Canadian Implementation of Genomic Evaluations

B.J. Van Doormaal[1], G.J. Kistemaker[1], P.G. Sullivan[1], M. Sargolzaei[2] and F.S. Schenkel[2]
[1]Canadian Dairy Network (CDN), Guelph, ON, Canada
[2]Centre for Genetic Improvement of Livestock (CGIL), Department of Animal & Poultry Science,
University of Guelph, Guelph, ON, Canada

Introduction

Effective August 2009, official genetic evaluations in Canada for the Holstein breed include genomic information in combination with traditional pedigree and progeny performance data. This follows the release by Canadian Dairy Network (CDN) of unofficial, research genomic evaluations on a monthly basis since April 2009. The main purpose of this paper is to summarize several elements of the official implementation of genomic evaluations in Canada.

North American Collaboration

Genomic evaluation services provided by CDN in Canada and AIPL-USDA in the United States are the product of a North American collaboration (VanRaden, 2009a) that began with the establishment of a Cooperative Dairy DNA Repository (CDDR). This CDDR traces back to 1992 when major A.I. organizations in North America started submitting DNA via frozen semen to the databank for every young sire sampled. Since 2008, a joint research and development effort has been in place involving scientists at AIPL-USDA, the Centre for Genetic Improvement of Livestock (CGIL) at the University of Guelph and CDN. This collaborative strategy included the sharing of all genotypes as well as the exchange of all computer code and research results from both countries. While the aim has been to use identical data and methods to compute genomic evaluations in each country, early research results have led to some current differences for official implementation.

Estimation of Direct Genomic Values

Computer code used for the estimation of Direct Genomic Values (DGV) was initially developed by AIPL-USDA and described by VanRaden (2008) and Schenkel (2009). Methods include four main steps; (a) de-regression of EBV to derive Daughter Deviation (DD) for each animal and trait, (b) estimation of allele frequencies in the base population, (c) formation of the genomic relationship matrix, and (d) set and solve mixed model equations, using a linear or non-linear model. Modifications to the original computer code and/or methods for estimating DGV and associated reliabilities were made at CGIL and CDN to reflect research results (Schenkel, 2009) and streamline data processing. Compared to the implementation in the US prior to August 2009, differences implemented by CDN include the following:

- Conducted full pedigree de-regression rather than using single generation pedigree.
- Used a linear mixed model.
- Polygenic effect of 20% rather than 5%.
- Applied discount factor of .5 to DGV reliability estimates prior to blending with EBV.
- Estimated SNP effects based solely on sires with official domestic EBV or MACE in Canada and excluded all cow evaluations. Appendix A provides a count of sires by evaluation type used for the estimation of SNP effects in Canada for August 2009.

To-date, all DNA genotyping within the CDDR collaboration has been done using the Illumina BovineSNP50^{TM} BeadChip, which includes nearly 57,000 single-nucleotide polymorphisms (SNP). In conjunction with the August 2009 genomic evaluation calculations in both the US and Canada, the minimum minor allele frequency required for inclusion of any SNP was decreased from 5% to 1%, which increased the total number of SNP analyzed from 38,416 to 43,385. Appendix A provides the count of genotyped sires used for the estimation of SNP effects, and a breakdown of the number with an official domestic EBV or a MACE evaluation, for each of the 63 traits analyzed. For most traits, roughly one-third of the bulls included had an official domestic EBV with the remainder being foreign proven sires (mainly US) that were genotyped. The entire Holstein genotype file included 28,042 animals, consisting of 2,375 bulls with an official LPI, 5,123 with a MACE LPI, 13,148 other males (mostly younger bulls) and 7,396 females (both cows and heifers).

Blending of DGV and Traditional EBV

VanRaden (2008, 2009b) used selection index to combine the direct genomic prediction from the SNP analysis, the PA or EBV computed from the subset of genotyped ancestors using traditional relationships and the published PA/PI or EBV. In Canada, an alternative blending approach was implemented based on research at CDN that examined the slope and R-Square of the regression equations for "reference" sires versus younger "prediction" bulls when current EBV is predicted from 4-year old evaluations with genomics (Sullivan, 2009; Kistemaker *et al.*, 2009). A weighted average approach based on the reliability of the traditional EBV/PA and that of the DGV, as described by Miglior *et al.* (2007), was found to yield similar R-square of prediction compared to the VanRaden selection index method but reduced differences in the regression slope such that young bull GPA are more fairly comparable to GEBV for progeny proven sires. This outcome is important since, with the incorporation of genomics, producers and industry personnel are more likely to directly compare bulls with GPA versus GEBV when making selection decisions.

Gains in Published Reliability

The average reliability of PA and GPA (and average gain with genomics) for selected key traits, including LPI and its three components, is presented in Table 1 based on genotyped Holstein bulls born between 2005 and 2009. In general, reliability gains with genomics are lowest for traits with no MACE information because (a) it is not a trait directly evaluated by Interbull (i.e.: Lactation Persistency, Dairy Strength, Rump), or (b) the US does not have national evaluations to submit to Interbull (i.e.: Milking Speed and Temperament).

Table 1. Average reliability of PA and GPA based on genotyped Holstein bulls born between 2005 and 2009 (August 2009 official evaluations).

Trait	Average Reliability (%)		
	PA	GPA	Gain
LPI	33	54	21
LPI – Production	34	58	24
LPI – Durability	34	53	19
LPI – Health & Fertility	29	46	17
Milk Yield	34	58	24
Fat Yield	34	58	24
Protein Yield	34	57	23
Fat Deviation	34	58	24
Protein Deviation	34	57	23
Conformation	35	54	19
Mammary System	36	55	19
Feet & Legs	34	51	17
Dairy Strength	36	55	19
Rump	34	50	16
Herd Life	29	48	20
Somatic Cell Score	36	56	20
Lactation Persistency	28	45	16
Daughter Fertility	28	45	17
Milking Speed	30	45	15
Milking Temperament	27	40	13
Calving Ability	38	58	20
Daughter Calving Ability	28	43	15

Table 2 shows the average reliability for LPI of population sub-groups for evaluations before and after the inclusion of genomics, as well as the average relative weight given to EBV/PA versus DGV for blending. As expected, the gains in average reliability for LPI are highest for young bulls, heifers, foreign dams and domestic cows, which now all average over 60% reliability with genomics incorporated and trickled down within the Canadian genetic evaluation system.

Table 2. Average reliability for LPI before and after incorporating genomic information (DGV) into traditional evaluations, by sub-group within the Canadian Holstein population.

Population Sub-Group	Average Reliability (%)			Blending Weight on DGV
	Before	After	Gain	
Young bulls and heifers born in 2007 to 2009	34	54	20	61%
Cows in 1st or 2nd lactation	52	62	10	54%
Foreign cows with MACE in Canada	42	57	15	58%
1st Crop proven sires in Canada	85	88	3	51%
Foreign sires with MACE in Canada	69	77	8	53%

Processing Steps and Information Trickling

A desired objective at CDN is to use the most recent genetic evaluation possible based on traditional calculation systems (i.e.: domestic EBV, MACE or PA), in combination with the most recent set of genotypes available, to derive the most accurate estimate of each animal's genetic merit. For the August 2009 release, for example, this included all official domestic EBV for cows and progeny proven bulls as well as the August 2009 MACE evaluations for foreign sires as provided by Interbull. For foreign dams from the US, August evaluations were converted using Interbull equations while April 2009 evaluations, when available, were used for dams from other countries.

At least one week prior to the official release date, CDN receives the complete file of all genotypes associated with the CDDR agreement from AIPL-USDA. In addition, pedigree information for all genotyped animals is exchanged. Once the Interbull pedigree file is received in the week prior to official publication, all pedigree processing can be completed. Agreements between CDN, AIPL-USDA and Holstein USA allow for the confidential pre-release of national traditional EBVs, excluding any genomic information, for males and females required by each country.

Upon receipt of the official evaluation files from Interbull, usually 5-6 days prior to the national evaluation release date in Canada, the post processing at CDN commences, including the following steps:

- In a single record, extract the publishable traditional evaluation and the associated reliability for each trait, which would be (in descending order of preference) an official domestic EBV or a MACE evaluation. Pas are not required at this stage.
- Extract a complete pedigree file and a file with all available genotypes.
- Launch the Genomic Evaluation System that uses only bulls with an official domestic EBV or an official MACE evaluation in Canada to estimate SNP effects on a trait-by-trait basis. For each trait, the DGV and associated discounted reliability are estimated for all genotyped animals.
- Process animals from oldest to youngest to compute evaluations for official publication, which involves three distinct but simultaneous functions. These include (a) blending of traditional evaluations with DGV for genotyped animals, (b) "trickling" of each genotyped animal's genomic information down to its non-genotyped descendants (currently excludes trickling to ancestors and collateral relatives), and (c) "trickling" of any MACE information for foreign sires and dams down to their descendants. Trickling of genomic and/or MACE information is trivial for young animals with PA only but for animals with their own performance and/or progeny data, the process first requires partitioning out the PA portion of the original EBV and then adjusts it before recombining into the modified EBV.
- Generate a single output record for every male and female with the final official evaluation to be published for each trait. For genotyped animals, an additional output file is generated including three pairs of evaluation and reliability, for the traditional evaluation after MACE trickling, the DGV, and the blended evaluation.

For the August 2009 official genetic evaluation release, the above required a total elapsed time of 15 hours.

Publication and Labelling

All officially published evaluations for every animal are available on the CDN web site (www.cdn.ca), including publicly accessible bull and cow files. Animals with their own genomic information included receive an evaluation labelled as GEBV, GMACE or GPA depending on the traditional evaluation information that was included. All animals with a genomic evaluation have a Genomic Evaluation Details page linked to its Genetic Evaluation Summary page.

Top bull lists for LPI (and other traits) or MACE LPI (for foreign sires) include only progeny proven sires and those with genomic information are labelled "G" in the Genomic Status (GS) column. For cows, official top lists for LPI and other traits are based solely on cows that have their own genomic information included. Similarly, a top heifer list for GPA LPI is produced by CDN to identify the highest genomically tested heifers. For females without genomic evaluations, CDN publishes separate reports for the highest LPI cows and the highest PA LPI heifers to identify the best candidates for future genotyping.

Delivery of Genomic Evaluation Services

Effective August 2009, Canadian official genetic evaluations for Holsteins include genomic information. CDN will provide monthly updates of genomic evaluations to Canadian owners of genotyped females. For genotyped males, unofficial genomic evaluations are provided to the authorized A.I. company that submitted the bull's DNA for genotyping. Official GPA for genotyped males will be published after being enrolled at NAAB or reaching two years of age,whichever comes first.

References

Kistemaker, G.J. & Sullivan, P.G. 2009. Experiences in validating genomic evaluations. *Report to Interbull Open Meeting, Barcelona, Spain, August 21-24, 2009.*

Miglior, F., Sullivan, P.G. & Van Doormaal, B.J. 2007. Optimum use of MACE information to predict composite type traits. *Interbull Bulletin 37*, 102-106.

Schenkel, F.S., Sargolzaei, M., Kistemaker, G., Jansen, G.B., Sullivan, P.G., Van Doormaal, B.J., VanRaden, P.M. & Wiggans, G.R. 2009. Reliability of genomic evaluation of Holstein cattle in Canada. *Interbull Bulletin 39*, 51-58.

Sullivan, P.G. 2009. Options for combining direct genomic and progeny test results. *Report to Genetic Evaluation Board of CDN, June 24, 2009 (Unpublished).*

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci. 91,* 4414-4423.

VanRaden, P.M., Wiggans, G.R., Van Tassell, C.P., Sonstegard, T.S. & Schenkel, F.S. 2009a. Benefits from cooperation in genomics. *Interbull Bulletin 39,* 67-72.

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S. 2009b. Invited Review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci. 92,* 16-24.

VanRaden, P.M. & Sullivan, P.G. 2009c. National and international genomic evaluations for dairy cattle. *J. Dairy Sci. E-Suppl. 1, 92,* 175.

Appendix A: Count of genotyped sires by evaluation type used for SNP effect estimation within trait for the August 2009 official genomic evaluations in Canada.

Trait	Total Sires Included	Official Domestic Evaluation		MACE Evaluation in Canada
		Count	% of Total	
Lifetime Profit Index (LPI)	7498	2375	31.7	5123
LPI - Production	7498	2375	31.7	5123
LPI - Durability	7498	2375	31.7	5123
LPI - Health & Fertility	7498	2375	31.7	5123
Milk Yield	8014	2375	29.6	5639
Fat Yield	8013	2374	29.6	5639
Protein Yield	8014	2375	29.6	5639
Fat Deviation	8013	2374	29.6	5639
Protein Deviation	8014	2375	29.6	5639
Somatic Cell Score	7970	2373	29.8	5597
Lactation Persistency	2375	2375	100.0	0
Conformation	7487	2367	31.6	5120
Mammary System	7488	2365	31.6	5123
Udder Depth	7494	2372	31.7	5122
Udder Texture	2375	2375	100.0	0
Median Suspensory	7490	2368	31.6	5122
Fore Attachment	7496	2371	31.6	5125
Front Teat Placement	7496	2371	31.6	5125
Rear Attachment Height	7494	2372	31.7	5122
Rear Attachment Width	2375	2375	100.0	0
Rear Teat Placement	5670	2373	41.9	3297
Teat Length	7481	2360	31.5	5121
Feet & Legs	7446	2366	31.8	5080
Foot Angle	7486	2366	31.6	5120
Heel Depth	2375	2375	100.0	0
Bone Quality	2375	2375	100.0	0
Rear Legs Side View	7494	2370	31.6	5124
Set of Rear Legs	2375	2375	100.0	0
Rear Legs Rear View	7480	2363	31.6	5117
Dairy Strength	7481	2366	31.6	5115
Stature	7499	2374	31.7	5125
Height at Front End	2375	2375	100.0	0
Chest Width	7493	2370	31.6	5123
Body Depth	7493	2375	31.7	5118
Angularity	7485	2370	31.7	5115
Rump	7461	2372	31.8	5089
Rump Angle	7497	2372	31.6	5125
Pin Setting	2375	2375	100.0	0
Pin Width	7496	2372	31.6	5124
Loin Strength	2375	2375	100.0	0
Herd Life	7580	2372	31.3	5208
Direct Herd Life	7006	2375	33.9	4631
Indirect Herd Life	6577	1553	23.6	5024
Milking Speed	2903	2663	91.7	240
Milking Temperament	2518	2309	91.7	209
Daughter Fertility	7457	2348	31.5	5109
Age at First Service	2361	2361	100.0	0
Non-Return Rate in Heifers	4968	2351	47.3	2617
Interval from 1st Service to Conception in Heifers	2361	2361	100.0	0
Non-Return Rate in Cows	5648	2354	41.7	3294
Interval from Calving to First Service	7532	2351	31.2	5181
Interval from 1st Service to Conception in Cows	7535	2353	31.2	5182
Days Open	7524	2346	31.2	5178
Calving Ability	8595	3331	38.8	5264
Calving Ease for Progeny Born from Heifers	8587	3320	38.7	5267
Calving Ease for Progeny Born from Cows	3351	3351	100.0	0
Calf Survival for Progeny Born from Heifers	4779	3295	68.9	1484
Calf Survival for Progeny Born from Cows	3348	3348	100.0	0
Daughter Calving Ability	5450	2612	47.9	2838
Calving Ease of Daughters at 1st Calving	5497	2620	47.7	2877
Calf Survival at Daughters' First Calving	6637	2607	39.3	4030
Calving Ease of Daughters at Later Calvings	2624	2624	100.0	0
Calf Survival at Daughters' Later Calvings	2613	2613	100.0	0