International Genomic Cooperation – North American Perspective

***B. Muir*[1]*, B. Van Doormaal*[2] *and G. Kistemaker*[2]**
[1]*Holstein Association of Canada, Brantford, Ontario,* [2]*Canadian Dairy Network, Guelph, Ontario*
bmuir@holstein.ca

Abstract

Centers in USA and Canada (AIPL-USDA, Animal Research Services, Beltsville, Maryland and the Canadian Dairy Network, Guelph, Ontario) have been sharing genotypes in order to enhance genomic evaluation procedures in both countries. Although genotypes and genomic methodology are shared, no direct comparison has been made between genomic breeding values published on the two country scales. January 2010 official proofs, direct genomic values and traditional EBV (termed PTA in USA) from both countries were compared. In general, genomic parent averages for young bulls were more correlated across countries than parent averages using traditional evaluations only. However, for LPI / Net Merit and Conformation, correlations across countries for first crop bulls and cows were stronger for traditional breeding values than for genomic breeding values. Accuracy of predicting future proofs was greater when all North American bulls were included in the SNP estimation, compared to only having domestic bull proofs available, with both countries realizing similar levels of accuracy. Gain in published reliability with inclusion of genomic information was greatest for young bulls and heifers evaluated in the USA, but this is also a function of differing methods used rather than differing levels of accuracy achieved with genomics. Proven bulls gained more published reliability in their second country of proof since the traditional EBV in that country would be a MACE evaluation of lower reliability so more gain is possible through genomics. This report summarizes the current benefits of collaboration in North America and challenges still to face.

Introduction

The North American Collaborative Dairy DNA Repository (CDDR) was formed following an initiative starting in 1992 when major A.I. organizations in North America began banking semen on young bulls. More than 15 years later, scientists in both USA and Canada initiated a joint research and development effort involving the sharing of genotypes, methodology and research results. Both countries officially launched national genomic evaluation services in 2009. The main objective of this paper is to summarize the details of the current collaboration in North America with special attention to the gains in accuracy of prediction realized from sharing genotypes and the increase in genetic correlations of published evaluations across countries.

Genotyping Service and Genomic Database

Currently, in both countries, male hair samples can only be submitted for genotyping through authorized A.I. organizations. In April 2013, public access to obtain male genotypes will be granted. Female hair samples are submitted for genotyping on behalf of the national Holstein associations (Holstein Canada and Holstein USA), upon request by the owner. To date, all North American genotyping for inclusion in genomic evaluations have been conducted using Illumina's BovineSNP50^{TM} BeadChip. All labs submit genotypes to the USDA who in turn forward all North American genotypes and associated pedigrees to the Canadian Dairy Network (CDN). In addition, the USDA and CDN exchange pre-genomic genetic evaluations for genotyped females and their ancestors. The total number of Holstein genotypes exchanged as of February 2010 is provided in Table 1. Of the 24,705 male Holstein genotypes exchanged, approximately 9,300 serve as the reference population for estimation of SNP effects in Canada. The USDA has slightly more (older) bulls available for SNP estimation that do not have a MACE evaluation in Canada provided by Interbull. In addition to the Holstein genotypes, 3,825 Jersey genotypes (3,149 male and 676 female) and 1,416 Brown Swiss genotypes (1,347 male and 69 female) have been exchanged.

Table 1. Holstein genotype database.

	Males		**Females**		
Nation of Reg.	**Proven**	**Young Bulls**	**Cows**	**Heifers**	**Sub-Total**
CAN	1,929	2,150	1,180	1,529	6,788
USA	6,769	12,456	3,452	5,112	27,789
Other	293	1,108	454	977	2,832
Sub-Total	24,705		12,704		
Total	37,409				

*Source: CDN, February 2010.

Genomic Evaluation Procedures

Although the two collaborating countries share data and methodology, genomic evaluation procedures still differ. In the USA, progeny proven bulls as well as cows are used in the estimation of SNP effects, whereas, in Canada females are excluded from the analysis. In Canada, accuracy of prediction decreased with the inclusion of female breeding values in the estimation of SNP effects because evaluations for elite cows are subject to upward bias due to preferential treatment. This very minor difference in genomic methodology leads to several interpretation issues when examining GEBV across countries. The assumed genetic variance explained by the BovineSNP50^{TM} BeadChip in the USA is 90% and in Canada is 80%. Therefore, polygenic effects were 10% versus 20%, respectively.

USDA uses a selection index approach to combine the direct genomic prediction from the SNP analysis (DGV); the PA/EBV computed using the subset of genotyped ancestors using traditional relationships and the published (or classical) PA/EBV. Canada blends DGV and PA/EBV weighted by their respective reliabilities to obtain a combined genomic breeding value for genotyped animals (Van Doormaal *et al.,* 2009). The blended approach allows for more appropriate direct comparison in Canada between GPA of young bulls and GEBV of proven sires (Sullivan, 2009).

The result of exchanging data and sharing methodologies is that every genotyped animal in North America receives a genomic evaluation in both the USA and Canada, regardless of where the animal was originally genotyped or evaluated. A.I. organizations receive GEBV on each country scale and can select and market young bulls accordingly. Access to female genomic evaluations is made public through national breed associations and evaluation centers. Increased ability to access information on animals due to the incorporation of genomics promotes greater opportunities for across border merchandizing of genetics.

Predicting Future Proofs

Several studies have shown that ability to accurately predict future proofs for young animals using genomics improves as the number of proven genotyped bulls included in the reference population for SNP estimation increases. Researchers at CDN recently conducted a validation analysis by predicting January 2010 de-regressed proofs using February 2006 Parent Averages (unpublished). Three types of starting values were compared, a traditional parent average (PA), a genomic PA (GPA, animal's genotype combined with traditional PA) that resulted from including only Canadian proven genotyped sires in the SNP estimation, and a GPA resulting from SNP estimation using all North American genotyped progeny proven sires. Average predictive abilities (squared correlations, R^2) for each scenario are summarized in Table 2. Accuracy of prediction (R^2) recently calculated by USDA in a separate study (Wiggans *et al.,* 2010b) are provided for comparison for common traits. The average increase in accuracy of prediction (across the traits shown in Table 2) in Canada by using GPA versus PA was 0.09 by including Canadian proven sires as the reference population and 0.16 by including all North American sires.

Predicting an animal's future performance has become more accurate with incorporation of genomic information, and this gain increases as the number of bulls in the reference population increases. By using more predictor bulls, the reliability increases by supplying more data to estimate each SNP effect (VanRaden, 2009a). For the six traits in common in Table 2, accuracy of prediction calculated in the USA (Wiggans *et al.,* 2010b) was slightly lower for most traits, except fat and fat and protein percentage, likely due to the use of a non-linear genomic evaluation model by USDA. The USA recently showed that adjusting traditional evaluations for yield in cows prior to inclusion of their indexes into SNP estimation increased their usefulness (Wiggans *et al.*, 2010a). However, it is important to note that average accuracy of prediction for most traits in Canada is still higher with the exclusion of cow evaluations in SNP estimation and using different methodologies, with averages of 0.47 in USA and 0.48 in Canada for the common traits in Table 2.

Table 2. Accuracy of predicting future proofs.

Trait	PA	Officially Proven in Canada (GPA)	North American Proven (GPA)	USDA* (GPTA)
LPI	.090	.209	.292	-
Milk	.243	.391	.469	.427
Fat	.173	.363	.442	.465
Protein	.209	.368	.449	.398
Fat %	.287	.419	.548	.655
Prot %	.272	.385	.526	.615
SCS	.255	.285	.361	.310
Conf	.229	.279	.348	-
RP	.308	.301	.317	-
MS	.215	.274	.376	-
FL	.281	.220	.222	-
DS	.278	.378	.418	-

*USDA results Wiggans *et al.* (2010b).

For this study, DGV, EBV and GEBV (January 2010) for more than 10,300 females and 23,000 males were made available for comparison from CDN and the USDA. Animals that were included in the analysis had GPA/GEBV published both on the Canadian and USA scales. Correlations between DGV, traditional PA/EBV and GPA/GEBV published on each country's scale were calculated for several groups of animals and traits.

Effect of Domestic Versus Foreign Evaluations

Table 3 shows the correlation of DGV, EBV and GEBV for LPI and Net Merit as national indexes in Canada and USA, respectively, for first crop bulls (born 2001 or later) according to their official LPI status in Canada in January 2010 ('Domestic' if the bull met minimum requirements for an official proof in Canada and 'MACE' otherwise). Results showed that when bulls were evaluated in Canada (regardless if they were foreign (0.98) or domestically (0.99) proven) their traditional EBV was highly correlated to their GEBV. When the same animal groups were evaluated in the USA, their EBV and GEBV were only correlated by .92 and .86, respectively. In the latter case, the roles would be reversed and the larger set of bulls (N=2636) that were designated MACE in Canada would be USA domestic bulls. If we assume most of the current 'MACE' bulls in Canada in this analysis were from the USA, we can postulate that because female indexes were included in the estimation of SNP effects, bull EBVs were not as correlated to their eventual GEBV as they would have been if they were evaluated in Canada. Also, the USA and Canada use different de-regression procedures as well as different models (linear versus non-linear) and methods (no rescaling of G-Matrix in Canada) in their genomic predictions that could contribute to differences in correlations. Correlations between EBV and GEBV were lower for foreign proven bulls (0.98 when evaluated in Canada and 0.86 in the USA) reflecting the lower reliability of MACE proofs.

Table 3. Correlations between DGV, EBV and GEBV for LPI (CDN) and Net Merit (USDA) by official status in Canada.

Canadian LPI Official Status	Type of evaluation	USDA ($NM)	CDN (LPI)
MACE	DGV vs EBV	.85	.94
(N=2636)	DGV vs GEBV	.98	.99
	EBV vs GEBV	.92	.98
Domestic	DGV vs EBV	.78	.97
(N=1001)	DGV vs GEBV	.97	.99
	EBV vs GEBV	.86	.99

Correlations between Genomic Evaluations

Correlations between genomic evaluations for the same trait in Canada and the USA for first crop bulls were higher if the bull was first proven in the USA (MACE in Canada) and higher for Protein and Conformation compared to Herd Life and LPI/Net Merit (Table 4). The higher correlations for bulls first proven in the USA demonstrated the domination of USA bulls in the reference population for estimating SNP effects. The lower correlations for traits such as Herd

Life with Productive Life reflected the lower heritability of these traits as well as the larger methodological differences across countries for traditional evaluation, which is especially true between LPI and Net Merit.

Table 4. Correlation of genomic evaluations.

Trait Pair	Proven Bulls (1st Crop*)		Cows	
(CDN/USDA)	CAN	USA	CAN	USA
Protein/Protein	.92	.96	.90	.95
Conformation/PTAT	.88	.94	.92	.97
Herd Life/Prod. Life	.83	.88	.80	.77
LPI/Net Merit	.76	.79	.78	.77

*born 2001 or later

Correlations for Young Bull Evaluations

Correlations between similar traits evaluated under two systems (CDN and USDA) for PA and GPA of young bulls were calculated for several traits and separately for bulls born in the USA versus Canada. For all trait pairs, the correlation across countries was consistently higher for GPA than for traditional PA (Table 5). Therefore, for young bulls (in the absence of any daughter information) incorporation of genomic information strengthened the connection in estimated genetic merit of that animal across the border. Correlations between trait pairs (both PA and GPA) were generally higher for USA-born young bulls compared to Canadian-born, with the exception of Herd Life/Productive Life. The countries exchange pre-genomic genetic evaluations for production, type and SCS, as well as pedigree information, and the extent to which this additional information is incorporated into national evaluations and the timing of these incorporations differs between Canada and the USA. CDN uses cow evaluations from USA and several European countries as they are made available to incorporate in the Canadian evaluations For descendants. The USA does incorporate foreign dam information from Canada but traditionally evaluations from the previous genetic evaluation run have been used.

Table 5. Correlations among young bull evaluations.

Trait Pair	Can-Born Young Bulls		USA-Born Young Bulls	
(CDN/USDA)	PA	GPA	PA	GPA
Protein/Protein	.72	.86	.92	.93
Conformation/PTAT	.67	.86	.90	.90
Herd Life/Prod. Life	.70	.78	.68	.76
LPI/Net Merit	.56	.74	.66	.78

In most cases, the genomic breeding values (GPA/GEBV) were more correlated than traditional values (PA/EBV) across countries for the traits investigated. This was intuitive, as countries are now sharing genotypes in addition to traditional breeding values, which incorporate phenotype and pedigree, and the resulting combined information should be more related. However, for some groups of animals and traits (not shown), breeding values were more correlated across countries before inclusion of genomic information. Specifically, traditional proofs LPI/Net Merit and Conformation for proven sires and cows were more correlated than GEBV.

Gain in Published Reliabilities of Genetic Evaluations

Table 6 shows the change in published reliabilities for LPI and Net Merit for several groups of animals. The reliabilities for young animals nearly doubled with the inclusion of genotypic information in both countries. Reliabilities for traditional PA calculated by CDN for Canadian-born young animals were generally higher than those calculated by USDA for USA-born animals, likely due to more inclusion of foreign dam genetic information.

Table 6. Average published reliability by group.

Animal Group	USDA (Net Merit)		CDN (LPI)	
	PA/ EBV	GPA/ GEBV	PA/ EBV	GPA/ GEBV
CAN Young Bull	32	**66**	38	**63**
USA Young Bull	34	**66**	34	**61**
CAN Heifer	33	**66**	38	**64**
USA Heifer	34	**67**	35	**62**
CAN Cow	43	**69**	53	**67**
USA Cow	47	**70**	42	**65**
CAN Proven Bull	69	**80**	85	**89**
USA Proven Bull	75	**82**	69	**79**

Different average published reliabilities between the two countries result from differences in methodology to calculate genomic evaluations and their associated reliabilities. This is shown by the higher average reliabilities in the USA versus Canada (Table 6) even though the accuracy of prediction was shown to be nearly identical in Canada on average (Table 2). Differences between reliability across country should be communicated to potential buyers in order to avoid confusion when the end user inappropriately considers reliability to be the measure of accuracy rather than the R^2 (accuracy of prediction).

Recent Research and Development Collaborations

A joint ongoing research initiative between USDA-AIPL, University of Guelph and the Canadian Dairy Network (CDN) began in early 2008. Several research topics have been or are currently being jointly investigated, or will be individually investigated and then shared. These topics include:

- Enhanced accuracy of the de-regression step
- Options for combining DGV and traditional evaluations
- Validation tests
- Scaling of GEBV for proven versus young bulls
- Need for rescaling of G-Matrix
- GMACE methodology for Interbull
- Multi-trait/country genomic evaluations
- Strategies for imputing genotypes
- Filling in the 50K panel for low call rates
- Imputing across generations
- Imputing across panels (i.e., 3K to the 50K)

Important Issues Needing Further Discussion

When countries share genotypes in addition to traditional proofs that incorporate performance and pedigree information, differences in genomic predictions can be accounted for, which should increase the genetic correlations across countries. Within country, a better prediction of expected performance of young bulls should result in enhancing abilities to merchandise genetics across borders.

Although research organizations in both countries have worked hard to improve methodologies to enhance genomic evaluation procedures, several practical differences still exist. The scale (range) in young bull evaluations (GPTA) in the USA is larger than those for GPA in Canada and therefore additional interpretation is required to compare values across countries. In addition, with similar accuracies of prediction, published reliabilities in the USA remain higher than those in Canada due to differences in methodology. In an international marketplace, proper extension of why differences in published reliabilities can occur, even with similar levels of accuracy of prediction, should be made a priority.

The influence of including MACE proofs in genomic prediction is still not completely understood. Although the road to GMACE has been paved, many questions still remain unanswered. Roughly two-thirds of the progeny proven sires included in the SNP estimation in Canada have MACE evaluations with an average LPI reliability of 70%. The remaining one-third has an official LPI in Canada with an average reliability over 85%. Incorporating the genotyped foreign sires with a MACE evaluation in Canada increases the accuracy of prediction but may complicate interpretation of genomic evaluations for these sires and associated progeny until GMACE has been made possible.

Acknowledgements

The authors would like to thank the USDA-AIPL Agriculture Research Service for providing GEBV for all US animals.

References

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

Van Doormaal, B.J., Kistemaker, G.J., Sullivan, P.G., Sargolzaei, M. & Schenkel, F. 2009. Canadian Implementation of Genomic Evaluations. Report to Interbull Open Meeting, Barcelona, Spain, August 21-24, 2009. *Interbull Bulletin 40*, 214-218.

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

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

Wiggans, G.R., Cooper, T.A. & VanRaden, P.M. 2010a. Cow adjustments for genomic predictions of Holsteins and Jersey bulls *J. Dairy Sci. 93(Suppl. 1)*, (in press).

Wiggans, G.R., Cooper, T.A., VanRaden, P.M. & Silva, M.V. 2010b. Increased reliability of genetic evaluation for dairy cattle in the United States from use of genomic information. *WCGALP*, Germany (submitted).