GENOMIC SELECTION: THE FUTURE OF ANIMAL BREEDING

T.H.E. Meuwissen
Norwegian University of Life Sciences, Box 5003, 1432 Ås, Norway

INTRODUCTION

Genomic information may be defined as 'all information contained in the genome', but with current and foreseeable future state of technologies this information is largely restricted to 1) known genes, which may become increasingly available through QTL mapping, fine-mapping, and causative mutation detection; 2) genetic markers, which may be used for marker assisted selection (MAS); 3) use of high-density marker maps in genomic selection (GS). The latter two points may sound very similar, i.e. they are both using markers for selection, but there is a fundamental difference: MAS uses only a fraction of total genetic variation, and thus must always be assisted by traditional animal breeding methodology, whereas GS tries to utilise all genetic variation and therefore makes selecting solely for genomic information possible. Hence, it achieves one of the holy grails of molecular genetics, namely to select directly for genomic information avoiding the traditional indirect selection for phenotypes of the animals and their relatives.

It is expected that in the near future animal breeding will increasingly use genomic information, and that this technology has come to a stage where it is ready for implementation in 1) breeding value estimation; and 2) novel designs of breeding schemes. The aim of this paper is to review the possibilities for MAS, GS and the novel designs breeding schemes.

MAS

The model: The main step in MAS is the development of marker assisted breeding value estimation (MA-BLUP), which was already developed by Fernando and Grossman (1989):

$$\mathbf{y} = \text{fixed effects} + \mathbf{a} + \mathbf{v}_p + \mathbf{v}_m + \mathbf{e}$$

where $\mathbf{y}$ = the phenotypes, $\mathbf{a}$ = the usual animal model effect, which represents the background genes here (ie. the non-marked genes); $\mathbf{v}_p$ ($\mathbf{v}_m$) = the effect of the paternally (maternally) inherited QTL allele. $\mathbf{a}$ is a random effect with Var($\mathbf{a}$) = $\mathbf{A}\sigma_a^2$, i.e. the relationship matrix; $(\mathbf{v}_p\ \mathbf{v}_m)'$ is a random effect with Var($\mathbf{v}_p\ \mathbf{v}_m$) $=\mathbf{G}\ \sigma_q^2$, with the (i,j)-th element of $\mathbf{G}$ being equal to the IBD probability of alleles i and j given the information on nearby markers; and σ_q^2 is variance explained by the QTL. Note that this approach does not require that the markers are very close

to the QTL, but as the distance between the markers and the QTL increases, the G matrix approaches the usual relationship matrix, A, and the MA-BLUP EBV approach the usual non-MAS EBV. This is how MA-BLUP is implemented in the MAS programmes of Germany (Bennewitz et al., 2004) and France (Boichard et al., 2002).

GENOMIC SELECTION

In genomic selection we will assume the availability of high-density marker information, implying a marker every cM or even more dense. Especially in the case of SNP markers a higher density will be required. In genomic selection we estimate the EBV as:

$$EBV_i = \Sigma\, CSE_{ij} \qquad [1]$$

Where CSE_{ij} is the estimate of the j-th chromosome segment effect of animal i, and summation is over all chromosome segments j. The chromosome segments are identified by the dense marker haplotypes, e.g. Meuwissen et al. used 2-marker-haplotypes to identify the segments. The estimates of the chromosome segments are obtained from a relatively small data set, e.g. an experiment containing 1,000 or more genotypes and phenotypes animals. **Chromosome segment effects are not tested for their statistical significance.** Such significance testing would imply that most of the segment effects would be found non-significant and are set to zero. It is expected that all the segments with no effect will on average have an estimated effect that is close to zero, and it is the average (or summed) effect that is of importance for equation [1]. Also, non-zero chromosome segments will often be over-estimated or under-estimated, but on average also their effect will be accurately estimated. Note that the estimation of the chromosome segment effects relies on the presence of LD between the marker haplotypes and the QTL that lie in-between the markers.

There will be a huge number of chromosome segments, typically 50,000 or more, and these have to be estimated in a data set of limited size, say 1000 or 2000 phenotyped and genotyped animals. Traditional statistics can not be applied to estimate all these effects, because there are not enough degrees of freedom to estimate so many CSEs. Meuwissen et al. (2001) compared 4 alternative methods 1) Least Squares: this is the traditional statistics approach but with some model selection step to reduce the number of effects to estimate. The chromosome segments were fitted one by one and those increasing the log-likelihood by more than 14 units, were assumed significant and were fitted simultaneously in the model. This resembles the QTL mapping approach: first detect the most significant QTL, and then fit only the most significant QTL in the MA-BLUP.

2)BLUP: In the BLUP approach, all CSEs are fitted simultaneously, but the degrees of freedom shortage is avoided by assuming that the CSEs all come from a distribution of effects with mean 0 and equal variance.
3) GIBBS: In the Gibbs approach, the CSEs are also assumed to have a prior distribution, but now the variance differs per position of the segment, i.e. in some positions the effects have more variance and thus can be bigger than in others. This approach was implemented by Gibbs sampling.
4) METROP: This approach is very similar to the Gibbs approach except that here with some probability P there was no variance at all at a particular position of the chromosome segments. This resembles the situation where there are no QTL at this segment.

The approaches were tested in a computer simulation of 10 chromosomes of 100 cM each. Each centimorgan had a micro-satellite marker at its beginning and end, and a potential QTL position in the middle. Whether the QTL position was explaining variation or not depended on whether there had been a mutation at the QTL position or not. 1000 generations were sampled to create a mutation drift balance and LD between the markers and the QTL. Next, in generations 1001 and 1002, a typical QTL mapping experiment was set up with 100 sires and 100 dams mated to produce 2000 offspring. The parents and their offspring were all genotyped and phenotyped and their data were used to estimate the CSEs. The 2000 animals in generation 1002 also obtained 2000 offspring in generation 1003, which were only genotyped, and their EBV were predicted using equation [1]. Since in the simulation study the true breeding values were known, the accuracies of these estimates were calculated and were: 0.36, 0.74, 0.80 and 0.84, for the methods Least Squares, BLUP, Gibbs and Metrop, respectively.

Thus, the least squares approach, which resembles the QTLmapping/MAS approach, yielded quite low accuracy probably because many of the QTL were not significantly detected, and thus only a fraction of the total genetic variance was used. The results for Gibbs, and Metrop show that quite high accuracies (similar to that of a progeny test) are possible by genomic selection. These high-accuracy EBV make the design of complete new breeding schemes possible, where phenotypic recording does not need to precede the selection steps.

BREEDING SCHEMES

Effect of genomic information: Firstly, since it provides extra information about the genetic value of animals, genomic information increases the accuracy of the EBVs. This implies that it is mainly useful for traits where the accuracy of EBVs is low in traditional breeding schemes: 1) traits with low heritability; 2) traits with recording difficulties (expensive recordings, recording after slaughter, or disease challenge testing, late in life recordings, and sex limited recordings). Secondly, genomic information may be used to reduce the generation interval, since the

genomic information can be recorded at very young age (even on embryos). This effect of genomic information may be bigger than that on the accuracy of selection, because accuracies are often already high in breeding schemes, whereas a factor 2 reduction in generation interval implies a factor 2 increase in rate of genetic gain. This reasoning resulted in the velogenetics schemes of Georges and Massey (1991), where oocytes were harvested from in-uturo calves, invitro matured and fertilised, selected based on markers, and implanted in recipient cows. Velogenetics schemes reduced the female generation interval from the usual minimum of 2 years to 6 months. Haley and Visscher (1998) took this idea to the extreme in their Whizzo genetics schemes. In these schemes, in-vitro meiosis was induced in embryos, followed by invitro maturation, and fertilisation, and selection based on markers, after which the selection cycle was repeated. The Whizzo genetics schemes could be run entirely in the lab, and the generation interval depends on the lab techniques.

In conclusion, genomic information seems most useful for difficult or expensive to record traits. Such traits could be recorded in an experiment together with dense marker information, and CSEs could be estimated. Next, the commercial/elite animals could be selected based on $EBV_i = \Sigma CSE_{ij}$., ie. which does not require any phenotypic recording. Use of genomic information will also result in reduction of generation intervals, which may well be its biggest impact on current breeding schemes. Substantial reductions are possible even with natural/conventional reproduction, but the most dramatic increases in genetic gain will be obtained when genomic selection is coupled to reproductive technologies that further reduce the generation interval. In order to make genomic selection possible, molecular genetics has made important advances recently in that genomes of important livestock species are being sequenced or have been sequenced, which makes large number of SNP markers available, and microarray technology has made genotyping for >10,000 SNPs possible at affordable costs of about 400 US$ per animal, which is cheaper than most current testing procedures.

REFERENCES:

Bennewitz, et al.(2004) *J. Anim. Breed. Genet.* **121**: 307-318.
Boichard D, et al. (2002) 8th WCGALP **22**: 03.
Fernando, R.L. and Grossman M. (198) *Gen. Sel. Evol.* **21**: 246-477.
Georges, M., and J.M. Massey (1991) Theriogenology 35: 151-159.
Haley, C.S., and P.M. Visscher (1998) J. Dairy Sci. 81(2): 85-97.
Hayes BJ, Goddard ME (2001) Genet. Sel. Evol. **33**:3-13.
Janns LLG, et al.(1995) Theor. Appl. Genet. 91: 1137-1147.
Meuwissen THE, and Goddard ME (2001). Genet. Sel. Evol. **33**:605-634.
Meuwissen THE, Hayes BJ, Goddard ME, 2001, Genetics 157: 1819.
Stricker C. et al.(2002) 7th WCGALP **21**: 12.
Windig, J.J. and Meuwissen, T.H.E. (2004) *J. Anim. Breed. Genet.* **121**: 26-39