Effect of Genomic Selection on National and International Genetic Evaluations

Sijne van der Beek
CRV, P.O. Box 454, 6800 AL Arnhem, The Netherlands
E-mail: Beek.S@hg.nl

Introduction

Sax (1923) was the first to show how genetic factors influencing quantitative traits can be identified using markers. It was only in the 1980's that the construction of linkage maps with many genetic markers allowed for the systematic screening for chromosomal regions influencing important traits (Botstein *et al.*, 1980). The 1980's en 1990's were the decades of linkage maps and QTL hunting. The general concept was that the most important chromosomal regions underlying quantitative traits would be identified and then used in selection. However, only few important genes (e.g. Grisart *et al.*, 2002) have been identified in commercial dairy cattle. As summarised by Dekkers (2004) commercial application of marker assisted selection was less successful than expected, mainly due to the lack of identified markers in linkage disequilibrium with important traits.

Haley and Visscher (1998) predicted that the development of cheap and high-density marker maps would move the selection based on polygenes plus individual loci to effective total genomic selection.

Meuwissen *et al.* (2001) developed the analytical framework to compute total genomic values given high-density marker maps and showed that with genomic selection the long awaited molecular revolution of animal breeding is within reach.

Genomic selection

Bovine high density genetic maps (http://www.ensembl.org/Bos_taurus/index.html), cheap typing technology (e.g. www.illumina.com) and the framework to compute a total genomic value (Meuwissen *et al.*, 2001; Meuwissen and Goddard, 2004) are now available. The computation and use of such total genomic values is commonly referred to as Genomic Selection.

Genomic Selection provides a unified concept. Because the whole genome is analysed simultaneously, there is no need for QTL or gene identification. The method just assumes that the whole genome explains all genetic variation. De Roos *et al.* (in press) showed that for high marker densities genomic selection without prior knowledge on the location of a large gene (in their case the DGAT gene) is as accurate as gene assisted selection in which explicit knowledge on a major gene is utilised. So given dense marker data, phenotypes, and a proper analytical tool, one can directly start to estimate breeding values without bothering about the identification of QTLs or genes.

Meuwissen *et al*. (2001) showed that with a dense marker map with one polymorphic multi allelic marker per centimorgan and a half sib structure with 100 offspring per sire, genomic selection yields an accuracy of selection of 0.73 when BLUP was used that assumed equal variance associated with each chromosomal segment. An accuracy of 0.85 was reached when a Bayesian method was used that assumed a prior distribution of the variance associated with each chromosome segment. So the use of a method that simultaneously estimates the variance associated with all chromosomal regions and all allelic or haplotype effects is required for efficient genomic selection.

Genomic selection of young animals with an accuracy of 0.85 is equivalent to a situation in which markers explain 50% of the variance with 100% accuracy and the other 50% of the variance is due to polygenes. Assuming a reliability of 40% for young animals for the polygenic part, the reliability of markers plus

polygenes will be 0.5×(1+0.4) = 0.7 which leads to an accuracy of 0.84. Schrooten *et al.* (2005) analysed the genetic progress for this situation. Genetic progress was 19 – 31% higher when markers explained 50% of the variance and number of progeny tested bulls was constant. When the number of bulls to be progeny tested was halved by preselection based on the markers, the genetic progress was hardly affected at all. When all selection was based on markers and progeny testing was abandoned, genetic progress increased by 70% due to the sharp reduction of the generation interval (Schrooten *et al.*, 2005).

To implement genomic selection a reference population is needed. The granddaughter design as it commonly is used in dairy cattle QTL mapping studies fits the purpose. Once the animals in the reference population are typed for the markers, genomic selection can immediately be applied to all traits for which estimated breeding values are available. This is a huge breakthrough compared to traditional marker assisted selection. In traditional marker assisted selection we relied on individual loci that were identified significantly using a genome wide significance test. The total amount of variance explained by the QTL identified was limited. In genomic selection we immediately explain most of the variance for all of the traits, without intermediate time consuming QTL hunting.

In October 2006 CRV (Holland Genetics) started to use genomic selection in her breeding programme. This application of genomic selection is based on 3000 SNPs using the methods of Meuwissen and Goddard (2004) and Windig and Meuwissen (2004). All current Holland Genetics test bulls are preselected based on genomic selection. The set of 3000 SNP markers will soon be replaced by a much larger set. This means that genomic selection with 0.85 or higher accuracy is within reach.

National genetic evaluation

Already for a number of years several breeding companies in various countries preselect young bulls based on genetic markers. The genetic marker information is not included in the national genetic evaluation system. Thus the assumption that the genetic evaluation model includes all data on which selection is based is violated. Marker assisted selection might therefore bias national genetic evaluation. However, the effect will have been limited since until recently the genetic markers explained only a fraction of the genetic variance.

Genomic selection will change this. Let's assume genomic selection with 65% reliability (81% accuracy) for all traits. Currently, a young bull has a parent average with about 40% reliability for production and about 30% reliability for durability and health. Genomic selection can be modelled as an additional source of information that explains 25% of the variance (65-40) for production and 35% of the variance (65-30) for durability. Let's further assume that based on genomic selection we select 1 out of 4 available young bulls that are already selected based on parent average. The young bulls are then tested using a standard progeny test and evaluated using a standard model that does not include genetic marker data. In the model, the expected value of the young bulls will be the parent average and the expected variance of the term young bull minus parent average [YB-PA] will be equal to the mendelian sampling variance. In reality, however, 1 in 4 genomic truncation selection with 65% reliability will change the mean and variance. Using simple computations in Microsoft Excel, it was computed that the mean of the young bulls shifts from parent average (PA) to PA + 0.5 σ_a and that the variance among sibs reduces from σ^2_{ms} to 0.25 $\times\sigma^2_{ms}$. Variance estimates in national and international genetic evaluation are largely based on offspring minus parent average terms. If genomic selection does not affect the parent average estimate then the average $[YB\text{-}PA]^2$ is 1.6 σ^2_{ms} where the model assumes this to be just σ^2_{ms}.

So the average true value breeding of genomically selected young bulls will be higher than the model assumes. Young bull breeding values will be biased downwards. For production, this effect will be limited because quickly the EBV of a bull will be based largely on daughter performance and not on pedigree information. For a trait like durability this is different. Currently, the EBV of a first crop bull will only get a reliability higher than 65%

if the daughters are in the third lactation. With genomic selection, you can reach 65% reliability right after birth. Pedigree information will for a long time have a large effect on the durability EBV of a bull, and thus for a long time genomically selected bulls will be underestimated for a trait like durability.

International genetic evaluation

Basically, if a bull is tested in country A than its breeding value in country B depends on the parent average in country B and on the mendelian sampling term in country A. Since national genetic evaluation underestimates the mendelian sampling term, this bias will be converted to country B.

To derive interbull EBVs, sire variances and genetic correlations are required. Genomic selection affects the quadratic terms that are used to compute the genetic parameters. Therefore, interbull genetic parameter estimates can be biased.

Interbull (trend) validation tests can also be affected by genomic selection. Given the nature of the information utilised, Interbull tests 1 and 2 are not likely to be affected. Interbull test 3, however compares evaluation runs over time. As shown above, genomic selection will cause a downward bias in early genetic evaluations especially for traits like durability. The new interbull test that monitors the mendelian sampling variance over time will probably sensitive to genomic selection. This sensitivity will however depend on the intensity of genomic selection.

The future

Several breeding companies will introduce genomic selection the next few years. Accuracy of genomic selection will be above 0.8. In some countries breeding companies will not submit marker data to a national database and the national genetic evaluation will not contain any marker information. In other countries breeding companies will work together to build up a reference population and an infrastructure to compute genomic selection breeding values. But still, in those countries the national genetic evaluation most likely will not include the marker information. In a third group of countries, the government or other bodies will be involved in analysing the reference population and building up the analytical framework for genomic selection. The latter group of countries might integrate marker information in the national genetic evaluation and submit national genetic evaluations that include marker information to the international genetic evaluation.

In any scenario, breeding companies will select and probably also market bulls based on genomic selection breeding values. Breeding companies would also like to convert those values to the base and scale in other countries.

Also in any scenario, international genetic evaluation will be affected. Some countries might submit national genetic evaluations that include marker information and Interbull needs to determine how to deal with that. And certainly, some countries will submit national genetic evaluations that do NOT include the marker information based on which bulls are selected. In the latter case Interbull needs to evaluate the robustness of the genetic parameter estimation procedures and (trend) validation tests. Also methods that deal with bias (Sullivan, 2002) will be required.

Conclusion

Genomic selection will be introduced in the near future and have a major impact on dairy cattle breeding. National and international genetic evaluation will be affected. Methods and procedures need to be derived both for the situation that the marker data will (partly) be incorporated in national genetic evaluation and for the situation that the marker data will not be incorporated in national genetic evaluation.

References

Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet. 32*, 314-331.

Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in

livestock: Strategies and lessons. *J. Anim. Sci. 82 (E. Suppl.)*,E313-E328.

De Roos, A.P.W., Schrooten, C., Mullaart, E., Calus, M.P.L. & Veerkamp, R.F. 2007. Breeding value estimation for fat percentage using dense markers on BTA14. *J. Dairy Sci. 90 (in press).*

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Nadine Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. & Snell, R. 2002. Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. *Genome Research 12*, 222-231.

Haley, C.S. & Visscher, P.M. 1998. Strategies to utilize marker-quantitative trait loci associations. *J. Dairy Sci. 81*, 85-97.

Meuwissen, T.H.E. & Goddard, M.E. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol. 36*, 261-279.

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.

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in Phaseolus Vulgaris. *Genetics 8, 552-560*.

Schrooten, C., Bovenhuis, H., Van Arendonk, J.A.M. & Bijma, P. 2005. Genetic progress in multistage dairy cattle breeding schemes using genetic markers. *J. Dairy Sci. 88*, 1569-1581.

Sullivan, P.G. 2002. Genetic evaluation strategies for multiple traits and countries. *PhD Thesis*, University of Guelph, Canada.

Windig, J.J. & Meuwissen, T.H.E. 2004. Rapid haplotype reconstruction in pedigrees with dense marker maps. *J. Anim. Breed. Genet. 121*, 26-39.