Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Challenges in Industry Application of Genomic Prediction Experiences from Dairy Cattle

E.A. Mäntysaari
MTT Agrifood Research Finland, Biotechnology and Food Science, Genetic Research

ABSTRACT: In dairy cattle breeding the genomic selection (GS) has been adapted fast. Many countries use almost only genomic selected young bulls, and almost all bulls entering AI have GS bulls as sires. The shift into full GS has been fastest in Holstein, while in the other breeds lower accuracy of genomic evaluations has hindered the change. The research on methodology has not proven advantages of variable selection methods or high density genotyping overwhelming, and most the evaluations are based on genomic relationship BLUP. The general trend is to genotype more and more cows and the cows are included in the reference population, and/or planning is to move into single-step type of evaluations. The use of younger bulls, and bull sires can lead into higher increase in inbreeding than before. Another danger to genetic diversity is the problems in implementing genomic selection with smaller, other than Holstein, breeds.
Keywords: dairy cattle; genomic selection; genomic evaluations

Introduction

In February 2006 Schaeffer submitted his revolutionary article "Strategy for applying genome-wide selection in dairy cattle" (Schaeffer (2006)). By deterministic calculations he showed that cattle breeding program based on genome-wide selection could double the genetic progress given by conventional progeny testing scheme with the 92% less cost. The calculations were based on genomic evaluation idea of Meuwissen, et al. (2001) who stated "Prediction of total genetic value using genome-wide dense marker maps" could reach an accuracy of 0.85, i.e., almost the accuracy of progeny test, the engine of dairy breeding programs. The reason why genome wide selection fits so well in dairy cattle is the universal AI breeding program. The number of AI bulls is small, each bull can have huge number of daughters, and even the tested but culled bulls will get very accurate breeding value estimates (EBV). Thus, the history data gives excellent data design to estimate genomic model effects, and when the genomic estimated breeding values (GEBV) become available, they can be used instead of progeny test. In the simplest dairy cattle genomic selection (GS) scheme the accuracy and selection differential are like in progeny test based AI program, but the generation interval is more like in pedigree based selection in pig breeding.

Meuwissen et al. (2001) speculated that "the advent of DNA chip technology may make genotyping of many animals for many of these markers feasible (and perhaps even cost effective)". Schaeffer (2006) had already got a taste of first 10000 SNP arrays from Affymetrix Inc (2005). The breakthrough of SNP technology really was the Illumina Bovine50k SNP (Van Tassel, et al. (2007)) that came into use in 2008 and into public market 2009. Already the first published results of applying SNPs in genomic evaluations (VanRaden et al. (2009)) proved the prediction accuracies of Meuwissen et al. (2011) to be reachable. Using SNP data from 3576 US Holstein bulls the accuracies for milk and protein genomic evaluations were estimated to be 0.76 and 0.75.

At the same time with publication of first real data results, dairy bull populations were rapidly genotyped all over the world. The Illumina Inc (San Diego, California, USA) company alone has now sold more than 1.7 million Bovine DNA chips (Andre Eggen pers. comm.). The genotypes of Holstein bulls with daughter performance are already available for more than 50,000 bulls and the number of genotyped bulls for other breeds is another 30,000. In 2012 Interbull received 57,902 Holstein young bull genome estimated breeding values (Sullivan and Jacobsen, (2013)). However, the number of selection candidates genotyped is increasing exponential, North American genome database alone received 30,000 Holstein bull genotypes last year (Wiggans 2013, unpublished).

The objective of this review is to summarize the means how genetic evaluation centers are using the genomic data and how breeding companies have taken the GEBV into use. Finally some new possibilities for the use of genomic information, and some possible dangers of genomic revolution are discussed.

Genomic evaluations

The first paper of genomic evaluations (Meuwissen et al. (2001)) presented three estimation methods that are still widely in use: BLUP estimation and Bayesian estimation models BayesA and BayesB. After that a plethora of methods have been developed. For the basic genomic prediction there does not seem to be a single uniformly best method (see review Daetwyler et al. (2013)). Moreover, for routine use, the methods have to be accurate in prediction, but also practical in use. Table 1 lists the methods that are in use in the countries that have submitted their genomic evaluations to Interbull for the GEBV validation test. The majority of the evaluations (9/17) are based on GBLUP, or BLUP estimation of SNP effects,

Table 1. Genomic evaluations for production traits in 17 countries that have submitted their evaluations into Interbull GEBV validation test and that filled the GENO form describing the GEBV methodology in use.

	Breed[1]	Model[1]	Ref Bulls(Cows)[3]	Genotype Consortium[4]	GENO form date
Australia	HOL JER	GBLUP GBLUP	3553* (9604) 948 (4247)		2012-12-19 2012-12-19
Belgium	HOL	SS-GBLUP, DRP	(total ref. 2429)		2013-01-21
Canada	HOL	GBLUP + 20%	22853* total: 272296	Intercontinental	2012-04-18
Denmark, Sweden and Finland	HOL RDC JER	GBLUP + 0%	23956* 7200 2184	EuroGenomics DFS & Norway US	1012-03-17
France	HOL Montbeliarde Normande	Haplotypes+55% Haplotypes+56% Haplotypes+57%	23241* 6835 4970	EuroGenomics	2010-02-23
Germany and Austria	HOL BSW SIM	GBLUP +1% GGLUP +15,15,20% GGLUP +20,25,10%	25875* 4300, total 10300 7500, total 33000	EuroGenomics Intergenomics	2011-05-05 2011-10-31 2011-05-04
Great Britain	HOL	SNP-BLUP +10%	21041*	Intercontinental	2011-10-25
Ireland	HOL	GBLUP	5095; 7081 in 2012	NZ	2010-11-02
Italy	HOL BSW SIM	GBLUP + 1% ~BayesA + 10% PCA-BLUP	20964* 5803 1165	Intercontinental Intergenomics	2014-02-14 2010-11-12 2011-10-25
Japan	HOL	GBLUP+20,30,20%(mpf)	3600		2012-04-12
New Zealand	HOL & JER	SS-GBLUP, DRP	HOL: 2626 JER: 642 (total ref. 85500)	Ireland	2010-05-21
Poland	HOL	GBLUP	2731*	EuroGenomics	2010-05-21
Slovenia	BSW	~BayesA +10%	5083	Intergenomics	2010-11-12
Spain	HOL	R Boost	21641*	EuroGenomics	18-2-2014
Switzerland	HOL HOL&SIM BSW	BayesC + 0%	3503* 5560 4480	 Intergenomics	
The Netherlands	HOL	Bayes SSVS +x%	21871*	EuroGenomics	2010-07-09
United States	HOL JER BSW	~Bayes A + 10%	22867* (58539) 4212 (15671) 5404 (494)	Intercontinental Viking Genetics Intergenomics	2013-02-20

[1] Breed codes: HOLstein, JERsey, BSW: Brown Swiss, RDC: Red Dairy Cattle, SIMmental; [2]Methods and the proportion of polygenic variance assumed: GBLUP: SNP random regression with homogenous SNP variance or genomic relationship matrix based computations; ~Bayes A: VanRaden (2008); SS-GBLUP,DRP: Single step GBLUP based on deregressed cow genetic evaluations. [3] Size of the bull reference group or (cows) if included in reference, can be outdated if the GENOFORM date old, *Jakobsen and Sullivan, (Trait Specific computation of shared reference population, Interbull Tech Document, 2013); [4]Genotype sharing consortiums: Alliances sharing bull genotypes for reference.

which in the Interbull GENOform have been considered the same (see, Strandén and Garrick, (2009)). Only two evaluation centers have implemented Bayesian variable selection models (Netherlands and Switzerland). Also GEBVs for a number of Brown Swiss subpopulations are based on BayesA type of non-linear SNP estimation model that originally was developed for US evaluations (VanRaden, (2008)). Two countries use single step evaluation concept (Christensen and Lund (2010), Aguilar et al. (2010)). In Spain the evaluations are based on machine learning algorithm called Random-boosting (González-Recio, et al. (2013)). While all others are relying on SNP random regression concept, in France genomic evaluations are based on haplotype block assisted genetic evaluation (Boichard et al. (2012)).

Most urgent call for further development of genomic evaluation models comes from small breeds. The accuracy of GEBVs for Holstein seems to be in a level that justifies new breeding programs. For all other breeds the accuracies are lower, and the lift in a genetic progress is not as significant. One hope is in multi-breed evaluations, i.e., gaining information from Holstein to smaller breed evaluations. This might be more successful with variable selection methods than with genomic relationship based methods. Another urgent need of development is in use of genomic information in large national genetic evaluations. Based on the expectations (means of GEBVs), the young GS bulls will lead into 50-80% bigger yearly genetic progress than what was seen before 2010. Because GS is not based on phenotypic data, the EBVs of young bulls in national evaluations become underestimated (Patry and Ducrocq (2011)), the environmental trend becomes over estimated, and thereafter the evaluations of older bulls with second crop daughters will drop (Tyrisevä et al. (2014)). One method for combining genomic information into full genetic evaluation has been named single-step evaluation (Christensen and Lund (2010) and Aguilar et al. (2010)). Till now the implemented single-step evaluations require inverses of genomic relationship matrices of size number of genotyped animals, although versions without inverse have been already presented (e.g., Legarra and Ducrocq (2012)).

For validation of reliability of GEBVs a standard suggested by Interbull (Mäntysaari et al. (2010)) has been well adapted. In the validation test GEBVs are calculated from truncated data, and then used to predict the phenotypic records for the youngest animals in the full data. The statistic measures to publish are $R^2_{validation}$ (the coefficient of determination of regression model $Y=b_0+b_1*GEBV$, divided by the reliability of phenotypic record as a predictor of genetic value) and variance inflation factor b_1. While this already offers a standardization for the figures to publish, to give a value for comparison, the GEBV test protocol also recommends to perform the same test with conventional parent average. Still missing is a standard how the individual animal GEBV reliabilities should be calculated and published. There is a clear discrepancy between the validation R^2 and the average of model based reliabilities. This is becoming more important, when more and more of the GS bulls have also sires that have no progeny records. Individual bull reliabilities are needed also for the international GEBVs.

Changes in Breeding Programs

Since the beginning (Schaeffer (2006)), the focus in GS in dairy cattle has been on shortening the generation interval in male selection pathway. With low reliability in GEBVs, most breeding programs have started GS by pre-selecting bull calves to AI and progeny test. However, even with a moderate R^2 the greatest genetic gains are obtained by intensive use of young GS bulls. Moreover, the genetic progress is directly related to proportion of bull dams inseminated with GS bulls, unless the R^2 is very low (only 5% improvement over PA) or all the cows are already all inseminated with GS bulls (Thomasen et al. (2013)).

Figure 1 shows the changes taken place in breeding schemes of North American and Nordic Holsteins, Nordic Red Dairy Cattle (RDC) and German-Austrian Simmental. The populations and corresponding genomic evaluations all represent different character. The Holstein schemes all have very high accuracy of GEBVs and the difference in adaptation of new breeding scheme comes more from the different business models in North America and Nordic countries. The relatively slower change in RDC and Simmental breeding schemes is clearly a reflection of lower reliability of GEBVs.

The adaptation of the breeding program to GS reflects also into the number of GS bulls as bull sires. In the Holstein breeding in US and in Viking Genetics the proportion of bull sires without daughters was already 85-95% in 2013. In Viking Genetics' RDC scheme almost half of the bull sires still had daughter records, and in German-Austrian Simmental around 70% of young bulls were sons of progeny tested sires.

The change in breeding program has so far been driven by sire ranking list. When the top bulls are young GS bulls, proven bulls are not used in breedings, and young bulls become used as bull sires automatically. In the first visions, GS was assumed to reduce the increase of relatedness in the breeding nucleus. Now some breeders are observing just the opposite. Miglior et al. (2014) used pedigrees of young GEBV bulls that were submitted to Interbull to quantify the effect of GS to diversity in AI sires. As has been expected, globally the number of bull sires was now (10%) more, and the number of sons per bull sire was (40%) lower. However the yearly rate of inbreeding was 4 times the rate before genomic era. Although, the change is not as large if considered by generation, still the authors suggest that actions are needed to keep the inbreeding rates in acceptable level. Best action would be use of the optimal contribution selection, especially with accounting genomic inbreeding (Sonesson et al. (2012)). Also more emphasis given to new sire families helps, maybe by giving more weight on health traits. Furthermore, breeders should limit how long one bull is kept in service.

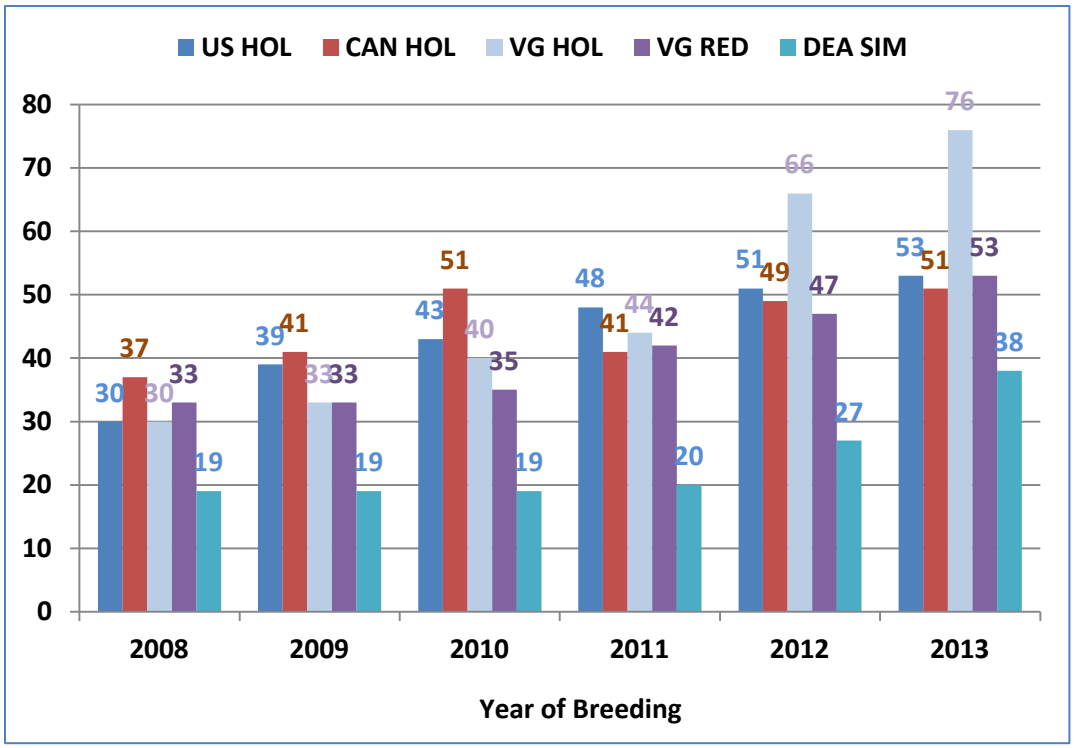

Figure 1. Proportion (%) of young non-progeny tested genomic selected bulls in breedings in North American and Viking Genetics (VG) Holstein, in VG Red and German-Austrian (DEA) Simmental populations.

Changes in operating environment

In attempt to improve the accuracy of GS the breeding organizations have formed genotype exchange alliances. Two large consortiums are exchanging Holstein genotypes: EuroGenomics includes Denmark-Finland-Sweden, France, Germany, The Netherlands, Spain, and Poland; and Intercontinental consortium includes USA, Canada, UK, and Italy. Both these compile into 20000-25000 bull genotypes reference population. Some Holstein operators (Ireland, New Zealand, Australia, Belgium, Switzerland) are not members of large consortia but sharing data case by case. Large consortium in number of members is Intergenomics, including seven Brown Swiss (BSW) countries. The BSW alliance operates in a different way than the others. They have contracted Interbull Center to compile the reference set of (currently ~6000 bulls) and to perform genomic evaluations. In other small breeds bilateral exchange is common, e.g. Nordic Jersey & US Jersey, Nordic Red & Norwegian Red.

Generally GS has chanced global trade and semen marketing to be more commercial. Before genomics the semen in the international marketing was certified by Interbull MACE proofs, now the marketing is based on skills of operators to persuade that their genomic evaluations are better, their traits are right, and/or they have the largest reference populations. The technological gap between exporters and importers is increasing and for an unexperienced buyer the label "genomic tested" can be sufficient mark of value. The only standard qualification for genomic evaluations is Interbull GEBV validation test (Mäntysaari et al. (2012)), but currently only un-biasedness is required. To protect the rights of the importers also an international standard for the accuracy of genomic evaluations should be agreed.

The GS has given Holstein breed global advantage over smaller breeds. The Holstein GEBVs are generally more reliable because of larger reference populations. Ironically global Holsteins also has smaller efficient population size than the "smaller breeds", making GS work better. The competitiveness of a breed is not only related to long term ΔG, but in a herd level, the breeders have a temptation to use young genomic tested Holstein bulls, when the top RDC, or Simmental or Jersey are only "old fashioned" progeny tested.

In North America the national genomic evaluations of bulls were restricted to owners of the reference genotypes until April 2013. After that date, every breeder, or companies in other sectors, can get GEBVs of their bulls. In Europe the genotypes are usually owned by farmers own cooperatives, and it is, in principle, their decision if they allow third party bull-calves to receive GEBVs. While in the future, running a breeding program does not necessarily require possession of cow resources, this opens possibilities for commercial breeding companies similar to ones in pig and poultry breeding.

Future challenges

There are two types of challenges in the future. Firstly, although the current GS has proven to be efficient, there are issues that have to be solved. Secondly, new technologies can give us completely new possibilities.

Optimistically we could expect that some of the new technologies can solve the current problems.

The GEBV methodological development could help on solving the problem of low reliability of GEBVs in small and admixed populations. Till now the computing methods, or/and higher genotyping densities have not shown promising results (e.g. Su et al. (2012)). Most promising seems to be to invest to large scale genotyping of females. In most populations this is done using lower density DNA chips (DFS, Australia, France, etc). Another methodological focus is in the development of single step type of evaluations. When the genetic progress becomes affected by the genomic selection, the multi-step approach cannot be used anymore, because of bias introduced by not accounting GS in candidates. Hopefully the computationally usable single-step algorithm, or alternative model, is presented in WCGALP 2014!

New future possibilities include the development of DNA technologies. The aim in the 1000 genomes project (see http://www.1000bullgenomes.com) is to collaboratively compile a reference set of full genome sequences of key bulls in breeding programs. This might be useful for the genome wide association studies but to be useful in genetic evaluations the number of sequences should be much more. Hickey (2013) discussed about the value of sequences, and suggested that they are useful only if millions of animals are sequenced. He speculated that a low coverage (0.1X) sequencing might actually be cheaper that the current DNA chips. More animals instead of denser genotyping policy has been perceived as winner in SNP arrays. Illumina Inc. introduced high density BovineHD chip in 2010. It has never become as popular as the BovineSNP50 chip. On the contrary, the sales of low density low cost chips have exploded so that LD sales represent now 85% of the total.

An increasing numbers of genetic defects are found in all breeds. This can be taken as a reminder of the importance of the within breed genetic diversity. Some defects are recessive lethals that were noticed by examining the haplotype segments (VanRaden et al (2011)), or deletions (Kadri et al. (2014)) which are difficult to recognize using the SNP data only. Recessive lethals can be seen as major genes affecting reproduction, and thus unraveling the gene actions. For the custom DNA chip designers the defects are great motivation. It is easier to justify the genotyping if the breeder gets, in addition to genomic prediction of total genetic merit, a gene test of one or few genetic defects. When the number of known defects become larger, it might be advisable to take a step backwards towards old breeding programs: design mating program such that excessive increase in relatedness is avoided.

Another attractive potential in GS is a possibility to develop genomic evaluations for novel traits that are difficult or expensive to measure from all animals. Examples of these are feed efficiency traits or detailed metabolic, health or reproduction traits. Some traits commonly mentioned are traits that can be included into breeding program also without genomics, like milk fatty acid decomposition or hoof trimming data. Feed efficiency complex is especially interesting because of its large and increasing economic value. The challenge is in obtaining a large and consistent cow reference population, and to maintain the data collection with new animals closely related to selection candidates. Because the heritabilities of the new traits are usually low, the number of cows phenotyped and genotyped must be in thousands, or more.

Conclusion

Genomic selection has changed the dairy cattle breeding enormously. In Holstein breeding many countries use almost only genomic selected young bulls, and almost all bulls entering AI have GS bulls - without progeny test result - as sires. In the other breeds, the change into full GS has been slower because of lower accuracy of GEBVs. The general trend is to genotype more and more cows and the cows are included in the reference population, and/or planning is to move into single-step type of evaluations. Because of cow genotyping the low density DNA chips are used much more than medium density chips. The use of younger bulls, and bull sires can lead into much higher increase in inbreeding than before. Another danger to genetic diversity is the problems in implementing GS with smaller – other than Holstein – breeds.

Acknowledgements

Many thanks to numerous colleaques in different organizations for their participation on the discussion, and for kindly contributing data and describing their operation in detail. Special thanks to Peter Sullivan, CDN, and Jacques Chesnais, Semex, Canada; George Wiggans, USDA, USA; Reiner Emmerling, Lfl, Germany; Lars Nielsen, Viking Genetics, Denmark. The data for the genetic evaluation practices was from Interbull Center, Sweden, but was clarified by individual evaluation centers

Literature cited

Affymetrix Inc. (2005). http://www.affymetrix.com/support/technical/datasheets/bovine10k_snp_datasheet.pdf. Accessed March 6, 2014.

Aguilar, I., Misztal, I., Johnson, D.L. et al. (2010). J. Dairy Sci. 93:743-752.

Boichard D., Guillaume F., Baur A., et al. (2012). Animal Prod. Sci 52, 115–120.

Christensen, O.F. and Lund, M.S. (2010). Genet. Sel. Evol. 42:2.

Daetwyler, H.D., Calus,M.P.L, Pong-Wong, R. et a. (2013) Genetics, Vol. 193, 347–365.

González-Recio, O., Jiménez-Montero, J.A., R. Alenda (2013). J Dairy Sci 96:614–624.

Hickey, J. M. (2013) J Anim. Breed and Genetics, 130:331–332.

Kadri, N. K., Sahana, G., Charlier, C. et al. (2014). PLoS genetics, 10(1), e1004049.

Legarra, A. and Ducroucq, (2012). J. Dairy Sci 95:1-17.

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001). Genetics 157: 1819-1829.

Miglior, F. Chesnais, J. Sargolzaei M, et al. (2014) In Advancing Dairy Cattle Genetics: Genomics and

Beyond 17-19.2 2014. Iowa State U. http://www.ans.iastate.edu/events/dairygenomics. Assessed 4.3.2014.
Mäntysaari, E.A., Liu, Z. and VanRaden, P. (2010). Interbull Bull. 41: 17-22.
Patry, C. and Ducrocq, V. (2011). J. Dairy Sci., 94:1011-1020.
Schaeffer, L. R. (2006). Journal of Animal Breeding and Genetics, 123(4), 218-223.
Sonesson, Anna K., John A. Woolliams, and T. H. Meuwissen. (2012) Genet Sel Evol 44: 27.
Strandén, I., and D. J. Garrick, (2009). J. Dairy Sci 92.6: 2971-2975.
Su, G., Brøndum, R., Ma, P. et al. J. Dairy Sci 95:4657-4665.
Thomasen, J.R. Egger-Danner, C., Willam, A. et al., (2014), J. Dairy Sci, 97:458-470.
Tyrisevä, A.-M., Lidauer, M.H., Aamand, G.P. et al (2014). The 10th WCGALP, 17.-22.8.2014, Vancouver, Canada.
VanRaden, P.M. (2008). J. Dairy Sci. 91:4414-4423.
VanRaden, P.M, Van Tassell, C.P., Wiggans, G.R, et al. (2009). J. Dairy Sci. 92:16–24.
VanRaden, P. M., Olson, K. M., Null, D. J., et al. (2011). J Dairy sci 94:6153-6161.
Van Tassell, C.P., Matukumalli, L.K., Taylor, C., et. al (2007). J. Dairy Sci., 90 (Suppl. 1), 421–422 (Abstr.).