Strategy for Applying Genome-Wide Selection in Dairy Cattle

L. R. Schaeffer

Centre for Genetic Improvement of Livestock

Department of Animal & Poultry Science

University of Guelph, Guelph, ON, Canada N1G 2W1

E-mail: lrs@uoguelph.ca

FAX: 519-767-0573

Date Mailed: February 7, 2006

Strategy for Applying Genome-Wide Selection in Dairy Cattle

Summary

Animals can be genotyped for thousands of single nucleotide polymorphisms (SNPs) at one time, where the SNPs are located at roughly 1 centiMorgan intervals throughout the genome. For each contiguous pair of SNPs there are four possible haplotypes that could be inherited from the sire. The effects of each interval on a trait can be estimated for all intervals simultaneously in a model where interval effects are random factors. Given the estimated effects of each haplotype for every interval in the genome, and given an animal's genotype, a 'genomic' estimated breeding value is obtained by summing the estimated effects for that genotype. The accuracy of that estimator of breeding values is around 80%. Because the genomic EBV can be calculated at birth, and because it has a high accuracy, a strategy that utilizes these advantages was compared to a traditional progeny testing strategy under a typical Canadian-like dairy cattle situation. Costs of proving bulls were reduced by 97% and genetic change was increased by a factor of 3 to 4. Genome wide selection will become a popular tool for genetic improvement in livestock.

(**Keywords:** progeny testing, SNPs, strategy, haplotype intervals, expected progress)

Introduction

MEUWISSEN, HAYES, and GODDARD (2001) proposed methods of predicting total genetic value using a genome-wide dense marker map from a limited number of phenotypic records using marker haplotypes. The markers were taken to be single nucleotide polymorphisms (SNPs) because over 500,000 SNPs are already known in the human genome and for dairy cattle there are chips with 10,000 SNPs, with more SNPs being discovered every day. Thus, the possibility exists of covering the entire genome with markers located no more than 1 cM apart. The markers are assumed to be in linkage disequilibrium.

For a genome of 3000 cM, only 3001 markers at 1 cM intervals are needed, but these need to be informative, so that a panel of 10,000 markers or more should increase the chances of success. At present, a chip for 10,000 SNPs is available for dairy cattle at a cost of less than $400 Cdn per animal. (All cost figures are presented in Canadian dollars.) For each contiguous pair of markers, the haplotypes inherited from the sire need to be constructed. Because SNPs have a single base pair difference there are just two alleles for each marker (usually), and therefore, for a pair of markers there are four possible haplotypes. The frequencies of each haplotype will depend on the frequency of alleles at each marker, and the distance between the markers as per recombination events. Enough animals need to be genotyped so that all haplotypes are represented in animals with records.

In a simulation study, MEUWISSEN, HAYES, and GODDARD (2001) (MHG-2001), compared least squares, BLUP, and Bayesian approaches for estimating the effects of each haplotype pair simultaneously. In a genome of 1000 cM QTLs were inserted evenly, and the true breeding value was the sum of the QTL effects. Markers were placed at 1 cM intervals throughout the

genome. The effects of marker haplotypes were estimated for each interval (1000 in total). Then using the genotype of the animal and the estimated haplotype effects, an estimated breeding value was calculated as the sum of the haplotype effect estimates corresponding to the genotype of the animal. This EBV will be denoted as GEBV, for genome wide EBV. The estimated haplotype effects are assumed to be general population estimates and not specific to any one animal or group of animals. Thus, GEBV could be calculated for resulting progeny as long as they were genotyped and marker haplotypes determined. The remarkable features of this approach are that the correlation of GEBV with true breeding values was 0.85 (regardless of heritability), and that animals could receive a GEBV at birth with this accuracy. Usually in dairy cattle, females seldom reach this level of accuracy, and bulls take six years or more to reach this accuracy in their EBVs.

A similar simulation study by Kolbehdari et al. (2006) verified the results of MHG-2001 using different heritabilities, and either evenly or randomly spaced QTLs. Correlations between GEBV and true breeding values of around 0.80 were found. Therefore, assuming that GEBV with high levels of accuracy are achievable at an early age, the question is how to take advantage of these properties. How can the traditional progeny testing scheme be modified (or replaced) in order to make faster genetic change? How much faster can it be? The purpose of this paper was to look at these questions.

Traditional Progeny Test Scheme

The Logistics

A traditional progeny test timetable is given in Table 1. The Canadian Holstein population

was used as the example in this paper. Roughly 1000 elite females are identified each year as dams of young bulls, and these are mated to specific sires. Usually an elite female has completed at least two lactations, is classified Very Good or better for type, has a daughter or two (or a son in AI), and has a solid family history of good breeding. The elite dams produce between 400 to 600 young bull calves which are purchased by AI and moved to the stud. By one year of age the young bulls are test mated to the population (500 - 800 matings) in order to have 100 daughters in their first EBVs for production, conformation, fertility, and longevity. About 43 months later the daughters from these matings complete their first lactations and the young bull EBVs for production are produced with an accuracy of about 75%. At this point the young bull is proven and may be culled or returned to service.

CHESNAIS (Personal Communication, 2005) suggests that the cost of proving one bull is about $50,000, which includes housing and feeding of the bull, collection and storage of semen, test matings, and incentives for producers to classify daughters and insure that each have test day records for production. Assume that 500 young Holstein bulls are tested per year. At $50,000 per bull, the cost to AI would be $25 million per year. The total time commitment per bull is 64 months from conception to first proof. If only 20 out of 500 bulls are returned to service, then the actual cost per bull returned to service is $ 25 million divided by 20 or $1.25 million. The very best bull, genetically, however, may bring in several million dollars in revenue over several years. The goal of an AI organization is to find that one bull per year that is attractive to the entire world.

Predicted Change

The four pathways of selection for a trait with heritability around 0.4 was considered. Table

5

2 contains the assumed values for intensity of selection, accuracy of evaluations and generation
intervals for each pathway. Similar numbers have been presented in several books or papers
(BOURDON 2000, SCHMIDT and VAN VLECK 1974). For each pathway, multiply the in-
tensity of selection times the accuracy of evaluation and sum results over pathways, and sum
the generation intervals over pathways. The result was 4.68 genetic standard deviations and 24
years, which gives 0.195 genetic standard deviations change per year.

Another scenario is given in Table 3 with greater selection intensities on sires of bulls and
sires of cows, and also a reduction of the generation intervals on sires of bulls and dams of bulls.
The resulting genetic change was 0.263 genetic standard deviations change per year. Generation
intervals are often longer than given in these tables due to selection on many traits and reluctance
of sire selection committees to use animals too early for parents of the next generation. Actual
genetic change would be lower than shown. The cost per one genetic standard deviation change
would be $128 million in Table 2, and $95 million in Table 3.

Genome-Wide Selection Scheme

The Logistics

There are many possible schemes that could incorporate the genome-wide selection strategy.
In this study, only changes to the traditional progeny test scheme were considered. At present,
10,000 SNPs can be genotyped at one time at a cost of less than $400 per animal. The cost was
assumed to be $500 per animal for this study. In the future, the cost of genotyping per animal
could drop dramatically as the volume of animals to be genotyped increases, and as molecular
genetic techniques improve. As will be seen later, the cost of genotyping is a minor factor in

this analysis.

The best possible statistical methods were assumed to be used to estimate haplotype effects and that sufficient data were available for this purpose (this will be examined more closely in the discussion section). The Bayesian method as suggested by MHG-2001, for example, could be used. The accuracy of a GEBV derived from the estimated interval effects was assumed to be 0.75 rather than 0.85 reported by MHG-2001.

An AI unit would begin by searching for elite dams. All possible candidates would be geno-typed and a GEBV with accuracy of 0.75 would be computed. The GEBV could be combined with the EBVs derived from actual data records. Assume that 1000 dams are genotyped for a cost of $500,000 per year.

Assume that 500 bull calves are born and each of those are genotyped to give a GEBV with accuracy 0.75. The top 20 are selected for purchase. The top 2 or 3 are designated for matings to elite dams in the next year. The cost of genotyping these bulls would be $250,000. The cost of buying 20 bulls might be $100,000. Space for 480 bulls would be freed up in the stud. The 20 bulls could be progeny tested, but could be used on the general population at one year of age just as any other proven young bull. Because the bull would go into service at one year of age, the semen fertility level would be at its highest point, and therefore, non-return rates should be high.

The total cost of genotyping dams and young bulls, plus buying 20 young bulls would be at most $850,000 per year. Some dams may be used more than once, particularly if the calf was female, and therefore, only newly chosen elite dams need to be genotyped each year, but this possibility is ignored for now. The cost of $850,000 represents only 3.4% of the $25 million used

7

¹²⁴ in the traditional progeny test scheme.

Predicted Change

¹²⁶ The availability of GEBV at birth with an accuracy of 0.75, can greatly reduce generation

¹²⁷ intervals if they are used at one year of age or less. The generation intervals of sires of bulls, sires

¹²⁸ of cows, and dams of bulls could all be reduced to 1 year. The dams of cows pathway would be

¹²⁹ unaffected because the general population would not be genotyped and selection pressure could

¹³⁰ not be increased. The accuracy of their EBVs could be argued to increase slightly if GEBV are

¹³¹ incorporated into the genetic evaluation system of all animals. However, usual EBVs in this

¹³² study were assumed to be based only on data records without GEBVs.

¹³³ The selection intensity or number of sires of bulls, sires of cows, and dams of bulls was

¹³⁴ assumed to be the same as in Table 2. If sires of bulls are chosen at one year of age on the basis

¹³⁵ of GEBV, then the accuracy is only 0.75 instead of 0.99 as is the case in the current progeny

¹³⁶ testing scheme. The accuracy of evaluation of sires of cows remains at 0.75, and the accuracy of

¹³⁷ dams of bulls increases from 0.60 to 0.75. Values are shown in Table 4. The pathway providing

¹³⁸ the most genetic change is the dams of bulls pathway rather than sires of bulls. Genetic change

¹³⁹ under a GEBV scheme would be 0.758 genetic standard deviations per year, which is 3.88 times

¹⁴⁰ greater than the results from Table 2. Given the costs of the GEBV scheme in the previous

¹⁴¹ section, the cost of one genetic standard deviation change would be $1.12 million. Compared

¹⁴² to $128 million (Table 2) or $95 million (Table 3) under the traditional progeny test scheme,

¹⁴³ genome-wide selection is very efficient. Increasing selection intensities, lowering the cost of

¹⁴⁴ genotyping an animal, and increasing the accuracy of GEBV only increases the advantage of

¹⁴⁵ genome-wide selection.

8

Discussion

Genome-wide selection has yet to be proven, but a simulation project could answer this question. Assuming that the results are positive, then there will be little choice (based on economics) except to adopt this strategy in place of traditional progeny testing because the savings could be $24 million per year to the Canadian AI industry. Emphasis will shift to the cow side of the pedigree and this could help reduce problems of inbreeding caused by too few sires of replacements. Refinements of the strategy on number of elite dams that are needed, number of bulls that need to be purchased, and number of sires of bulls could all be subject to major changes.

A nucleus herd or a consortium of herds should be established by the AI organization with about 10,000 cows in total. The AI organization may decide to own the herds or have binding contracts with the herd owners, and this could be financed by the $24 million savings from the genome-wide selection strategy. Every cow in the consortium would be genotyped ($5 million initially) and data on more traits would be collected on these animals than from cows in the general population (through incentives to owners). The data would be used for re-estimating the haplotype interval effects every year or two, with either the same SNP panel or new sets of SNPs. Every year, the new potential female replacements would be genotyped, and GEBV used to select the better females. Young bulls would eventually come from the consortium rather than the general population or breeder herds, because the females in the consortium would be genetically superior to all females outside the consortium by a wide margin. The remaining money could go towards research projects aimed at efficiency of production, locating major QTLs (with the largest haplotype interval estimates), or reducing the price of semen because

168 individual bulls will not be as important as they are today.

169 Routine genetic evaluations of bulls and international comparisons of bulls that are common
170 today may become less important under genome-wide selection. GEBVs may be calculated
171 within an AI organization and perhaps not shared with competitors. Comparisons would be
172 between country schemes rather than between bulls. The country that collects the best data
173 and estimates the haplotype interval effects most accurately will succeed better. Of interest to
174 researchers would be the estimates of haplotype interval effects. Holstein populations in different
175 countries may have different frequencies of haplotypes, and different estimates of interval effects.
176 However, if the haplotype interval effects were similar, they would provide a means of ranking
177 animals internationally.

178 Methods for estimating haplotype interval effects will need to be developed to handle the
179 large number of intervals, genotypes per animal, and multitude of traits. One set of animal
180 genotypes can be used with any trait. The estimates of interval effects would differ by trait. By
181 taking the sum of absolute values of haplotype effects within an interval, those intervals with
182 the larger values likely contain one or more QTL. The majority of intervals, however, should
183 have relatively small effects (as in the infinitesimal model). MHG-2001 removed the intervals
184 with small effects, but found a large reduction in the accuracy of GEBV.

185 Conceptually, the GEBV model is not very complicated. The animal effects in a genetic
186 evaluation model are replaced by haplotype interval effects, which are random. There is no
187 relationship matrix or IBD matrix required. The question is whether the same variance ratio
188 should be used for all intervals or if there should be interval specific variances. In MHG-2001,
189 there was little difference in accuracy of GEBV when a common ratio was used for all intervals or

when separate ratios were estimated. Ideally, if many traits are recorded, multiple trait analyses may be better for estimating haplotype effects. If the spacing of SNPs can be made more dense (every 0.5 cM), then maybe the accuracy of GEBV can be increased to 0.90. The problems will be exciting to study.

Conclusions

The advantages of a genome-wide selection scheme are too great to ignore. Genetic change can be 3 to 4 times greater than the current progeny testing schemes, and the savings in logistical costs could be 97% of today's costs. The country that adopts this strategy the earliest will have a major start over other countries, and this could help achieve international dominance in dairy genetics. Genome-wide selection has greater potential than nucleus, MOET, or marker assisted schemes for making genetic change. Costs of genotyping are also likely to decrease over time which would make genome-wide selection easier to administer.

There will be an initial start-up period for a country in which animals will need to be genotyped in order to estimate the haplotype interval effects. Work will be needed to identify informative SNPs and to write software for constructing haplotypes from the SNP genotypes. The best method of analyzing the data also needs study, but even non-optimum methods appear to give good accuracy of GEBVs. The above costs have not been built into the previous comparisons because they are start-up costs similar to the development of new software for a random regression model. Milk recording programs should continue, but may focus more on management assistance rather than collecting data for genetic evaluation. Consortium herds would be a source of high quality and complete data on all manner of traits, but because the number of such herds will be small, the cost of data collection can be supported by AI. AI units

may want to own the herds or at least have exclusive rights to ensure quality data collection and access to future young bulls.

Dairy producers that are not part of the consortium could also want GEBVs on their cows, but this would be at their own expense. This might help them to stay competitive with the consortium herds. Alternatively, some herds may wish to genotype all cows, but not collect data, and other herds might want genotypes and provide data. An AI unit might benefit from having a large consortium of 50,000 to 100,000 cows that have GEBVs, but costs will dictate how large this group might go.

MEUWISSEN, HAYES, and GODDARD (2001) are to be commended for making this highly significant and very exciting proposal for genetic improvement of livestock. Genome-wide selection will become common place in the not too distant future.

References

BOURDON, R. M. 2000. *Understanding Animal Breeding.* Prentice-Hall, Inc. Second Edition.

KOLBEHDARI, D., L. R. SCHAEFFER, J. A. B. ROBINSON. 2006. Estimation of genome-wide haplotype effects in half-sib designs. J. Anim. Breed. Genet. (submitted).

MEUWISSEN, T. H. E, B. J. HAYES, M. E. GODDARD. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.

SCHMIDT, G. H., L. D. VAN VLECK. 1974. *Principles of Dairy Science.* W. H. Freeman and Company.

Table 1.

Timetable of Progeny Testing Activities.

Time Months	Activity
0	Elite dams chosen and bred.
9	Bull calves born from elite dams.
21	Test matings of young bulls made.
30	Daughters of young bulls born.
45	Daughters of young bulls bred.
54	Daughters calve and begin first lactation.
57	First EBVs for young bulls from test day model.
64	Daughters complete first lactations,
	Keep or cull young bulls.

Table 2.

Four Pathways of Selection, Scenario 1 Progeny Testing.

Pathway	Selection		Accuracy,	Generation	
	Percentage	i	r_{TI}	Interval, L	$i \times r_{TI}$
Sire of Bulls	5	2.06	0.99	10	2.04
Sire of Cows	20	1.40	0.75	6	1.05
Dams of Bulls	2	2.42	0.60	5	1.45
Dams of Cows	85	0.27	0.50	3	0.14
Totals				24	4.68

15

Table 3.

Four Pathways of Selection, Scenario 2 Progeny Testing.

Pathway	Selection Percentage	i	Accuracy, r_{TI}	Generation Interval, L	$i \times r_{TI}$
Sire of Bulls	2	2.42	0.99	9	2.40
Sire of Cows	5	2.06	0.75	6	1.54
Dams of Bulls	2	2.42	0.60	3	1.45
Dams of Cows	85	0.27	0.50	3	0.14
Totals				21	5.53

Table 4.

Four Pathways of Selection, Genome-wide strategy.

Pathway	Selection		Accuracy,	Generation	
	Percentage	i	r_{TI}	Interval, L	$i \times r_{TI}$
Sire of Bulls	5	2.06	0.75	1	1.54
Sire of Cows	20	1.40	0.75	1	1.05
Dams of Bulls	2	2.42	0.75	1	1.82
Dams of Cows	85	0.27	0.50	3	0.14
Totals				6	4.55

17