Reliability Increases from Combining 50,000- and 777,000-Marker Genotypes from Four Countries

P.M. VanRaden[1], K.M. Olson[2], D.J. Null[1], M. Sargolzaei[3], M. Winters[4] and J.B.C.H.M. van Kaam[5]
[1]*Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD, USA*
[2]*National Association of Animal Breeders, Columbia, MO, USA*
[3]*Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON, Canada*
[4]*DairyCo Agriculture and Horticulture Development Board, Kenilworth, Warwickshire, UK,*
[5]*Associazone Nazionale Allevatori Frisona Italiana, Cremona, Italy*

Abstract

Genomic predictions were compared on U.S. scale after combining 50,000 (50K) and 777,000 (HD) marker genotypes across countries. The genotyped Holsteins included 161,341 animals with five marker densities including 1,510 with HD. Imputation was more accurate with FImpute than with findhap across the five densities. Reliability of HD predictions minus reliability of 50K predictions averaged only 0.4%. For 50K predictions, reliability for the combined reference population was higher than for domestic only, with the gain averaging 2.6% for Holsteins and 3.2% for Brown Swiss when 50K genotypes were added for 3,593 foreign Holstein and 732 Brown Swiss bulls that had no U.S. daughters prior to August 2008. Gains should be larger with 7,974 foreign Holstein and 870 Brown Swiss bulls in the May 2012 reference population. Multi-trait methods were not more accurate than single-trait for Holsteins, but multi-trait reliability averaged 1.4% higher than single-trait for Brown Swiss, perhaps because of lower genetic correlations between Brown Swiss populations. International exchange improves accuracy and decreases cost of obtaining additional phenotypes and genotypes.

Key words: genomic evaluation, international exchange, marker density, imputation

Introduction

International partners are sharing costs of creating the large data sets of genotypes and phenotypes needed to make selection accurate for young animals (Cromie *et al.*, 2011; Lund *et al.*, 2011; Zumbach *et al.*, 2011). Nearly all reports, including initial studies from Canada (Schenkel *et al.*, 2009) and the United States (Olson *et al.*, 2011), concluded that predictions improve when 50K genotypes of foreign bulls are added to reference populations.

High density genotypes have provided small gains in reliability compared with 50K (Su *et al.*, 2012), no gains (Harris and Johnson, 2011), or gains for only some traits (Solberg *et al.*, 2012). The current report compares 50K and HD predictions computed after exchanging Holstein genotypes across four countries, and also compares Brown Swiss predictions with and without 50K genotypes of foreign bulls.

Methods

Genotypes from the United States (USA), Canada (CAN), Italy (ITA), and the United Kingdom (GBR) were combined for Holstein reference bulls. The May 2012 USA reference population currently includes 10,534 USA, 3,040 CAN, 2,902 ITA, 571 GBR, and 1,461 other progeny-tested bulls totaling 18,508, plus 22,800 USA cows with lactation records. For Brown Swiss, genotypes from Switzerland (CHE), Germany (DEU), and Austria (AUT) were exchanged to obtain a May 2012 USA reference population of 1,682 bulls (812 USA, 403 CHE, 318 DEU, 52 AUT, and 97 other), plus 374 USA cows.

Addition of 50K (Illumina BovineSNP50) genotypes from foreign bulls was tested for observed reliability on the USA scale using both single-trait (VanRaden *et al.*, 2009) and multi-trait (VanRaden and Sullivan, 2010)

prediction methods. In the multi-trait model, bulls with ≥10 daughters in the USA evaluation were treated as domestic even if first progeny tested in another country. Information from all other bulls was modeled as one foreign trait. Assumed correlations between domestic and foreign performance were weighted means of genetic correlations of USA with the other countries as estimated by Interbull. Predictions based on 50K genotypes were compared with predictions based on HD genotypes from the same combined population.

Validation studies were conducted to evaluate the effect of including foreign bulls in the reference population on USA genomic evaluations. The truncated August 2008 Holstein reference population included 3,593 foreign bulls, 7,125 domestic bulls, and 5,124 domestic cows with records. The validation data set consisted of 3,404 Holstein bulls with USA identification numbers that had no daughter records in August 2008 but had USA daughter records by April 2011. Corresponding Brown Swiss numbers were 732 foreign and 741 domestic reference bulls, 171 reference cows, and 115 validation bulls.

Illumina BovineHD (HD) genotypes were obtained for: 460 ITA, 305 USA, 284 GBR, 93 CAN, and 196 other bulls; 138 cows from the USDA Beltsville Agricultural Research Center research herd; and 34 other cows, totaling 1,510 Holsteins. Two preliminary studies used 342 and 1,074 HD genotypes to test effects of additional animals on imputation and reliability. The final analysis included only 311,725 of the 777,962 HD markers because about half of those were redundant (Harris *et al.*, 2012).

Lower density genotypes also were included in the data sets. As of March 2012, Holstein genotypes were available for 82,358 animals with 50K genotypes, 1,797 animals with 8K (GeneSeek Genomic Profiler), 20,177 animals with 6K (Illumina BovineLD), 52,270 with 3K (Illumina Bovine3K), and 3,229 0K (nongenotyped) dams imputed from ≥4 genotyped progeny. Total numbers genotyped were 161,341 Holstein and 3,096 Brown Swiss.

Programs FImpute (Sargolzaei *et al.*, 2011) and findhap (VanRaden *et al.*, 2011) were tested for imputing across a range of densities for a simulated chromosome with 1,112 HD genotypes. Genomic evaluations were computed using both a linear model with normal prior and a nonlinear model with heavy-tailed prior for marker effects and compared with 50K evaluation. VanRaden *et al.* (2012) report further details on the HD evaluation.

Results and Discussion

Single-trait predictions from the combined 50K reference populations truncated in 2008

Table 1. Gains (percentage points) in 50K reliability over parent average reliability with foreign genotypes excluded or included using single-trait (ST) or multi-trait (MT) prediction.

Trait	Holstein			Brown Swiss		
	ST, foreign excluded	ST, foreign included	MT, foreign included	ST, foreign excluded	ST, foreign included	MT, foreign included
Milk yield	26.6	28.7	28.6	15.6	20.9	20.6
Fat yield	29.0	31.3	31.1	10.3	11.4	12.5
Protein yield	20.2	22.5	22.3	13.5	16.2	16.6
Productive life	19.5	21.8	22.2	4.9	6.5	7.5
Somatic cell score	23.7	27.2	27.2	5.0	4.2	6.6
Daughter pregnancy rate	17.5	21.3	21.2	9.8	6.8	7.3
Final score	22.2	24.0	24.1	2.7	9.3	15.7
Stature	29.4	34.0	33.9	12.8	21.5	24.2
Sire calving ease	8.8	10.4	9.7	1.9	8.7	6.6
Mean (9 traits)	**21.9**	**24.6**	**24.5**	**8.5**	**11.7**	**13.1**

averaged a 24.6% gain in reliability above parent average for Holstein and 11.7% for Brown Swiss as compared to 21.9% using only Holstein bulls with domestic daughters and 8.5% for Brown Swiss bulls with domestic daughters (Table 1). Increases from combining 50K data using single-trait methods should be larger now than for 2008 truncated data because the number of bulls with only foreign daughters increased from 3,593 to 7,974 for Holsteins and from 732 to 870 for Brown Swiss.

Multi-trait evaluations were not more reliable than single-trait evaluations for Holsteins, but multi-trait reliability averaged 1.4% higher than single-trait for Brown Swiss (Table 1). Reliability differences were very stable across traits for Holsteins but varied much more for Brown Swiss because of fewer validation bulls. Treating traits for foreign bulls as being different measurements than for domestic bulls may be important for Brown Swiss because genetic correlations were lower and genetic backgrounds may differ more across countries; treating Holstein traits from all four countries as the same trait gave good results. In general, countries and traits with high genetic correlations should be treated as a single population, whereas more diverse populations with differing trait definitions or differing genetic backgrounds may benefit from implementing multi-trait methods.

Imputation was more accurate from 50K to HD than from 3K or 6K to HD. After imputing missing markers with FImpute or findhap, respective percentages of correct genotypes were 99.96 and 99.89 from HD, 99.3 and 99.0 from 50K, 94.7 and 94.6 from 6K, and 91.1 and 90.5 from 3K genotypes. Accuracy for 3K and 6K genotypes further improved about 2% with FImpute and 1% with findhap if imputed first to 50K and then to HD instead of imputing all genotypes together. The reference population includes imputed 3K and 6K genotypes from cows, but nearly all of the validation bulls had 50K genotypes.

Mean HD reliability from the combined Holstein reference population truncated in 2008 was 61.1% with a nonlinear model and 60.3% with a linear model (Table 2); mean 50K reliability was 60.7% with a nonlinear

Table 2. Reliability of HD linear, HD nonlinear, and 50K nonlinear predictions including foreign Holstein bulls with single-trait methods.

Trait	HD linear	HD non-linear	50K non-linear
Milk	64.0	65.2	65.5
Fat	66.9	68.7	68.5
Protein	59.8	60.1	60.6
Productive life	77.3	78.6	77.0
Somatic cell score	65.3	65.7	65.2
Dtr pregnancy rate	66.9	67.3	66.6
Final score	54.9	55.0	55.2
Stature	65.5	66.2	64.4
Sire calving ease	29.6	31.5	33.8
Mean (28 traits)	**60.3**	**61.1**	**60.7**

model. Thus, including 6 times more markers resulted in an increase of only 0.4% reliability, a little less than the 0.9% expected from simulation (VanRaden *et al.*, 2011).

Preliminary imputation tests indicated that too few HD animals can decrease reliability but that additional HD genotypes may not lead to further improvements. Mean reliability for preliminary HD evaluation was 0.5% lower than 50K reliability with only 342 HD genotypes and 0.5% higher with 1,074 HD genotypes. The final evaluation with 1,510 HD genotypes had mean reliability only 0.4% higher than 50K reliability and 0.1% less than the 1,074 HD test. These results indicate that imputation from 50K to HD is already sufficiently accurate. However, further improvements in imputation from lower densities to HD would be useful if HD is implemented.

Future research may identify the most useful subsets of HD markers to include on future chips of less density. Jersey breeders could benefit from genotype exchange, but lower correlations between northern and southern hemisphere traits could require using a multi-trait evaluation. Additional HD genotypes for Ayrshire, Brown Swiss, and other breeds could improve benefits from multi-breed evaluation, but implementation could be difficult if genomic databases are isolated or if traditional evaluations are published only within breeds. A more pressing

need may be to account for the effects of genomic preselection on the traditional evaluations.

Conclusions

International exchange continues to improve reliability of evaluations and reduce cost of obtaining additional genotypes. For 50K evaluations, reliability for young Holstein bulls improved on average by 2.6% when foreign reference bulls were added to USA evaluations. That increase underestimates the true value of genotype exchange because many of the high-reliability foreign bulls were treated as domestic if they had USA daughters. Benefits for Brown Swiss were slightly greater at 3.2% because of fewer domestic bulls, but less than in the combined InterGenomic reference population (Zumbach et al., 2011). Multi-trait and single-trait predictions were equally accurate for Holsteins, but multi-trait prediction resulted in an additional increase of 1.4% in mean reliability for Brown Swiss when foreign evaluations were treated as correlated rather than the same trait.

High density genotypes provided only a 0.4% average increase in reliability of Holstein evaluations across traits. A nonlinear model with heavy-tailed prior for marker effects gave only a 0.8% increase over the linear model. Imputation from 50K to HD was very accurate, but imputation from lower densities to HD was more difficult. Increasing the number of HD genotypes from 342 to 1,074 improved reliability of genomic predictions, but a further increase to 1,510 did not further improve reliability. Added costs of obtaining more reference bulls and more HD genotypes were reduced by sharing genotypes among several cooperating partners.

Acknowledgements

References

Cromie, A.R., Berry, D.P., Wickham, B., Kearney, J.F., Pena, J., van Kaam, J.B.C.H., Gengler, N., Szyda, J., Schnyder, U., Coffey, M., Moster, B., Hagiya, K., Weller, J.I., Abernethy, D. & Spelman, R. 2011. International genomic co-operation: Who, what, when, where, why and how? *Interbull Bulletin 41*, 72–78.

Harris, B.L., Creagh, F.E., Winkelman, A.M. & Johnson, D.L. 2012. Experiences with the Illumina high density bovine beadchip. *Interbull Bulletin 44*, 3-7.

Harris, B.L. & Johnson, D.L. 2011. The impact of high density SNP chips on genomic evaluation in dairy cattle. *Interbull Bulletin 42*, 40–43.

Lund, M.S., de Roos, A.P.W., de Vries, A.G., Druet, T., Ducrocq, V., Fritz, S., Guillaume, F., Guldbrandtsen, B., Liu, Z., Reents, R., Schrooten, C., Seefried, F. & Su, G. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet. Sel. Evol. 43*, 43.

Olson, K.M., VanRaden, P.M. & Null, D.J. 2011. Impacts of inclusion of foreign data in genomic evaluation of dairy cattle. 2011. *J. Dairy Sci. 94 (E-Suppl. 1)*, 164–165 (abstr. 35).

Sargolzaei, M., Chesnais, J.P. & Schenkel, F.S. 2011. FImpute–An efficient imputation algorithm for dairy cattle populations. *J. Dairy Sci. 94(E-Suppl. 1)*, 421(abstr. 333).

Schenkel, F.S., Sargolzaei, M., Kistemaker, G., Jansen, G.B., Sullivan, P., 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.

Solberg, T.R., Heringstad, B., Svendsen, M., Grove, H. & Meuwissen, T.H.E. 2012. Genomic predictions for production- and functional traits in Norwegian Red from BLUP analyses of imputed 54K and 777K SNP data. *Interbull Bulletin 44,* 240-243.

Su, G., Brøndum, R.F., Ma, P., Guldbrandtsen, B., Aamand, G.P. & Lund, M.S. 2012. Genomic prediction using high-density SNP markers in Nordic Holstein and Red. *Interbull Bulletin 44,* 157-161.

VanRaden, P.M., Null, D.J., Sargolzaei, M., Wiggans, G.R., Tooker, M.E., Cole, J.B., Sonstegard, T.S., Connor, E.E., Winters, M., van Kaam, J.B.C.H.M., Van Doormaal, B.J., Faust, M.A. & Doak, G.A. 2012. Genomic imputation and evaluation using high density Holstein genotypes. *J. Dairy Sci.* (submitted).

VanRaden, P.M., O'Connell, J.R., Wiggans, G.R. & Weigel, K.A. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol. 43,* 10.

VanRaden, P.M. & Sullivan, P. 2010. International genomic evaluation methods for dairy cattle. *Genet. Sel. Evol. 42,* 7.

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

Zumbach, B., Jorjani, H. & Dürr, J. 2011. Brown Swiss genomic evaluation. *Interbull Bulletin 42,* 44–51.