Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters

W.M. Muir

Department of Animal Sciences, Purdue University, West Lafayette, Indiana, USA

Keywords
- genomic selection
- BLUP
- GEBV

Correspondence
W.M. Muir, Professor Genetics, Department of Animal Sciences, Room G406 Lilly Hall, Purdue University, W. Lafayette, IN 47907, USA.
Tel: 765-494-8032;
E-mail: bmuir@purdue.edu

Received: 31 May 2007; accepted: 5 October 2007

Introduction

The attraction of biotechnology for breeding programmes is to directly utilize genotypic information, termed marker-assisted selection (MAS), with the intent of greatly enhanced genetic progress, both in speed and accuracy. Translation from concept to application has been attempted for many decades, first with blood groups (e.g. Neumann-Sorensen & Robertson 1961). The advent of direct DNA markers, such as DNA fingerprints and restriction fragment length polymorphisms (RFLP) (Southern 1975, 1982; Selsing et al. 1976) and polymerase chain reaction (PCR)-based methods (Saiki et al. 1988) led to further advances of MAS methods. These methods have become widely acknowledged and used in breeding programmes (Aitken et al. 1994; Boichard et al. 1998; van Marle-Koster & Nel 2003; Muir 2003; Frey et al. 2004; Liu & Cordes 2004).

These technical advancements were bolstered by theoretical results presented by Lande & Thompson (1990) who showed that when MAS is used with...
traits of low heritability ($h^2$), the combined efficiency could be over 300% of that achievable by breeding based on the phenotype alone. This was an exciting result because least heritable traits are exactly where classical quantitative genetics approaches need help. However, upon further reflection it was obvious that finding quantitative trait loci (QTL) for traits of low $h^2$ had the same problem as quantitative genetics, i.e. the signal-(genetics)-to-noise (environment) ratio is very poor and finding QTL for such traits is equally difficult. Thus the prospect of finding even a moderate amount of the genetic variation accounted for by markers is not realistic in those situations.

Nevertheless, there are a number of other traits with moderate heritabilities where MAS could be beneficial, such as sex-limited traits, traits that are expensive to measure, or can only be measured on relatives (Meuwissen & Goddard 1996). However, implementation of MAS in breeding programmes is problematic in that separate markers are usually needed for each trait. If the markers are linked to the QTL, rather than the causative quantitative trait nucleotide (QTN), linkage phase variants cause the makers to be incorrect in some families or breeds. Finally, pleiotropic effects of linked QTL on other traits must also be considered (see Dekkers & Hospital 2002 for review). Breeders need a robust method of selection that can be used for all traits, across families, and in an optimal manner taking into account genetic correlations among traits.

One possible solution is genomic selection based on genome-wide predicted breeding values (GEBV), which was first proposed by (Meuwissen et al. 2001) and later variations suggested by Xu (2003) and ter Braak et al. (2005). This theoretical development coincides with the new single nucleotide polymorphism (SNP) technology which is high throughput, accurate, and relatively inexpensive. The concept of genomic selection is to estimate effects of all markers simultaneously, without first subjecting to variable selection (model building) with associated type II errors. In addition, by fitting allelic effects as random, rather than fixed effects, the problem of degrees of freedom for estimating the effects is overcome. Genomic selection requires dense markers spaced across the genome (equal spacing being optimal without prior knowledge of QTL positions), thereby taking advantage of all available genetic variation in population-wide linkage disequilibrium (LD) with those markers. Conceptually, the more dense the markers and greater the population-wide LD, the better the method works. Thus, for the first time there appears to be a combined solution that addresses both theoretical and technical issues of MAS in breeding programmes.

However, there remains a number of unanswered technical questions regarding GEBV that need to be examined before advancing to implementation. Using simulations Meuwissen et al. (2001) examined the efficiency of GEBV under a number of restricting assumptions, such as equally spaced QTL always centered between two markers, and other assumptions which may not be possible or valid, such as assuming commercial populations will be in a mutation-drift equilibrium (MDE) and a trait with $h^2 = 0.5$. Other unresolved issues include: (i) least heritable traits, (ii) distributional assumptions regarding maker and QTL frequencies, (iii) the number of generations of data needed on phenotype and genotype for accurate prediction, (iv) importance of new versus existing LD between markers and QTL, (v) marker density and number, (vi) number of animals genotyped per generation, and (vii) impact of selection on new and existing LD. Most of these issues will be examined in a preliminary non-exhaustive manner.

**Materials and methods**

A gene level simulation programme previously developed by Muir (1997, 2000) was utilized for this study. This programme allows for any number of loci, number of alleles per locus, gene action, proportion selected from each sex, initial gene frequency distribution, environmental effects distribution, population structure and size, and mutation rate ($u$). Gene effect distributions can be chosen from normal, negative exponential, and uniform. For these simulations, gene effects for each allele at a locus were assigned randomly from a normal distribution with a mean zero and unit variance, with additive effects within and between the loci. A normal distribution for allele effects was used because of programme limitations, a gamma distribution was utilized by Meuwissen et al. (2001) but differences in these distributions are subtle, the gamma being more versatile and depending on the parameters used, can resemble distributions from the negative exponential to normal. Regardless, both normal and gamma result in rare alleles with large effects and common alleles with small effects, but the gamma is skewed such that there are more alleles with small effects than the normal. Thus simulation presented here will tend to have a few more alleles with larger effects. These minor differences in gene effect distribution assumptions are not expected to
change substantially the main conclusions of the research.

The genetic variance was computed as the variance among genotypes. Given the genetic variance, environmental effects were generated from a normal distribution with a mean of zero and a variance set to give the desired specified heritability. The phenotype was assumed to be the sum of the genotypic and environmental effects, i.e. no genotype–environment interactions.

Meuwissen et al. (2001) used $h^2 = 0.5$; 1000 generations of size 100 to establish MDE, following which two generations of training with 200 and 2000 animals were genotyped and phenotyped respectively, and 2000 animals genotyped only in the third generation (one generation of prediction based only on genotypes); seven alleles per locus (50 haplotypes per two loci); 10 chromosomes each of 100 cM, and a marker density of 1/cM. Generations for which both genotypes and phenotypes are available were termed training generations (TG) because the data is being used to train the model to differentiate random noise from genetic signal. Generations beyond the TG are the prediction generations because only the independent variables are obtained (the marker information) and no data is generated for the dependent variables (phenotypes). Because of limitations of the computer memory, this study could not accommodate that total number of genotypes, thus the simulations presented here are not directly comparable, but we were able to confirm a number of their conclusions. However, the scale of the simulations was shown to be primarily an issue as related to long-term selection response.

For the research presented here, genome size ranged from 1 to 10 000 cM randomly allocated to three chromosomes. The number of QTL loci was always set to 100 and randomly spaced throughout the genome. Marker loci were set equally distant at a range of spacing between 0.1 and 100/cM. The number of alleles per marker locus was set at two, simulating SNP. Initial allele frequency was simulated in two different ways: (i) randomly set at each locus (marker and QTL) between 0 and 1 based on the uniform distribution and in multi-locus Hardy-Weinberg equilibrium (HWE) both within and between the loci, i.e. genotypic and gametic frequencies were directly proportional to their allelic frequencies, resulting in both zygotic and gametic phase equilibrium and (ii) in MDE. The MDE was established using similar assumptions as Meuwissen et al. (2001) but with a greater number of generations to establish equilibrium. Specifically, a completely inbred population of 128 animals, with equal numbers of both sexes, were randomly mated for 5000 generations with $u = 2.5 \times 10^{-5}$ per QTL locus and $u = 2.5 \times 10^{-3}$ per marker locus. These were the same assumptions as given by Meuwissen et al. (2001) except they used a population size of $N_e = 100$ and 1000 in the first two generations to train the model. The high mutation rate for marker loci may be justified for microsatellites which are known to be hypervariable (Dallas 1992), but is problematic for SNP, which have an estimated $u = 2.5 \times 10^{-9}$ in mammals (Kumar & Subramanian 2002). Impact of these assumptions on the mutation-drift frequency distribution will be addressed in the discussion section.

A population in two locus HWE is a worst case for LD as all population-wide LD is generated new as a result of the drift process during the TG. In contrast, a population in MDE will have pre-existing LD for those loci remaining polymorphic at both the QTL and marker loci and are closely linked. This results because a new mutant allele initially exists in only one individual in complete LD with the chromosome, and then spreads through the population, or becomes extinct, by the drift process. However, because a large number of generations of random mating, and recombination, must occur before the mutant allele can become common through the drift process, only those mutations in close proximity with the QTL will remain in LD. Furthermore, only those marker–QTL combinations that have become common will be useful to MAS. Definition of common and uncommon allele frequencies is relative to the population size being genotyped. For example, if only 1000 animals are genotyped, a marker with an allele frequency of 0.01 may be common enough to estimate effects, depending on the number of total loci being genotyped and co-linearity issues, but clearly the same allele frequency in a population of 100 would be too uncommon to be useful.

Neither of these situations (HWE or MDE) is likely to be realistic in actual breeding programmes. All breeding programmes include selection and finite population sizes, both of which cause domestic populations to differ greatly from the assumptions of HWE and MDE. For example, MDE is a concept applied to historical ancestral populations in which loss of genetic diversity because of genetic drift (inbreeding) is perfectly balanced by new mutations. As a result, the majority of allelic diversity in an MDE population is rare and close to fixation. In practice, the reduction in effective population size from that of the ancestral population, as a result of
the domestication process, was shown to result in a minimum loss of 60% of all allelic diversity in broilers and 80% in layers as compared with the ancestral MDE population (Muir & Cheng 2007). A consequence of this loss in allelic diversity is also the loss of historic haplotype LD block structures established through new mutations. These smaller blocks merged into larger LD blocks.

The real issue is how large are the haplotype blocks in various species. The optimal length of an LD block is an intermediate. If the block is very small, as in humans (Altshuler et al. 2005), then a high SNP density is needed to capture the LD. If the LD blocks are very large, as in partially inbred lines, then fewer SNP would be needed, but then there would not be as much genetic variation to capture. These are similar issues faced when deciding on a resource population for coarse versus fine mapping QTL, i.e. F2 versus F10 but in this case the genetic variation is constant between generations while haplotype blocks are broken down by recombination. In the case of an ancestral population in MDE, haplotype blocks are constantly broken down by recombination but regenerated by new mutations. But the loss of historic haplotype LD block structures was variable. For the first 10 generations all males and females, i.e. effective population size (Wright 1968) ranged from 16 to 128. The evaluation and training populations of which between 16 and 128 were used to genotype and phenotype. Then in the second 10 generations, only genotypes were available and phenotype (1 ≤ t ≤ 5), these were the TG. Then in the next 10-t generations, only genotypes were obtained and EBV predicted. For the first t generations, individuals were chosen for breeding randomly, while in the following 10-t generation individuals were chosen either at random (Random) or directionally (Direct) selected. Directional selection was based on breeding values estimated either by traditional BLUP (Henderson 1984), or as total genetic merit (Equation 3). For traditional BLUP, an infinitesimal model was assumed; all pedigree

Best Linear Unbiased Prediction (BLUP), but to avoid confusion with traditional BLUP (Henderson 1984) we will call this method Ridge Regression (RR). Nevertheless, RR is quick, easy to program, and as Meuwissen et al. (2001) demonstrated, RR performed almost as well as the much more advanced and time-consuming Bayesian methods. Meuwissen et al. (2001) showed that the accuracy of RR was 0.732 while BayesB was 0.848, or performed 86% as well. Their result suggests that RR should be a good indicator of what is possible with more advanced methods of GEBV. Thus, although RR is not optimal, it presents a conservative assessment of what is possible with GEBV. However, results of Meuwissen et al. (2001) comparing RR with BayesB were for their specific simulation conditions and assumptions and may not extrapolate well to the other scenarios.

The mixed model equation (MME) for RR is

\[ Y = X\beta + ZG + \epsilon. \]  

(1)

The MME are:

\[ \begin{bmatrix} X'X & X'Z & X'Z + \frac{n\sigma_e^2}{\sigma_g^2}I \\ Z'Z + \frac{n\sigma_e^2}{\sigma_g^2}I & \frac{n\sigma_e^2}{\sigma_g^2}I & \frac{n\sigma_e^2}{\sigma_g^2}I \end{bmatrix} \begin{bmatrix} \beta \\ G \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}. \]  

(2)

where \( X \) is a vector of 1s, \( G \) is a vector of genetic effects due to the \( i \)th marker allele or haplotype, \( Z \) is an incidence matrix and has a 0, 1, and 2 for the number of alleles of type \( G_i \) present in the \( j \)th animal.

Total merit, or GEBV, was estimated as

\[ \hat{Y}_j = \sum_i Z_jG_i. \]  

(3)

Population size ranged from 128 to 1024 individuals of which between 16 and 128 were used to generate the next generation with an equal number of male and females, i.e. effective population size \( (N_e) \) ranged from 16 to 128. The evaluation and training structure was variable. For the first \( t \) generations all individuals were both completely genotyped and phenotyped \( (1 \leq t \leq 5) \), these were the TG. Then in the next \( 10-t \) generations, only genotypes were obtained and EBV predicted. For the first \( t \) generations, individuals were chosen for breeding randomly, while in the following \( 10-t \) generation individuals were chosen either at random (Random) or directionally (Direct) selected. Directional selection was based on breeding values estimated either by traditional BLUP (Henderson 1984), or as total genetic merit (Equation 3). For traditional BLUP, an infinitesimal model was assumed; all pedigree
information was included for all generations, phenotypic data from the TG were incorporated into the MME, and missing values were used for the 10-t generations beyond the TG. True values for the genetic and environmental variance in the base population were used in the MME. The assumption was made, for both BLUP and GEBV, that total genetic and environmental variances remained constant. This assumption is known to be false under selection for both models.

Each simulation was replicated 30–60 times starting over each replicate with the same genetic architecture and in the case of HWE, different initial allele frequencies, or in the case of MDE, the same initial haplotype frequencies. Because of the large number of factors examined, not all combinations were included. Comparisons between methods were based on the correlation (accuracy) between the estimated and true breeding values.

Results and discussion

The MDE

The theoretical distributions of gene frequencies for neutral alleles was developed by Wright (1930, 1931, 1935, 1937) and showed that these distributions could be described in terms of $4N_e \mu$. When $4N_e \mu$ is close to 1 a uniform distribution results, but when $4N_e \mu < 1$, a U-shaped distribution emerges. The U-shaped distribution occurs because of the balance between inbreeding, driving alleles to fixation, and new mutations, driving allele back in. The U-shaped distribution implies that the majority of the loci have a minor allele frequency close to zero.

Figure 1a gives the distribution that results for $N_e = 100$ after 1000 generations and $\mu = 2.5 \times 10^{-3}, 2.5 \times 10^{-4}, 2.5 \times 10^{-5}$ (the number of loci for these simulations was increased to 2000). The effective population size and mutation rates correspond to those of the makers and the QTL in these simulations in Meuwissen et al. (2001). For markers a uniform distribution results, as predicted by Wright for $4N_e \mu = 1$, but for the lowest mutation rate, that associated with QTL, a J-shaped distribution resulted. If the number of generations of random mating is increased to 75 000, a more U-shaped distribution does emerge (Figure 2), but an excess of 100 000 generations would be needed to produce the characteristic U-shaped distribution. Nevertheless, the proportion of polymorphic loci in generations 1000 through 75 000 remained fairly constant at 4%, thus 1000 generations is adequate to establish a steady-state equilibrium for these simulations, if not true MDE.

These results may provide an important clue to differences in accuracy for GEBV in simulations using HWE versus MDE as starting conditions. A population in MDE has only 4% as many QTL–marker associations as a population in HWE. As such, there are many fewer real effects to estimate and potential to be confounded with other effects, so accuracy is expected to be better for a population in MDE, regardless of other factors.

These results also point to the issue of scale for simulations. With only 4% of loci segregating, a simulation with 10 QTL loci could not capture the distribution. A simulation with at least 100 loci is needed, at minimum, to capture the first significant digit of the frequency distribution, and will have approximately four QTL loci and 100 marker loci segregating. The simulation of Meuwissen et al. (2001) would have had approximately 40 QTL and 1000
marker loci segregating. Thus no substantial differences are expected between simulations with either 100 or 1000 loci if selection is not considered, to a first approximation they will both give similar conclusions. However, for a selected population, the rate of change in allele frequency is proportional to the amount of genetic variance accounted for by each locus (Lynch & Walsh 1998). Thus rate of fixation and changes in genetic variances will be greater with small number of QTL. Therefore comparisons of response to selection with traditional BLUP versus GEBV for populations in MDE with few segregating QTL may only be valid for a few generations of the selection (discussed in a later section).

Wright did not discuss the minimum \( N_e \) needed to produce the theoretical distributions, but differing number of generations are needed to achieve the same distribution for different \( N_e \). For example, Figure 1b gives the frequency distribution for the same \( 4N_e\mu \) as Figure 1a, but with a larger \( N_e \) and 5000 generations of random mating. Clearly none of the populations are close to MDE. For genomic selection to be commercially applied, the most economical and scaleable technology is currently based on SNP. SNP have a mutation rate on the order of \( u = 2 \times 10^{-9} \) (Kumar & Subramanian 2002). For \( 4N_e\mu = 1 \), as assumed for markers, a population size in the billions would be needed, with corresponding millions of generations to establish MDE. Such large populations and evolutionary time may not be realistic for commercial (or any) populations. Finally commercial populations are derived from wild populations that may have at one time been in MDE, but the domestication process quickly eliminates rare alleles as was observed in a large SNP biodiversity project in layers (Muir & Cheng 2007). Inbreeding as low as 10% during domestication can result in over 50% allele loss (Muir & Cheng 2007). All of these issues bring into question if the MDE should be assumed for SNP.

On the other hand, SNP markers are discovered by sequencing relatively few individuals. One is actually starting out with billions of potential SNP (the genome) and through sequencing finding only those that are polymorphic. The discovery process results in an ascertainment bias (Nielsen et al. 2004; Clark et al. 2005) resulting in common alleles being over-sampled and rare alleles under-sampled. For genotyping commercial animals this is a favourable result as SNP discovered in this way have a higher probability of being informative in all populations. In fact, the SNP frequency distribution will be a uniform distribution if discovery was based on only two individuals (Clark et al. 2005), as is the case for poultry. Thus although \( u = 2 \times 10^{-9} \) for SNP, the frequency distribution of discovered SNP will be the same as that of a population with \( N_e\mu = 1 \) in MDE.

Training the model

Estimation of marker–QTL associations requires phenotypic and genotypic data both within and across generations, which is the purpose of the TG. At each TG, all data collected up to and including that generation were used to estimate breeding values. Thus, with four TG there are four separate rounds of estimation, one for each TG, but all data up to the final TG were used for prediction in the remaining 10-t generations.

Accuracy of GEBV with alternative number of TG and a constant population size of 128 (\( N_e = 16 \)) with GEBV is given in Figures 3 and 4. Results show that accuracy increases as more generations are allowed to generate observations from which effects are estimated. Even more exciting is the observation that GEBV works very well for traits with low heritability. Populations starting in MDE (Figures 3b and 4b) had even higher accuracies at each TG and greater persistency of accuracy. The difference between HWE and MDE conditions was most notable following two TG. The number of generations of high accuracy beyond the TG is critical to the economic efficiency of GEBV. From Figure 3a, a trait of low heritability in HWE can achieve accuracy greater than the heritability of the trait for at least seven generations with at least three TG. A population in

![Figure 2](image-url)  
**Figure 2** Allele frequency distribution for \( 4N_e\mu = 0.01 \) (\( \mu = 2.5 \times 10^{-5} \), \( N_e = 100 \)) with alternative numbers of generations of random mating, 2000 bi-allelic loci and 1000 cM.
MDE (Figure 3b) achieves accuracy greater than the heritability of the trait for at least seven generations with only two TG.

For a highly heritable trait, starting in HWE (Figure 4a), regardless of the number of TG, the accuracy cannot exceed the heritability in the prediction generations. The accuracy is already at such a high level that it is hard to improve on. For a population starting in MDE, at least three generations of training were needed to achieve accuracy greater than the heritability for at least three generations beyond TG. For the same heritability, Meuwissen et al. (2001) found an accuracy of 0.732 for RR with two TG, 2200 phenotypes/genotypes, and starting in MDE. In our simulations we observed accuracies of 0.73 and 0.83, respectively for populations starting in HWE and MDE (Figure 4a,b), with only a total of 512 phenotypes/genotypes following four TG. Thus our results compare even more favourably for GEBV than those of Meuwissen et al. (2001). This result is most likely because of the increased number of TG.

This conclusion was supported by results given in Figure 5. Here the same total numbers of phenotypes/genotypes obtained in TG of both cases was
2048, but in one case (1024/generation) with two TG and in the other (512/generation) with four TG. Clearly, accuracy with four TG and smaller population size is superior to that of two TG with a larger population size in all post-TG generations. Similar results were found for other combinations (where total numbers were held constant but TG differed).

**Marker density**
The effect of marker density on accuracy of GEBV is seen in Figure 6. In these simulations the total number of markers is held constant to 100, but the genome size varied over which the markers were spread, as for example, certain regions of the genome were saturated with markers based on prior knowledge. The main factor being examined here is the recombination rate, holding all other factors constant. As expected, as the recombination rate increases, the accuracy decreases for population starting in either HWE or MDE. These results support the conclusion that higher density is better, but only if the total number of makers is held constant. The following section examines impact of increasing total number of markers with recombination rate held constant.

**Marker number**
Marker density was held constant at 10/cM, but the genome size probed varied, resulting in differing amounts of the genome being queried. These simulations were designed to look at the effect of total number of marker loci, while holding recombination rate and total number of QTL constant. The population was started in either HWE or MDE, \( h^2 = 0.1 \), 128 individuals genotyped/phenotyped each generation, \( N_e = 16 \), the genome increased in size from 1 to 300 cM, corresponding to between 10 and 3000 marker loci; however, in each case the total number of QTL loci was 100. Accuracy was measured in each of the four TG for each case. Results are given in Figure 7. These results show, as expected, that as the TG increase, regardless of the number of makers, the accuracy increases. This result is expected because the total number of observations increases with each TG. However, for populations starting in HWE (Figure 7a), for all TG greater than 1, there is an intermediate optimum at 1000 marker loci. This optimum is most likely because of the relative number of QTL versus marker loci, the optimum occurring at a ratio of 10:1 marker:QTL loci. When the number of QTL loci was greater than the marker loci, the markers cannot capture all the genetic variance. On the other hand, when the numbers of markers far exceed the number of QTL, confounding (co-linearity) of effects in the data cannot be ade-

---

**Figure 6** Effect of marker spacing (recombination rate) on accuracy with four generations of training, constant number of markers (100), and a heritability of 0.1, starting in Hardy-Weinberg or mutation-drift equilibrium, in all cases the total population size was 128 with an effective population size of 16, 100 quantitative trait loci (average over 60 replicates, SEM = 0.02).

**Figure 7** Relationship between number of marker loci and accuracy of selection using a constant recombination rate (marker density of 10/cM), a heritability of 0.1, population size was 128 with an effective population size of 16, accuracy was estimated in each of the fourth training generation, 100 quantitative trait loci (average over 30 replicates, SEM = 0.03). (a) Starting in H-W equilibrium; (b) starting in MDE.
quately separated. This conclusion was supported by results of populations started in MDE (Figure 7b). Here the optimum for all TG was with the minimum 10 marker loci, corresponding to a ratio of 2.5:1 marker:QTL; at 100 markers, the ratio is 25:1 and clearly suboptimal.

**Effective population size**

The effect of $N_e$, and associated generation of new LD through the drift process, is shown in Figure 8 for populations starting in either HWE or MDE. In these simulations a population of size 128 was genotyped for four TG and phenotypes measured for a trait of $h^2 = 0.1$. The numbers of breeders, chosen at random in each generation, were 16, 32, 64, and 128. As expected, because of pre-existing LD, populations starting in MDE have higher accuracy for all $N_e$ than those starting in HWE, the differences between starting conditions becoming greater as $N_e$ increased. These results show that GEBV is effective for populations without pre-existing LD, but only for relatively small $N_e$, not exceeding $N_e = 64$.

Nevertheless, even for populations starting in MDE, generation of new LD was beneficial for improving accuracy. If the rate of decay in accuracy holds constant, for populations starting in MDE, accuracy of GEBV should remain higher than the heritability up to approximately $N_e = 2048$. The populations simulated by Meuwissen et al. (2001) started with a population size of 100, expanded to 200, then to 2000. The effective population size after the initial bottleneck was in fact $N_e = 1/2 \left( \frac{1}{100} + \frac{1}{200} + \frac{1}{2000} \right)$, or $N_e = 194$, which is close to the largest $N_e$ used in these simulations (128).

Comparisons of accuracy to traditional BLUP

In these simulations accuracy of traditional BLUP was compared with GEBV. Starting with a population in HWE, phenotypes for a trait with $h^2 = 0.1$, and genotypes at 100 loci were collected for each of the 10 generations on 256 individuals, of which 32 (16 males and 16 females) were chosen at random as parents of the next generation, i.e. all generations were TG. Accuracy was determined at each generation. The objective of this simulation was to determine if and when the accuracy of GEBV could surpass traditional BLUP. Results are given in Figure 9. The accuracy of BLUP is initially much greater than that of GEBV, but the difference quickly diminishes and GEBV surpasses BLUP by the fourth TG and continues to improve, whereas BLUP quickly reaches a plateau after two generations. These results suggest that with adequate TG, response with GEBV can far exceed traditional BLUP, even when starting in HWE.

**Selection**

The last factor considered here is selection. Measuring accuracy of GEBV in randomly mating populations may not reflect what will occur with selection because selection changes variances, allele frequencies, and generates unfavourable LD (Bulmer 1971) which is not useful to GEBV. All population-wide associations between markers and QTL are because of a difference in the frequencies of coupling versus repulsion phase double heterozygotes. Selection on markers utilizes those associations that are in the
favourable phase, leaving those in the unfavourable phase; much like selection on a trait can change the magnitude and even sign of a genetic correlation between traits (Bohren et al. 1966). For these simulations, populations were started in either HWE or MDE and phenotyped for a trait with $h^2 = 0.1$, and genotyped at 100 loci for each of the four TG on 256 individuals, of which 32 (16 males and 16 females) were chosen at random as parents of the next generation. In the generations following the TG, parents were either chosen randomly or directionally selected based on GEBV. Accuracy was determined at each generation pre- and post-TG.

Results are given in Figure 10. Random selection under these conditions resulted in high accuracy for GEBV at the end of training and continued high accuracy for all generations following training, particularly for populations in MDE (Figure 10b). In contrast, directional selection following TG resulted in a rapid decline in accuracy, the decline was more rapid for populations in MDE. These results show that high accuracies in the TG do not necessarily transfer to populations under selection.

However, the real question is how rapidly GEBV can improve a trait as compared with traditional BLUP. Figure 11 presents changes in the trait phenotype for the same conditions as the previous study except with the addition of BLUP selection. BLUP EBV were based on phenotypes collected in all generations. The objective was to compare how much progress a breeder could make using traditional state-of-the-art methods as opposed to GEBV selection.

For populations starting in either HWE or MDE, response of the first generation of selection based on GEBV was greater than that of BLUP (Figure 11a,b), as would be expected since GEBV had a higher accuracy than BLUP under these conditions. However, in each of the subsequent generations the gains per generation were higher with BLUP than GEBV.
reflecting the decrease in accuracy of GEBV shown in Figure 10. For a population starting in HWE, GEBV selection was only effective for approximately five generations post-TG, while BLUP continued to show steady progress. By the end of 10 generations BLUP had amassed five times the response as GEBV. For a population starting in MDE, response to BLUP selection plateau was after four generations of selection as did GEBV. In both cases genetic variability was exhausted because only four QTL were segregating.

Scale issues

These last set of simulations examines the issue of scale. One set of simulations was completed using a genetic scale one order of magnitude larger than the previous simulations. For these simulations, 1000 bi-allelic marker loci were evenly distributed, and 1000 QTL were randomly distributed, over 1000 cM on 30 chromosomes and closely approximated with the simulations and conditions given by Meuwissen et al. (2001). Assuming the same mutation rates for markers and QTL as in the previous simulations, MDE was established following 1000 generations of random mating with an effective population size of 128. Selection via BLUP and GEBV was accomplished as in the previous simulations. Results are given in Figure 12 for accuracy and selection response. These results differ quantitatively, but not qualitatively from the smaller-scale simulations. The accuracy here is less than before because the total numbers of phenotypes measured were the same as before, but the numbers of genotypes was increased by 10-fold, emphasizing that phenotypes need to increase with a corresponding increase in genotypes. Nevertheless, issues of scale do not substantially change results of these simulations or inference of the simulations using smaller number of loci in non-selected populations.

General discussion

All MAS methods, including GEBV, require LD. LD can either pre-exist because of new mutations in ancestral populations, or as a result of new LD generated by the drift process. In the case of pre-existing LD, those will be primarily closely linked because of the large number of generations between occurrence of the mutation and detection in commercial populations. For those marker–QTL closely linked, but in LE, genetic drift as a result of reduction in population size can return those loci to LD. Population-wide LE and LD is simply the relative frequency of coupling versus repulsion heterozygotes. Any genetic force, including migration, selection, and drift, can change these frequencies.

Results presented here are conservative for what is possible with GEBV because we used RR which required an unrealistic assumption that all markers account for an equal proportion of the genetic variance. More advanced methods, such as those suggested by Meuwissen et al. (2001) or Xu (2003), or Markov chain Monte Carlo (MCMC) will perform somewhat better as already seen by Meuwissen et al. (2001).

Nevertheless, a critical issue for even those methods is the approximate bounds on the relative amount of phenotypic versus genotypic information, across generations, such that true marker effects in LD with QTL can be separated from other confounding effects. Adding makers without also increasing the total number of phenotypes may not increase, and can decrease accuracy. This result is expected to

![Figure 12](image-url)
be true for all methods. In the end, no amount of statistical sophistication can make up for lack of phenotypic data. Confounding due to co-linearity can only be removed by design.

However, others have argued that these conclusions may be method-dependent, true for RR, but perhaps may not be true for other methods, such as BayesB. The latter may be true if, e.g. there are multiple markers accounting for the same haplotype block, and then all markers in the same block are equally predictive regardless if they are confounded. At the heart of this debate is whether a filtering method is needed to reduce the number of independent variables. With classical prediction methods, adding additional independent variables could result in fitting noise and poorer prediction. This will be a topic for further research and discovery.

As an idea of the number of phenotypes needed for independent estimation of genotypic effects, in designed studies where all possible interactions are of interest, the number of observations needed \((B)\) (without replication) is \(B = M^N\) where \(M\) is the number of levels of a factor (alleles per markers) and \(N\) is the number of factors (loci). This number becomes impractical even for small \(N\). If one is only interested in main effects, assuming no interactions, there needs to be at least \(M\) observations for a given factor, holding all others constant, i.e. \(B = MN\) observations. For a genome size of 3000 cM and 100 SNP/cM, this would require 300 000 loci \(\times 2\) alleles per locus = 600 000 phenotypes, and these would be best measured over multiple generations to separate loose from tightly linked QTL.

Alternatively, and as a compromise, the best way to utilize GEBV is use combined selection methods. As seen from Figures 11 and 12, GEBV can only improve a trait to the extent the markers and QTL are in the correct phase, even if adequately estimated, and not all QTL will be in LD with surrounding markers, even if closely linked, those will be in linkage equilibrium (LE). Classic phenotypic selection (BLUP) can capture QTL regardless of LD and LE, i.e. does not need or use markers. One method to combine GEBV with classic BLUP is to add a second random effect to the mixed model for the polygenic part. The first random effect is associated with the markers while the second effect is the usual polygenic effect accounted for by the relationship matrix.

\[
Y = XB + ZG + Wu + \varepsilon
\]

The first three terms are as defined in Equation 1), \(W\) is the animal effect and \(u\) is a vector of polygenic effects. This model will capture loci in either population-wide LE or in LD. The downside of this model is that continued phenotypes are needed in all generations, but as seen from Figure 9, accuracy of GEBV can exceed that of BLUP, thus by combining the two methods accuracy can be maximized.

**Conclusions**

Assumption of a MDE is problematic for SNP marker loci because of the population sizes needed and/or generations required to establish, and or utility of the predominant very rare SNP in such a MDE (Figure 2). Nevertheless, due to the way SNP are discovered, a near uniform distribution of informative SNP results, the distribution of which resembles an MDE for \(N_{e}u = 1\), as assumed in these simulations and those of Meuwissen et al. (2001).

GEBV excels for traits of low heritability regardless of initial equilibrium conditions (Figures 3 and 4), as opposed to traditional MAS, which is not useful for traits of low heritability. Accuracies estimated from populations starting in HWE resemble those in MDE in shape, but the intercepts are lower and the persistency of accuracy is less. TG are very beneficial for increasing accuracy of selection even for the same fixed total number of phenotypes/genotypes measured (Figure 5). TG separates those associations because of loose versus tight linkage. In addition, TG allow for random genetic drift to generate new LD which is also beneficial to GEBV. However, only two TG are necessary to estimate loose versus tight linkage. Thus, the reasons for the importance of TG beyond two is perhaps purely for generation of LD. The debate of importance of historical LD versus LD generated by drift is a matter for debate and continued research.

Increasing number of markers has conflicting effects on accuracy. As the number of markers increase, both the density and the number of independent variables increase. Increased density, with associated reduced recombination rate, increases accuracy (Figure 6) but increasing the number of markers without increasing the number of phenotypes is counter-productive (Figure 7) as co-linearity issues confound effects and can reduce accuracy. This effect could be particular to RR and might be overcome by BayesB or other such estimation methods.

Effective population size is critical for population HWE, in which case GEBV is only effective for very small \(N_{e}\) and accuracy is not very persistent. On the other hand, populations that start from MDE still benefit from smaller \(N_{e}\) but the importance is much reduced, and persistency of accuracy extends for
more generations (Figure 8). Thus GEBV is expected to achieve high accuracy in most commercial species, but perhaps more so for smaller than large $N_e$.

In comparison with traditional BLUP, GEBV can exceed the accuracy of BLUP (Figure 9). The accuracy of BLUP is theoretically limited when based on individual and ancestor’s information. The accuracy of BLUP can approach one only with progeny testing or inbreeding. On the other hand, with adequate TG, the accuracy of GEBV can approach one without progeny testing.

Unfortunately, selection rapidly reduces the accuracy of GEBV as the favourable LD is utilized. Regardless of starting conditions, HWE, or MDE, accuracy rapidly declines after selection is initiated. In comparison with BLUP, early selection based on GEBV is greater than that of BLUP, but within as few generations BLUP gains exceed GEBV. Even still, GEBV could have an advantage over traditional BLUP in cases such as sex-limited traits, traits that are expensive to measure, or can only be measured on relatives.

Acknowledgements

This research was partially funded by a USDA/NRI research grant 2004-05434. The manuscript was greatly improved and discussion added as a result of two wonderful insightful anonymous reviewers.

References


