Reliability of Genomic Evaluation of Holstein Cattle in Canada

F. S. Schenkel[1], M. Sargolzaei[1], G. Kistemaker[2], G.B. Jansen[3], P. Sullivan[2], B.J. Van Doormaal[2], P. M. VanRaden[4] and G.R. Wiggans[4]

[1]University of Guelph, Guelph, N1G 2W1 Ontario, Canada
[2]Canadian Dairy Network, N1K 1E5 Ontario, Canada
[3]Dekoppel Consulting, 10010 Chiaverano (TO), Italy
[4]Agricultural Research Service, United States Department of Agriculture, 20705-2350 Beltsville, MD, USA

Introduction

Researchers in Canada and the United States have been collaborating to develop and implement genomic evaluations aiming to fully integrate them into their national genetic evaluations for dairy cattle in 2009. Collaboration started with the Cooperative Dairy DNA Repository (CDDR) through contributions of semen from seven major artificial insemination (AI) companies in North America. In collaboration with Illumina, Inc., the BovineSNP50 BeadChip was developed to genotype about 57,000 single nucleotide polymorphisms (SNP) in a single assay (Matukumalli *et al.*, 2008).

A total of 38,416 SNP have been selected from the Illumina BovineSNP50™ chip for use in genomic evaluations. The discarded SNP are either problematic to score, have a minor allele frequency of less than 5%, or are highly correlated with other SNP (VanRaden *et al.*, 2009).

In 2009, genomic evaluations are expected to become official in both the United States and Canada and additional accuracy from genomic evaluations of genotyped animals will be propagated to relatives that have not been genotyped yet.

The objective of this investigation was to validate genomic predictions in Canada using Canadian domestic proofs and MACE proofs of foreign bulls, as part of collaboration amongst researchers from the Centre for Genetic Improvement of Livestock (CGIL)-University of Guelph, the Animal Improvement Program Laboratory (AIPL-USDA) and the Canadian Dairy Network (CDN).

Material and Methods

Genotyped Animals

Holstein bulls (11441) and cows (1472) were genotyped using the Illumina BovineSNP50™ Chip. These animals included bulls with semen from the Cooperative Dairy DNA Repository (Ashwell and Van Tassell, 1999), the National Center for Genetic Resources Preservation, ARS, USDA (Fort Collins, CO), and bulls and cows from research projects, contributing organizations, and from the recently established genomic evaluation program in the United States. The SNP in the chip were identified and selected to be polymorphic across a wide variety of breeds included in the International Bovine HapMap Project (Matukumalli *et al.*, 2008; Van Tassell *et al.*, 2008). Extraction of DNA and genotyping was conducted by: the Bovine Functional Genomics Laboratory, ARS, USDA (Beltsville, MD); the Division of Animal Sciences, University of Missouri (Columbia, MO); the Department of Agricultural, Food and Nutritional Science, University of Alberta (Edmonton, Canada); GeneSeek (Lincoln, NE); Genetics & IVF Institute (Fairfax, VA); and Illumina, Inc. (San Diego, CA). Scoring of marker genotypes was performed using Illumina's Beadstudio software. Genotypes were coded by the number of counted alleles, giving values of 0, 1, or 2. A code of 5 was used to indicate a missing genotype.

Traits

A total of 44 traits were analyzed and are presented in this report. These traits included 5 production traits, 28 type traits, and 11 functio-

nal and reproductive traits. The required proofs were provided by Canadian Dairy Network (CDN). Canadian domestic and MACE proofs were used in this investigation.

Deregression

Daughter deviations (DD) were estimated using four-year old proofs for prediction (DD_p), and using current 2008 proofs for validation (DD_v).

For DD_p, November 2004 MACE proofs of foreign bulls and February 2005 official Canadian domestic proofs were considered in addition to generations of ancestors going back to 1950. Two scenarios were considered: Use of Canadian domestic proofs only (Domestic) and use of the available official information (Official), i.e., domestic proofs for Canadian proven bulls and MACE proofs for other bulls registered in the CDN database.

For DD_v, November 2008 Domestic evaluations were used.

The DD were estimated using an iterative procedure for general pedigrees developed by Dr. Gerald Jansen, which utilizes a strategy to compute approximate reliabilities directly in a reduced animal model similar to the algorithm reported by Jamrozik *et al.* (2000).

The average reliability of DD_p from the four-year old proofs was 0.93, 0.88 and 0.89 for production traits, SCS, and type traits using domestic proofs only and 0.82, 0.68 and 0.67 using the available official information. Reliabilities were lower when all available information was used because many additional bulls were included with MACE evaluations.

For traits not yet evaluated in 2004/2005

Proofs for some reproduction and functional traits were not yet available in 2004/2005. Therefore, CDN carried out evaluations with four-year old data and current software to generate the August 2004 domestic proofs for 10 of those traits, which were used in this study. In this case August 2008 proofs were used to estimate DD_v for validation. The average reliability of DD_p from the four-year old proofs was 0.72 for these traits.

Genotype data

For traits evaluated in 2004/2005

The analyses included 12913 genotyped animals with registration identification in the CDN data base. From those, bulls that were born in 2000 or earlier with available February 2005 Canadian domestic proofs or available official proofs (February 2005 Canadian domestic proofs or November 2004 MACE proofs) were defined as “predictor bulls” for the Domestic and Official scenarios, respectively. Younger bulls born between 2000 and 2004 with available 2005 Canadian domestic parent average (PA), but no proof from daughters, were defined as “validation bulls”. Table 1 shows the distribution of genotyped bulls by category of predictor (for both the Domestic and Official scenarios) and validation bulls for milk production traits.

For traits not yet evaluated in 2004/2005

The analyses included 9,243 genotyped animals with registration identification in the CDN data base. From those, bulls born in 2000 or earlier and with available August 2004 Canadian domestic proofs were defined as predictor bulls. Younger bulls born between 2000 and 2003 with available 2004 Canadian domestic PA, but no proof from daughters, were defined as validation bulls. August 2008 DD_v of validation bulls were used to test the 2004 genomic predictions. Table 2 shows the distribution of genotyped bulls by category of predictor and validation bulls for maternal calving ease of heifers.

Genomic predictions

Predictions were computed by a linear genomic model, in which the traditional additive genetic relationship was replaced by a genomic relationship matrix (VanRaden, 2008). This model is equivalent to estimating genetic marker effects assuming equal genetic variance for all markers.

A total of 38,416 SNP were used in the genomic evaluation. This set of markers has been previously selected from the Illumina BovineSNP50™ chip for use in genomic

evaluations (VanRaden *et al.*, 2009). As suggested by Wiggans *et al.* (2008), in the future the number of SNP used might be increased by lowering the minimum minor allele frequency required, as the number of genotyped predictor bulls will increase.

Daughter deviations (DD_p) of predictor bulls were used to obtain four-year old direct genomic predictions (GBV) of current DD_v for validation bulls. As the genomic predictions include less information than the domestic PA, because genotypes are available for only a subset of the sires and grandsires, PA were also computed using traditional additive relationship matrix with only the subset of genotyped ancestors (subset-PA). The final genotype-enhanced PA (GPA) was then computed by an index that combined the PA, the predicted genomic breeding value (GBV) and the subset-PA, using the respective reliabilities of the three components to determine the appropriate index weights. The inclusion of subset-PA in the index allowed the difference between genomic and traditional evaluations to be considered.

Validation

Validation bulls had many daughters in DD_v, and zero daughters in the DD_p used for genomic predictions. Therefore, these bulls provide a realistic test for the ability to predict younger bull's genetic merit using genomic information.

The squared correlation (r^2) between four-year old PA and 2008 DD_v (r^2_{PA}) for validation bulls was calculated. Observed genomic reliabilities (Rel_{Obs}) were then obtained as the r^2 between GPA and 2008 DD_v (R^2_{GPA}) of younger bulls, after adjusting for error variance in the 2008 DD_v and for prior selection on pedigree, i.e.:

$$Rel_{Obs} = (r^2_{GPA} / Rel_{2008\ DD}) + [Rel_{PA} - (r^2_{PA} / Rel_{2008\ DD})]$$

Results

For traits evaluated in 2004/2005

Genomic predictions (GPA) for production, SCS and type traits showed increased r^2 compared to r^2 of traditional PA of validation bulls by using predictor bulls with domestic proofs only or official proofs (domestic + MACE), as shown in Table 3. The increases in r^2 were more pronounced when MACE proofs were also used. For production traits the average increase in r^2 of GPA was 0.27 points when Official prediction set was used in contrast to 0.15 points when Domestic prediction set was used. The comparable increases were 0.16 and 0.10 points for type traits.

Observed reliability of GPA for the validation bulls are shown in Table 4. For all traits, except Rear legs-rear view, the observed reliability of GPA based on both prediction sets (Domestic and Official) was higher than the reliability of PA. The gain in reliability over PA was more pronounced when MACE proofs were included. For production traits the average gain was 0.29 and 0.16 points using Official and Domestic prediction sets, respectively. For type traits the average gain was 0.18 and 0.12 points, respectively.

For traits not yet evaluated in 2004/2005

The increase in r^2 was small for the 10 functional/reproductive traits. The average increase in r^2 of GPA was 0.04 points using the Canadian domestic proofs (Table 5). The corresponding increase in observed reliability was also small. The average increase in reliability of GPA was 0.05 points (Table 6). MACE proofs were not available for these traits.

Implications

Genotype-enhanced PA showed increased r^2 and observed reliability for 42 out of 44 traits analyzed. Exceptions were the traits rear leg-rear view and interval from calving to first service, for which there was no observed gain from the genomic information.

The use of MACE proofs of foreign bulls for prediction increased r^2 and observed reliability when compared to the use of Canadian domestic proofs only. The use of MACE quadrupled the number of bulls in the

prediction set, which likely led to the observed increase in prediction ability of GPA, regardless of the fact that the average reliability of MACE proofs was lower than the corresponding reliability of domestic proofs. This illustrates the importance of large genotype data files for estimation of markers effects, which is a goal that can be more easily achieved through collaboration and sharing of genotype information.

The functional and reproductive traits analyzed in this research showed the smallest gains in using genotype enhanced PA from genomic evaluations based on domestic proofs only. For these traits, MACE proofs were not available in 2004 and, therefore, the effect of incorporating MACE information could not be assessed. However, MACE results are currently available for some of these traits, and based on the consistent benefits found from using MACE information, it is reasonable to assume that MACE information will also be useful to improve genomic predictions of functional and reproductive traits.

For prediction of GPA of current young bulls, predictor and validation bulls used in this study will be jointly used as predictor bulls (i.e. predicting future DD in 2012). This will provide around 1700 predictor bulls with domestic proofs and around 4700 predictor bulls with MACE proofs, which should result in even larger gains in observed reliabilities than the gains reported from the present study.

Acknowledgements

- Ontario Centre for Agricultural Genomics and L'Alliance Boviteq Inc. fund.
- Natural Sciences and Engineering Research Council of Canada and L'Alliance Boviteq Inc. Collaborative Research and Development grant.
- National Research Initiative grants, USDA.
- National Association of Animal Breeders (NAAB, Columbia, MO):
  ABS Global (DeForest, WI)
  Accelerated Genetics (Baraboo, WI)
  Alta (Balzac, AB)
  Genex (Shawano, WI)
  New Generation Genetics (Fort Atkinson, WI)
  Select Sires (Plain City, OH)
  Semex Alliance (Guelph, ON)
  Taurus-Service (Mehoopany, PA)
- American and Canadian Holstein Associations.

References

Ashwell, M.S. & Van Tassell, C.P. 1999. The Cooperative Dairy DNA Repository –a new resource for quantitative trait loci detection and verification. *J. Dairy Sci. 82(Suppl. 1)*, 54. (Abstr.).

Jamrozik, J., Schaeffer, L.R. & Jansen, G.B. 2000. Approximate accuracies of prediction from random regression models. *Livest. Prod. Sci. 66*, 85–92.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M., Heaton, M.P., O'Connell, J., Sonstegard, T.S., Smith, T.P.L., Moore, S.S. & Van Tassell, C.P. 2008. Development and characterization of a high density SNP genotyping assay for cattle. *Genome Res.* (submitted).

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

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

Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenchild, C.D., Moore, S.S., Warren, W.C. & Sonstegard, T.S. 2008. Simultaneous SNP discovery and allele frequency estimation by high throughput sequencing of reduced representation genomic libraries. *Nature Methods. 5*, 247-252.

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Schenkel, F.S. & Van Tassell, C.P. 2008. Selection of single nucleotide polymorphisms and genotype quality for genomic prediction of genetic merit of dairy cattle. *J. Dairy Sci.* (submitted).

Table 1. Distribution of genotyped bulls by category of predictor and validation bulls.

Bull category	Birth year	No. bulls	
		Domestic[1]	Official[2]
Predictor	1952 to 1969	5	5
	1970 to 1979	22	22
	1980 to 1989	107	514
	1990 to 1994	149	408
	1995	201	686
	1996	162	670
	1997	181	845
	1998	65	340
	1999	162	470
	2000	43	167
	Total	***1097***	***4127***
Validation[3]	2000	114	
	2001	269	
	2002	76	
	2003	48	
	2004	17	
	Total	***524***	

[1]Distribution of predictor bulls when only bulls with Canadian domestic proofs were used in the prediction.
[2]Distribution of predictor bulls when all bulls with domestic and MACE official proofs in Canada were used in the prediction.
[3]Validation bulls: Bulls with Canadian domestic proofs in November 2008, but only a PA in February 2005.

Table 2. Distribution of genotyped bulls by category of predictor and validation bulls for traits not evaluated yet in 2004/2005.

Bull category	Birth year	No. bulls
Predictor	1952 to 1969	13
	1970 to 1979	41
	1980 to 1989	160
	1990 to 1994	175
	1995	211
	1996	161
	1997	184
	1998	66
	1999	153
	2000	15
	Total	**1179**
Validation[1]	2000	99
	2001	231
	2002	51
	2003	24
	Total	***405***

[1]Validation bulls: Bulls with Canadian domestic proofs in August 2008, but only a PA in August 2004.

Table 3. Squared correlations (r^2) of 2005 genomic predictions (GPA) using domestic proofs only (Domestic), and domestic and MACE proofs (Official)[1], and gain in r^2 over the r^2 of parent average (PA)

Trait	No. validation bulls	r^2 2005 PA	Domestic GPA_D	Official GPA_O	Gain GPA_O- PA	Gain[2] GPA_O-GPA_D
Milk	524	0.28	0.41	0.50	0.22	0.09
Fat	524	0.21	0.35	0.47	0.26	0.12
Protein	524	0.33	0.45	0.52	0.19	0.08
Fat %	524	0.23	0.42	0.59	0.36	0.18
Prot %	524	0.24	0.41	0.54	0.30	0.13
0verall - Production	*524*	*0.26*	*0.41*	*0.53*	*0.27*	*0.12*
Somatic cell score	499	0.24	0.29	0.39	0.15	0.10
Conformation	505	0.27	0.39	0.44	0.18	0.06
Dairy strength	514	0.17	0.28	0.32	0.14	0.04
Rump	502	0.25	0.31	0.34	0.09	0.03
Feet and legs	487	0.18	0.24	0.22	0.04	-0.02
Mammary system	504	0.19	0.33	0.43	0.24	0.10
Stature	524	0.33	0.53	0.58	0.25	0.04
Height at front end	505	0.16	0.34	0.34	0.18	-
Chest width	499	0.25	0.30	0.35	0.10	0.05
Body depth	510	0.25	0.32	0.39	0.14	0.07
Loin strength	504	0.22	0.34	0.34	0.12	-
Pin width	513	0.15	0.29	0.37	0.22	0.08
Pin setting	469	0.11	0.12	0.13	0.03	0.01
Rump angle	518	0.23	0.39	0.50	0.27	0.11
Bone quality	510	0.22	0.36	0.36	0.14	-
Foot angle	473	0.19	0.26	0.26	0.07	0.00
Heel depth	463	0.09	0.19	0.19	0.10	-
Set of rear legs	366	0.10	0.13	0.13	0.03	-
Rear legs- side view	502	0.18	0.30	0.35	0.17	0.05
Rear legs- rear view	478	0.22	0.20	0.21	0.00	0.01
Udder depth	521	0.16	0.33	0.43	0.27	0.10
Udder texture	482	0.07	0.25	0.25	0.18	-
Median suspensory	482	0.21	0.28	0.39	0.18	0.11
Fore attachment	505	0.16	0.32	0.39	0.23	0.07
Front teat placement	510	0.23	0.40	0.46	0.23	0.06
Teat length	507	0.19	0.30	0.41	0.22	0.10
Rear attachment height	502	0.25	0.32	0.36	0.11	0.04
Rear attachment width	492	0.13	0.24	0.24	0.11	-
Rear teat placement	508	0.11	0.21	0.21	0.10	-
0verall – Type	*502*	*0.21*	*0.31*	*0.37*	*0.16*	*0.06*

[1] The number of predictor bulls for production traits and SCS was 1097 and 4127 for Domestic and Official GPA, respectively. For type traits the same features were 1113 and 3966, respectively.

[2] For few type traits MACE evaluations were only available as parent averages and, therefore, were not used. In this case Official and Domestic GPA were the same and the difference in r^2 was not calculated. For consistency these traits were excluded from the overall means given in the table.

Table 4. Observed reliability of 2005 genomic predictions (GPA) using domestic proofs only (Domestic), and domestic and MACE proofs (Official)[1] and gain in reliability over the reliability of parent average (PA)

	No.	Reliability					
	validation	2005	Domestic	Official	Gain	Gain[2]	2008
Trait	bulls	PA	GPA_D	GPA_O	GPA_O-PA	GPA_O-GPA_D	DD[3]
Milk	524	0.39	0.53	0.63	0.24	0.09	0.90
Fat	524	0.39	0.55	0.68	0.29	0.13	0.90
Protein	524	0.39	0.51	0.60	0.21	0.09	0.90
Fat %	524	0.39	0.59	0.79	0.40	0.19	0.90
Prot %	524	0.39	0.57	0.72	0.33	0.15	0.90
0verall - Production	*524*	*0.39*	*0.55*	*0.68*	0.29	*0.13*	*0.90*
Somatic cell score	499	0.37	0.43	0.54	0.17	0.11	0.86
Conformation	505	0.37	0.51	0.58	0.21	0.07	0.85
Dairy strength	514	0.37	0.48	0.53	0.16	0.05	0.88
Rump	502	0.37	0.44	0.47	0.10	0.03	0.84
Feet and legs	487	0.37	0.45	0.43	0.06	-0.02	0.79
Mammary system	504	0.37	0.53	0.65	0.28	0.11	0.85
Stature	524	0.37	0.59	0.64	0.27	0.05	0.90
Height at front end	505	0.37	0.57	0.57	0.20	-	0.85
Chest width	499	0.37	0.43	0.49	0.12	0.06	0.84
Body depth	510	0.37	0.45	0.53	0.16	0.08	0.87
Loin strength	504	0.37	0.51	0.51	0.14	-	0.85
Pin width	513	0.37	0.52	0.62	0.25	0.10	0.88
Pin setting	469	0.37	0.39	0.41	0.04	0.02	0.73
Rump angle	518	0.37	0.55	0.67	0.30	0.12	0.88
Bone quality	510	0.37	0.53	0.53	0.16	-	0.87
Foot angle	473	0.37	0.46	0.46	0.09	0.00	0.76
Heel depth	463	0.37	0.52	0.52	0.15	-	0.71
Set of rear legs	366	0.38	0.43	0.43	0.05	-	0.66
Rear legs- side view	502	0.37	0.52	0.57	0.21	0.06	0.85
Rear legs- rear view	478	0.37	0.35	0.37	-0.01	0.02	0.77
Udder depth	521	0.37	0.56	0.67	0.30	0.11	0.89
Udder texture	482	0.37	0.60	0.60	0.23	-	0.79
Median suspensory	482	0.37	0.47	0.60	0.23	0.14	0.79
Fore attachment	505	0.37	0.55	0.64	0.27	0.09	0.86
Front teat placement	510	0.37	0.56	0.63	0.26	0.07	0.87
Teat length	507	0.37	0.50	0.62	0.25	0.12	0.86
Rear attachment height	502	0.37	0.45	0.50	0.13	0.04	0.84
Rear attachment width	492	0.37	0.50	0.50	0.13	-	0.83
Rear teat placement	508	0.37	0.49	0.49	0.12	-	0.86
0verall - Type	*502*	*0.37*	*0.49*	*0.55*	*0.18*	*0.07*	*0.84*

[1] The number of predictor bulls for production traits and SCS was 1097 and 4127 for Domestic and Official GPA, respectively. For type traits the same features were 1113 and 3966, respectively.

[2] For few type traits MACE evaluations were only available as parent averages and, therefore, were not used. In this case Official and Domestic GPA were the same and the difference in r^2 was not calculated. For consistency these traits were excluded from the overall means given in the table.

[3] Average reliability of the daughter deviations of the validation bulls in November 2008.

Table 5. Squared correlations (r^2) of 2004 genomic predictions (GPA) using domestic proofs (Domestic)[1] and gain in r^2 over the r^2 of parent average (PA).

Trait	No. validation bulls	r^2 2004 PA	Domestic GPA_D	Gain GPA_D-PA
Daughter fertility	379	0.09	0.12	0.03
Age at 1st service	382	0.04	0.08	0.04
Days open	370	0.11	0.16	0.05
Calving to 1st service	342	0.14	0.14	0.00
Non return rate-cow	379	0.08	0.10	0.02
No. services-cow	392	0.10	0.14	0.04
1st serv. to concept.-cow	374	0.10	0.14	0.04
Gestation length-cow	380	0.17	0.21	0.04
Maternal calving ease (CE)-heifer	405	0.12	0.18	0.06
Maternal calving ease (CE)-cow	371	0.06	0.09	0.03
Overall	*377*	*0.10*	*0.14*	*0.04*

[1] The number of predictor bulls ranged from 1070 to 1179 depending on the trait.

Table 6. Observed reliability of 2004 genomic predictions (GPA) using domestic proofs (Domestic)[1] and gain in reliability over the reliability of parent average (PA).

Trait	No. validation bulls	Reliability 2004 PA	Domestic GPA_D	Gain GPA_D-PA	2008 DD[2]
Daughter fertility	379	0.29	0.34	0.05	0.72
Age at 1st service	382	0.30	0.35	0.05	0.78
Days open	370	0.29	0.36	0.07	0.77
Calving to 1st service	342	0.29	0.29	0.00	0.77
Non return rate-cow	379	0.29	0.32	0.03	0.72
No. services-cow	392	0.29	0.34	0.05	0.77
1st serv. to concept.-cow	374	0.29	0.34	0.05	0.77
Gestation length-cow	380	0.29	0.34	0.05	0.83
Maternal calving ease (CE)-heifer	405	0.30	0.37	0.07	0.78
Maternal calving ease (CE)-cow	371	0.28	0.32	0.04	0.75
Overall	*377*	*0.29*	*0.34*	*0.05*	*0.77*

[1] The number of predictor bulls ranged from 1070 to 1179 depending on the trait.
[2] Average reliability of the daughter deviations of the validation bulls in August 2008.