinghorn *et al.* (2006). This uses an evolutionary algorithm to select animals on an objective function of two components: “Distance” - a measure of phenotypic dispersal (following Davarsi and Soller 1992) across traits, and “Relationship” – a metric of both decreased relationship between phenotypically similar animals (the contrast between phenotypically divergent groups is weakened by a small

number of parents contributing to each group) and increased relationship between phenotypically dissimilar animals (which helps contrast QTL effects, as in some sib- pair designs). “Relationship” is aimed at cleaner designs that reduce the number of false positive calls.

Three treatments were: full emphasis on “Distance” denoted by “D”; full emphasis on “Relationship” denoted by “R”, and equal emphasis on each, denoted “DR”, which used a function to target the averages of the D and the R results achieved. As for Experiment 1, the marker effects were estimated with the same five levels of genotyping and with or without GPs, with sub treatments using different GPI thresholds when GPs are used. Twenty replicates were run for each treatment.

RESULTS AND DISCUSSION

Experiment 1. Random selection and use of genotype probabilities. It was observed that estimated marker effects were close to the true value when GPI thresholds up to 30% were used compared with using information on genotyped animals only (“No GP”, Figure 1). Standard deviations (SDs) between replicates reflect the standard error of one replicate and hence accuracy of estimation. SDs were higher for estimates obtained without using GPs (No GP) than with GPs. The minimum SD was obtained when all records were used (All GPI), even those with very low GPIs. At 5% genotyping, SD was ±3.775 with “No GP” while it was ±3.398 with “All GPI”.

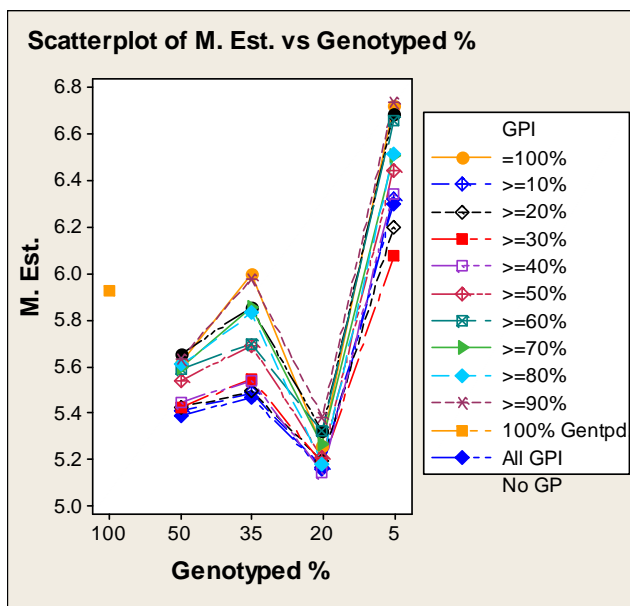


Figure 1. Marker effect estimates (M. Est.) with different proportions of genotyping.

Experiment 2. Selective Genotyping. Under selective genotyping, the estimates obtained were biased upwards in comparison to random selection, as predicted by Darvasi *et al.* (1992). This overestimation was reduced with the inclusion of relationship together with distance as a selection criterion (DR and R, Figure 2). SDs also reduced when the relationship was taken into account, while minimum SD observed when equal emphasis was given to Distance and relationship (DR). At 5% genotyping SD of estimated marker effect under DR, with “All GPIs” is 3.667 which is very close to the SD under random selection (3.398, Figure 2) although the marker effect is overestimated.

The results for selective genotyping were corrected according to the following equation;

$$\delta = D_T / \gamma_p \quad \text{Equation 1}$$

Where “ D_T ” is the raw estimated marker effect by selective genotyping, “ δ ” is the corrected estimate of the marker effect and $\gamma_p = 1 + Z_{1-p/2} \cdot i_{p/2}$. Where “ Z ” denotes the truncation point, “ p ” is

the proportion of the animals that have been genotyped and “i” is the selection intensity (Darvasi *et al.* 1992).

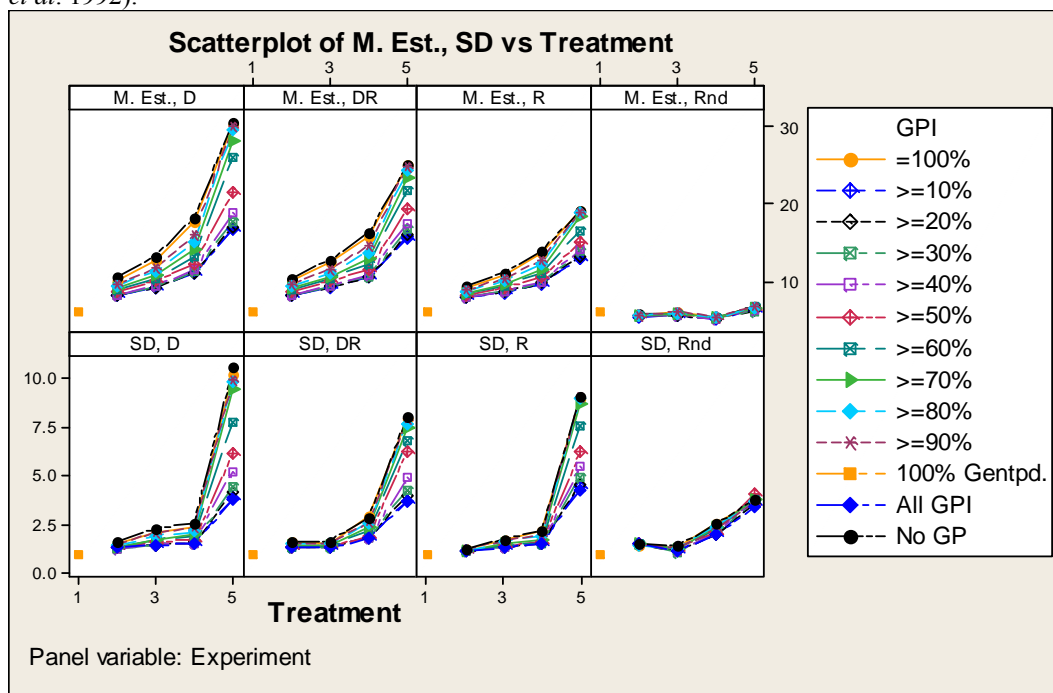


Figure 2. Comparison of marker effect estimates(row 1) and standard deviations (row 2) obtained from Experiment 1(Rnd) and Experiment 2; D, DR and R.

When the correction described above was made to the estimates, they were overcorrected especially at low selection proportions. This was true for “No GP” (Figure 3) as seen by (Muranty *et al.* 1997). Under selective genotyping the magnitude of overestimation is reduced by using GPs, which partly fill the missing information, making the strategy more similar to genotyping all animals, whence no correction is needed. This is why equation 1 results in the over correction seen in Figure 3.

Under D, the corrected results for “No GP” were less biased compared to the “No GP” under DR and R (Figure 3, upper row). This may be because for D, the selection is purely based on the phenotypic distance and this is where the formula for correction more precisely applied. For DR and R relationship was also taken into account and the formula is not designed to accommodate that criterion. The SDs for corrected marker effects were smaller (< 2) in all 3 selective genotyping methods compared to the standard deviation of the marker estimates obtained from random selection of animals (Figure 3), indicating higher accuracy.

Under random selection of animals for genotyping, marker effect estimates are more accurate if genotype probabilities are used; with increased accuracy related to the population average GPI. Under selective genotyping, including relationships in the selection function gives less bias results compared to phenotypic dispersal alone, and bias is further reduced if GPs are also used. Although Darvasi and Soller’s equation tends to overcorrect this reduced bias, it more substantially reduces the SD of the estimate compared to random selection. Therefore accommodating relationships and

GPs increases the accuracy of estimating a positive marker effect even though it is underestimated after simple correction using equation 1.

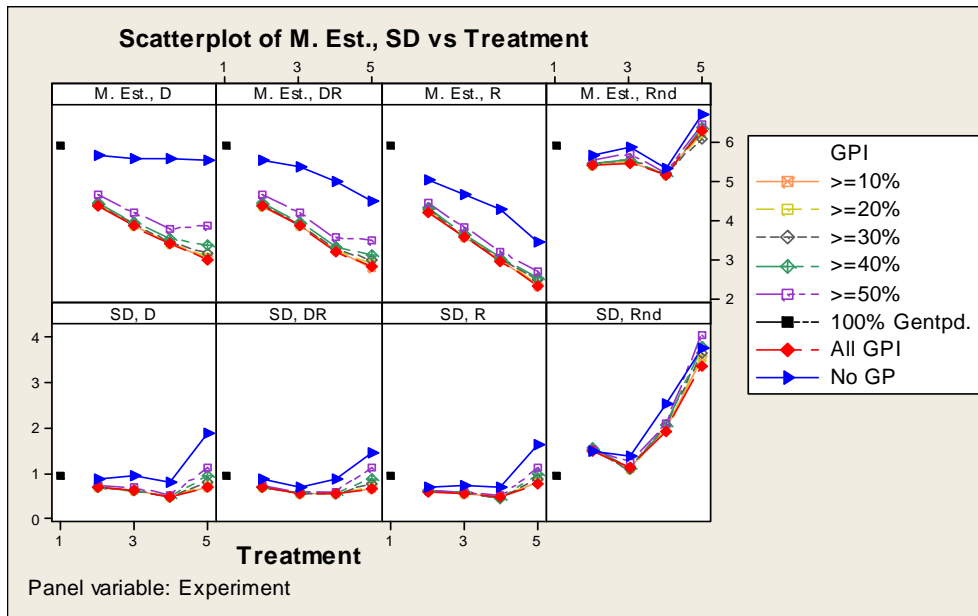


Figure 3. Corrected marker effect estimates (row 1) and their standard deviations (row 2) in selective genotyping (D, DR and R) compared with random selection (Rnd).

CONCLUSIONS

Accommodation of relationships as carried out in this paper reduces the bias in marker estimates that are caused by correlations with polygenic effects. Use of genotype probabilities gives further improvement, by including animals that would not otherwise be used in the analysis. Bias due to relationships is more important than bias due to selective genotyping *per se*, because the latter affects all markers in a similar manner, such that ranking of markers is less affected. Therefore, this combined approach should help when choosing discovery project markers for validation. In such cases, general trend in estimation bias is less critical than capturing markers that in fact have true effects.

ACKNOWLEDGEMENT

To the IRQUE project of the Faculty of Veterinary Medicine and Animal Science University of Peradeniya, Sri Lanka for funding this study.

REFERENCES

- Baruch, E., J. I. Weller (2008) *J. Dairy Sci.* **91**: 4365.
 Darvasi, A., Soller, M. (1992) *Theoretical and Applied Genetics* **85**: 353.
 Hayes, B. J., Chamberlain, A. J. *et al.* (2007) *Genetics Research* **89**: 215.
 Kerr, R. D., Kinghorn, B.P. (1996) *J. Anim. Breed. Genet.* **113**: 457.
 Kinghorn, B.P. (1997) *Genetics* **145**:479.
 Kinghorn, B.P., Bastiaansen, J.W.M., van der Steen, H.A.M., Deeb, N., Yu, N., and Mileham, A.J. (2006) *Proc. 8th Wrld. Cong. on Gen. Appld. Livst. Prod.*, Paper 20-05. 4 pages.
 Lander, E. S., Botstein, D. (1989) *Genetics* **121**: 185.
 Muranty, H., Goffinet, B. (1997) *Biometrics* **53**(2): 629.