View to the Future:
Could Genomic Evaluation Become the Standard?

M. E. Goddard
University of Melbourne and
Department of Primary Industries, Victoria
Australia

Genomic selection

Genomic selection means marker assisted selection based on a panel of markers that cover the genome so densely that every quantitative trait locus (QTL) is in linkage disequilibrium with at least one marker (Meuwissen *et al.* 2001). Genomic evaluation is the calculation of estimated breeding values (EBVs) based on this marker data. The main requirement for genomic selection is a 'reference population' consisting of animals that have been evaluated for the trait, genotyped for the markers and from which an equation to predict breeding value from genotypes can be derived. This prediction equation can then be applied to animals that have only DNA marker genotypes but no phenotype or offspring. This allows selection to be combined with short generation intervals and hence achieve faster genetic improvement.

Genomic selection will become the standard

Selection of dairy cattle based on DNA markers will become the standard because it will lead to faster genetic gain than other methods and the bulls with the highest EBVs will be young bulls without milking daughters. Meuwissen *et al.* (2001) showed in simulation how breeding value could be predicted with accuracies as high as 0.8 and Schaeffer (2006) showed how this could double the rate of genetic gain. The promise displayed by these simulation studies is starting to be born out in real data. USDA now claim the accuracy of evaluating young Holstein bulls is approaching 0.8 for milk production traits.

Detailed studies could be done to show how best to design a breeding program taking advantage of genomic selection but a very simplistic approach is enough to appreciate that it will increase the rate of genetic gain. Consider a nucleus selecting bulls for progeny testing and cows as bull dams and as dams of replacements within the nucleus. Assume that heifers are used at 12 months of age as embryo donors without any selection except that they were born in the nucleus. Table 1 lists the selection intensities and generation intervals. The rate of gain $0.2\sigma_g$ per year is similar to that seen in well run traditional breeding programs. A breeding program using genomic selection could select both males and females at 12 months based on DNA markers and achieve a rate of genetic gain of $0.42\sigma_g$. In practice the gains may be more or less than this, but the potential to increase rate of gain by using short generation intervals and selecting in both sexes is obvious.

Table 1. Rate of genetic gain (ΔG) in a traditional breeding program and in a breeding program using genomic selection. r=accuracy of selection, i=standardised selection intensity, L=generation length, $\Delta G = \Sigma(ir)/\Sigma L$.

	Traditional			Genomic selection		
	r	i	L	r	i	L
males	0.8	2	6	0.6	2	2
females 0	0	2		0.6	0.8	2
ΔG	0.2			0.42		

Cattle breeders do not make decisions about semen purchase on the basis of long term rates of genetic gain but on comparisons between bulls available to them. Based on the assumptions used for table 1, bulls of one year of age, selected on DNA markers will be as good as proven bulls initially. However, as the rate of genetic improvement increases, the young bulls will become clearly superior to the proven bulls (Table 2).

Table 2. Mean breeding value of bulls of 1 year of age selected on DNA markers and proven bulls 5 years of age. The genetic merit is expressed in units of genetic standard deviation relative to the mean of the progeny test team from which the proven bulls where selected.

		Traditional	Genomic selection
ΔG		0.2	0.42
Bull age	selected on		
1	DNA	1.68	2.88
5	progeny test	1.6	1.6
5	both	2.8	2.8

The transition from the old rate of gain to the new, higher rate of genetic gain will happen in stages. Initially it will come about because the dams of young bulls have been selected on the basis of DNA markers whereas the dams of proven bulls, being older, have not been selected on DNA. Then the sires of the young bulls will be selected on DNA markers. That is, during a transition phase, the young bulls will have superior sires and dams to the proven bulls and this will more than compensate for their less accurate genetic evaluations giving them a greater advantage than shown in the 'traditional' column of table 2. Eventually the proven bulls would also have parents selected on the basis of DNA markers but by then the rate of gain will have increased and the young bulls will still be superior to proven bulls.

Table 2 includes the scenario where bull calves are selected for progeny testing based on DNA genotypes but when 5 years of age are available with progeny test results. In calculating the merit of these bulls I have simply added the gains from genomic selection and from traditional progeny test. This results in an overestimation of the merit of these bulls because the variance available at the second selection stage (after progeny testing) is reduced and because the proportion selected at the second stage must be greater than is possible in a single stage based on DNA genotypes. However, the table suggests that there will be some proven bulls that are only slightly below the merit of the young bulls available at the same time. Thus, if the number of bulls progeny tested did not decline, there would continue to be a pool of competitive proven bulls for farmers who wished to buy semen from them. However, AI companies are likely to reduce the number of bulls they progeny test as they utilise genomic selection to select and market a team of young bulls. The optimum design to maximise genetic gain is likely to involve very extensive DNA testing of bull calves and a drastic reduction in the number of bulls progeny tested. This will have a positive side effect in countries where the number of daughters per bull is low in that progeny tests for lowly heritable traits such as fertility will become more accurate.

Accuracy of genomic evaluations

The theoretical accuracy that can be achieved from genomic selection has been investigated by Goddard (2009), Hayes *et al.* (2009a and b). The least favourable scenario is where there are a very large number of genes affecting the traits of interest, all with small effects. In this case, the best analysis is the so-called 'BLUP' model of Meuwissen *et al.* (2001) which assumes all SNP effects are drawn from the same normal distribution. The theoretical

accuracy depends on $\lambda = TR/(N_eS)$ where T=the number of animals in the reference population, R = the reliability of the EBV available on animals in the reference population, N_e = effective population size and S= the length of the genome in Morgans. From this parameter and the theory developed in the papers quoted above, we can predict the accuracy of EBVs calculated from many SNPs covering the whole genome. This theory assumes constant N_e which is not what has occurred in the history of modern cattle, but it is probably the N_e in the last few generations that is relevant, so I will use N_e =100. Then with T=7000 animals, R=0.8, S=30, we expect an accuracy of 0.8 and that is approximately as observed.

Therefore countries other than USA and breeds of cattle other than Holstein should be able to utilise genomic selection provided they can assemble a large reference population. In fact, there are benefits of increasing the reference population well beyond 7000 animals and of updating it with new information. This is especially true for traits where existing EBVs are less reliable than is typically the case for milk production traits. For instance, if longevity EBVs have a reliability of only 0.4 then we would need to double the size of the reference population to maintain the same accuracy of genomic EBVs. To keep expanding and updating the reference population we may need to make use of cows as well as progeny tested bulls. Cows have EBVs with lower reliability than bulls, so we will need many more of them than bulls. Taken together, these considerations convince me that the dairy industry would benefit from using very large and growing reference populations.

Uses of genomic evaluations

Although the first use of genomic selection has been to select among bull calves the best ones to progeny test, there are many related uses. They include selection of bulls as sires of cows, selection of bulls as sires of sons, selection of cows as dams of sons, selection of cows as dams of cows, selection among heifer calves of those to retain as replacements, selection of mates to minimise unfavourable non-additive gene effects such as inbreeding, allocation of heifers to the management system and dairy product stream that maximises their profitability, warning of cows prone to a particular disease, determining pedigree.

The breeding value of cows can be evaluated as accurately as bulls using genomic evaluation. Given the multiple uses of DNA marker genotypes, as the cost of genotyping decreases, I predict that most cows will have a DNA profile and this will be the most important element in calculation of their EBV. However, phenotypic information will continue to be collected as part of herd recording for management purposes and this, together with the genotypes of the cows, will allow for continued updating of the reference population and therefore the prediction equations.

Future of national genetic evaluations

If both bulls and cows will be selected on DNA markers in the future, is there a need for national genetic evaluations? An alternative would be that either companies selling bulls or companies selling DNA testing carry out genetic evaluations. I believe that a national evaluation is better because it uses all national data to calculate the most accurate EBVs and they are all expressed on the same scale which makes their use by dairy farmers much easier. In addition, the herd recording system in many countries is the natural channel through which data (genotypes and phenotypes) and reports (eg EBVs) flow from and to dairy farmers. There is also a need for updating the prediction equations and this requires performance data and EBVs calculated through herd recording and national evaluation systems.

For the foreseeable future, EBVs will be calculated using traditional performance data as well as any DNA genotypes that might be included. There are several ways that this can be done. For instance, the traditional EBVs can be calculated as at present and, separately from them, breeding values predicted solely from the genotypes and the prediction equation. Then a selection index method can be used to combine them. The predictions based solely on DNA genotypes have been called direct genomic values or molecular breeding values

and the EBV resulting from combining the two sources of information has been called a genomic EBV, but this terminology is not universally agreed upon and may change. This method of combining EBVs from traditional and marker data is not ideal but is practical when only a small proportion of animals have DNA genotype data. Another approach would be to calculate the relationship between animals based on marker genotypes, if they exist, and on the pedigree if they do not exist. This is equivalent to genomic selection because using relationships calculated from markers is equivalent to using the markers in the BLUP model of genomic selection (Goddard 2009). However, this approach is not practical if there are many genotyped animals because the relationship matrix must be inverted and there is no quick way to do this comparable with forming the inverse of the numerator relationship matrix based on pedigree.

In the future we need a method that can utilise large number of genotypes (millions of markers) on cows as well as bulls. I suggest that as the number of animals with genotypes increases, national genetic evaluations will abandon traditional methods and use a statistical model for phenotype that includes the effect of a large numbers of markers. This will remove the need to fit relationships altogether. The marker genotypes will either be from laboratory genotyping or will be imputed from genotyping on the animal itself and/or on its relatives. I will call this a gene based model (Goddard 1998) although, initially at least, it is a marker based model. However, with time, more and more causal mutations will be discovered and included in the genotype data. The number of markers will increase rapidly and this will lead to models that assume that only some markers have an effect on the trait ie the model called Bayes B by Meuwissen *et al.* (2001).

Future of International Genetic Evaluations

How should the role of Interbull change in the future? Interbull could continue to offer their existing service which uses phenotypes but no genotypes to calculate EBVs. I fear this will be a diminishing role as selection based on DNA markers becomes the norm. If Interbull is to utilise DNA genotype data how should it do so? One criterion to use in answering this question is "what method would yield the most accurate EBV"? I suggest that the best method by this criterion would use the genotype data directly. Interbull would combine data from around the world to construct a large reference population and use this to estimate prediction equations for all traits in all countries. This could be considered as a continuation of their current role but merely incorporating a new source of data (ie genotypes).

This approach will be prevented in the short term by countries or companies being unwilling to share the genotypes of animals in their reference population because they believe this knowledge gives them a competitive advantage over other countries or companies. However, I believe this competitive advantage will be short lived as others use the continually improving technology to catch up. Therefore, I suggest we keep this long term objective in mind as we formulate shorter term solutions. A series of options for the short term have recently been considered by an Interbull taskforce (Banos *et al.*, 2009).

If national evaluations move to 'gene based models' Interbull may also have to change its approach or the national and Interbull evaluations will not mesh together as they do at present. Perhaps we could use a gene based model at Interbull but with a dispersed computing strategy so that each country's data were largely processed nationally with only a minimum exchange of intermediary calculations between the country and Interbull. Instead of these intermediary calculations being DYDs as at present, they might be national solutions for all marker effects.

Conclusions

In the near future, EBVs based on DNA marker data will become the main tool for selection of bulls and cows. This will increase the rate of genetic gain and lead to a variety of other uses for the DNA genotypes including management of the herd. However, to obtain these benefits we will need large reference

populations that are continually updated. Commercial use of markers on cows can provide this as long as the DNA data and marker data are captured and brought together, for instance, at a national genetic evaluation centre. We are currently in a transitional phase during which practical but sub-optimal solutions will be found and used. It is difficult to predict the solutions that may emerge in the longer term. I have suggested that we will use data sets with millions of cows each with millions of genetic markers. This data could be analysed using models that directly include the effect of the markers and ignore pedigree relationships. Interbull could perhaps use a similar model with completely different summary data being transmitted between national evaluations and Interbull from the DYDs that we currently transmit. This new integration of national and international evaluations might resemble a distributed computing strategy for a world-wide analysis based on raw genotype data of bulls and cows.

References

Banos, G., Calus, M., Ducrocq, V., Durr, J., Jorjani, H., Liu, Z., Mantsaari, E., Sullivan, P. & VanRaden, P. 2009. Preliminary report from the Interbull Task Force on the role of genomic information in genetic evaluation.

Goddard, M.E. 1998. Gene based models for genetic evaluation – an alternative to BLUP? *Proc 6th Wrld Congr. Genet. Appl. Livest. Prod. 26*, 33-36.

Goddard, M.E. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica. 136,* 245-257.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci. 92,* 433-443.

Hayes, B.J., Visscher, P.M. & Goddard, M.E. 2009. Increased accuracy of artificial selection by using eh realised relationship matrix. *Genetics Research 91,* 47-60.

Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Prediction of total genetic value using genome wide dense marker maps. *Genetics 157,* 1819-1829.

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim Breed Genet. 123,* 218-23.