Incorporation of Genotype Effects into Animal Model Evaluations when only a Small Fraction of the Population has been Genotyped

E. Baruch and J. I. Weller

ARO, The Volcani Center, Bet Dagan, Israel

Animal Model

How identified QTL can be used in selection program ?

Traditional approach: progeny test.
Genetic gains of ~0.1 phenotypic standard deviations per year can be obtained in modern progeny test schemes for a trait with a heritability of 0.25.

Alternative Approach
Estimate breeding value of all individuals in the population, but not directly estimate the QTL.

Most modern dairy cattle breeding programs are based on progeny test scheme, and genetic evaluations are generally derived by application of the "animal model".

Once a segregating quantitative trait locus (QTL) has been detected via linkage to genetic markers, application is difficult.

•Gene-marker phase varies among individuals.

•Only a small fraction of the population is genotyped.

•Necessary to correctly weigh marker, pedigree, and trait information

The basic problem for MAS

Only a very small fraction of the recorded population will have genotypes for genetic markers.

Thus the model of Fernando and Grossman (1988) cannot be directly applied.

Most studies have proposed to analyze DYD or EBV of genotyped animals, but the properties of these statistics are not well defined.

Fernando and Grossman (1989) proposed a "gametic" model extension to the AM that assumes that the two QTL alleles of each individual are random effects sampled from a distribution with a known variance.

Breeding values are estimated for all individuals in a population, including QTL effects via linkage to genetic markers.

[This method is suitable for any population structure, accommodates the fact that QTL-marker phase is generally unknown, and also can incorporate non-linked polygenic effects and other "nuisance" effects, such as herd or block.

The disadvantages are that the method assumes that all animals have been genotyped, and that both recombination frequency and the variance due to the QTL are known *a priori.*]

Model of Israel and Weller (1998)

- ➢ All animals with records are included in the analysis.
- ➢ The QTL effect is considered fixed.
- ➢ The QTL effect for individuals that were not genotyped is the probability of receiving each possible genotype, based on their relationships to the genotyped animals.

The model of Israel and Weller (1998) assumes complete linkage between the QTL and a single marker, and only two QTL alleles are segregating in the population.

The model further assumes that either a daughter or granddaughter design has been applied to determine QTL genotypes of the family ancestors. The QTL effect is then included in the complete animal model analysis as a fixed effect.

For individuals that are not genotyped, probabilities of receiving either allele are included as regression constants. These probabilities can be readily computed for the entire population using the segregation analysis method of Kerr and Kinghorn (1996).

Model of Israel and Weller (1998)

$$Y_{ijk} = a_i + p_i + h_j + m_k + q + e_{ijk}$$

- Additive Effect
- Permanent Environmental Effect
- Fixed HYS Effect
- Lactation Effect
- Fixed QTL Effect
- Residual Error Effect

Estimation of QTL effects

☑ Simulations of daughter and
   granddaughter design:
   Unbiased estimates of QTL
   effects

❌ Actual Data
   QTL effects biased downward
   (*Weller et al., 2003*)

The method of Israel and Weller (1998, 2002) has been tested extensively on simulated populations, and was able to yield virtually unbiased estimates of QTL effect, even though only 25% of the individuals were genotyped. Two and three generation populations were analyzed. However, when this model was applied to actual data from the Israeli Holstein population for the *DGAT1* locus segregating QTL on chromosome 14 that affected milk production traits (Grisart *et al.*, 2002), the QTL effect was strongly underestimated relative to alternative estimation methods.

Reasons for discrepancies between results on simulated and actual data

Simulations do not adequately describe actual data:

- Smaller fraction of genotyped individuals (25% vs. 1%)
- Lower rare allele frequency (0.2 vs. 0.1)
- More generations included in the analysis (3 vs. 8)

The actual data set differed from the simulated data sets in three aspects. A much smaller fraction of the total population was genotyped in the actual data, less than 1% of the total population; frequency of rare allele was very low, about 10%, in the actual data; and the actual data included about 8 generations, while the simulated data included on 2-3 generations.

Effect of allele frequencies on the estimate of QTL substitution effect

Initial frequency of negative allele	Estimated QTL effect	Standard deviation
0.1	0.12	0.06
0.2	0.21	0.08
0.3	0.31	0.05
0.4	0.37	0.06
0.5	0.39	0.07
0.6	0.42	0.05
0.7	0.39	0.06
0.8	0.42	0.04
0.9	0.45	0.05

Bias is greatest when the initial frequency of the negative allele is lowest, making it more difficult to accurately estimate the QTL effect. Bias also increased with the number of generations included in the simulations

Hypothesis

Bias is apparently due to confounding between the QTL genotype probabilities and the relationship matrix.

Apparently, as the fraction of animals with inferred genotypes increases, the genotype probabilities tend to "mimic" the effect of relationships.

[QTL effect was underestimated in all cases, but bias was greater for extreme allelic frequencies, and increased with the number of generations included in the simulations.]

Estimate of QTL substitution effect
(without relationship)

$$Y_{ijk} = c_i + h_j + m_k + q + e_{ijk}$$

- **Cow Effect**
- Fixed HYS Effect
- Fixed Lactation Effect
- Fixed QTL Effect
- Residual Error Effect

KEEP OUT
Relationships

Cow Effect = sum of additive polygenic and permanent environmental effect..

This model differs from the model of Israel and Weller (1998) in that only cows with production records are included, and covariance among cow effects are assumed to be zero. That is the relationship matrix is not included. This will referred as the cow model.

Estimate of QTL substitution effect
(without relationship)

- Relationship between individuals was not included in estimation of the polygenic effect.

- Thus the permanent environmental and polygenic effect were combined into a single "cow effect," and only animals with records were included in the analysis.

11

Estimation of QTL substitution effects from the cow model

Initial frequency of negative allele	Estimated QTL effect	Standard deviation
0.1	0.50	0.24
0.2	0.54	0.10
0.3	0.52	0.08
0.4	0.51	0.10
0.5	0.51	0.08
0.6	0.53	0.06
0.7	0.47	0.08
0.8	0.52	0.06
0.9	0.57	0.05

Results of the Cow Model QTL effect estimates as a function of initial QTL allelic frequencies are presented in this table. Estimates of the QTL effect were nearly unbiased for all initial QTL frequencies. As expected, the standard deviations increased with reduction of the initial frequency of the negative allele, the allele whose frequency decreases during selection.

Application

Although the "cow model" can be used to derive unbiased estimates of the QTL effect, it cannot be used to derive EBV, because the relationship matrix is not included.

Although the cow model can be used to obtain unbiased estimates of QTL effects, it cannot be used for routine genetic evaluation, which requires incorporation of the relationship matrix. Therefore, the following algorithm was devised:

Practical Algorithm for EBV calculation

1. Compute genotype probabilities for animals with unknown genotypes.
2. Estimate QTL effects by cow model.
3. Subtract the known or inferred QTL genotype effect, based on the cow model estimated QTL effect, from the cows' production records.
4. Compute EBV for all animals from the adjusted cow records.
5. Summing the EBV with the inferred or known QTL effect of each animal.

Selection of bull calves

Mean of 20 best calves
(based on ten simulations)

	Traditional Model	Modified Model
EBV	0.453	0.494
Records	0.469	0.603

The mean EBV of the 20 best calves selected by this model is higher by 0.12 SDU from the mean of the 20 best calve according to the initial model. Alternatively, suppose that the SD for milk production of a given herd is 1000 kg, then the advantage of the this scheme is 150 kg higher at year 10.

Gain in QTL effect by year
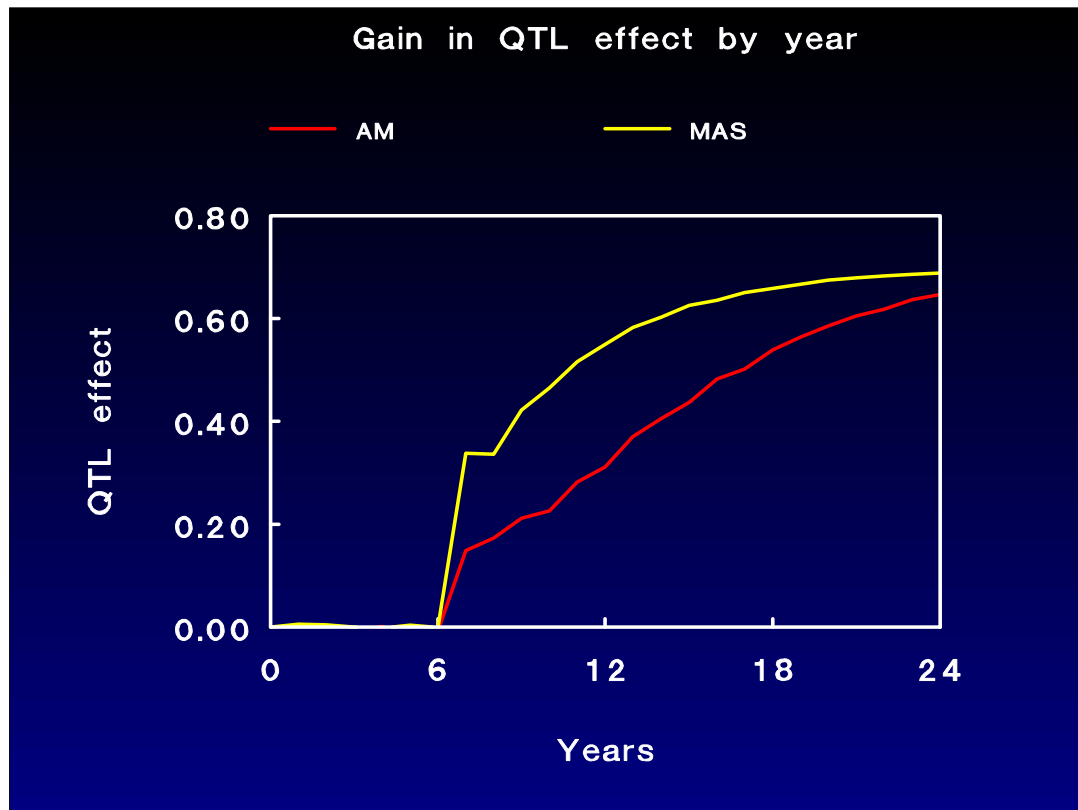
The gain of 0.15 by MAS at year 10 represents an increase of nearly 20% over the traditional scheme.

The initial frequency was 0.3 and the total gain is 0.7.
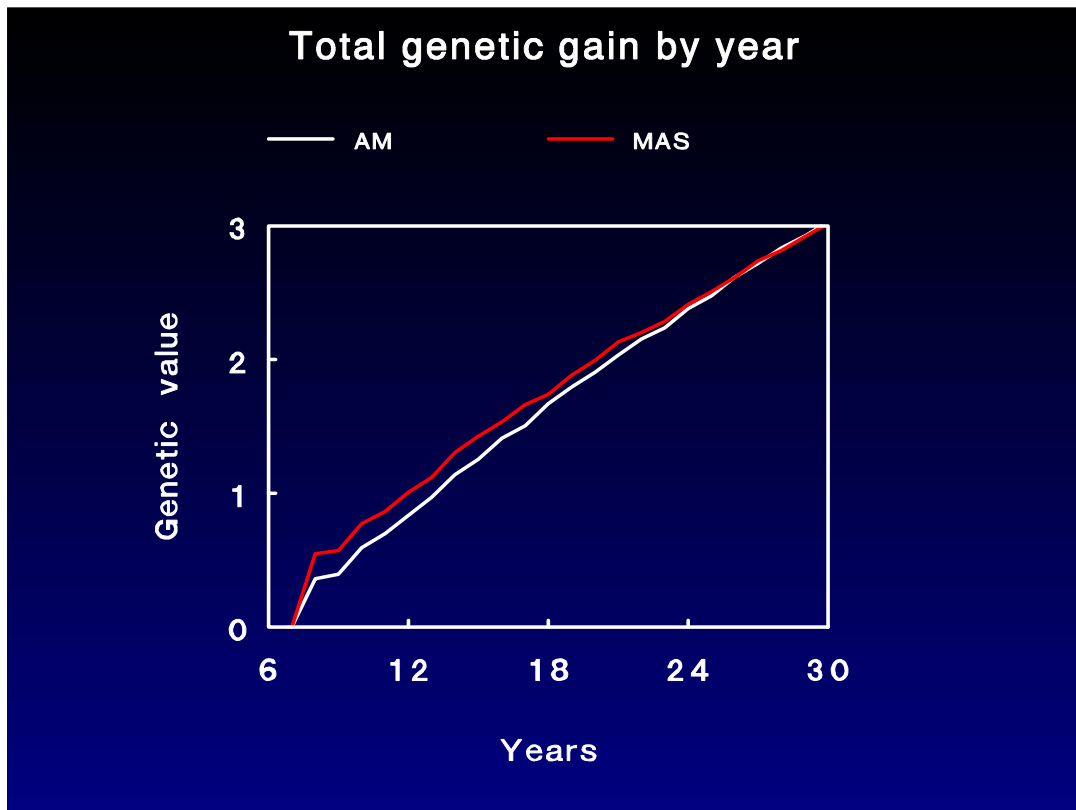
Total and polygenic genetic gain and genetic gain in mean QTL value as a function of year of simulation.

Conclusions

1. At year 10 total genetic gain was 20% greater by the proposed algorithm.
2. Even at year 30 the MAS scheme is still higher than the traditional scheme.
3. Thus genetic gain with trait-based selection did not surpass MAS over the long-term as noted by Gibson (1994).