Genomic evaluations in the United States and Canada: A collaboration

G.R. Wiggans[1], T.S. Sonstegard[1], P.M. VanRaden[1], L.K. Matukumalli[1,2], R.D. Schnabel[3], J.F. Taylor[3], J.P. Chesnais[4], F.S. Schenkel[5] & C.P. Van Tassell[1]

[1]*Agricultural Research Service, USDA, Beltsville, MD, USA*
[2]*Bioinformatics and Computational Biology, George Mason University, Manassas, VA, USA*
[3]*Division of Animal Sciences, University of Missouri, Columbia, MO, USA*
[4]*Semex Alliance, Guelph, ON, Canada*
[5]*University of Guelph, Guelph, ON, Canada*

Summary

The United States and Canada have collaborated in developing genomic evaluations that utilize genotypes obtained from high-density single-nucleotide polymorphism (SNP) chips. The DNA for genotyping was obtained from semen from 7 major artificial-insemination (AI) companies in North America through the Cooperative Dairy DNA Repository (CDDR) and from other sources. In collaboration with Illumina, Inc., the BovineSNP50 BeadChip was developed to genotype >50 000 SNP in a single assay. To determine the accuracy of genomic prediction, evaluations calculated in 2003 for bulls with evaluations (born mostly before 1999) were used to predict April 2008 evaluations for bulls without evaluations in 2003 (mostly born in 2001 and 2002). The accuracy of genomic evaluations is closely tied to the number of animals in the predictor group. For Holsteins, 7 AI organizations have had several thousand young bulls genotyped for initial application of genomic evaluation. Over 12 500 Holstein, 1 250 Jersey and 350 Brown Swiss genotypes have been processed to check for conflicts between parent and progeny DNA and to impute some of the missing genotypes. The SNP effects were estimated using current evaluations for over 8 700 genotyped bulls and cows that were based on lactation records from >10 million cows.

Genomic evaluations of the nominated animals were calculated by combining SNP effects with parent averages for yield, functional, calving and type traits as well as the net merit economic index. For Holsteins, genomic information provided a substantial improvement in accuracy over parent average for all traits, and the amount of improvement increased as the number of bulls in the predictor group increased. Jersey and Brown Swiss results indicated that more predictor animals were necessary for adequate accuracy, and additional genotypes are being obtained. The first release of genomic evaluations was in April 2008 for Holsteins. Estimated SNP effects are updated 3 times each year with the national genetic evaluations. Additional updates provide evaluations for newly genotyped animals based on previously estimated SNP effects. In 2009, genomic evaluations are expected to become official in the United States and Canada and will also affect evaluations of

descendants that have not been genotyped. Genomic evaluations of bulls currently are released only to the owner and AI organization until the bulls are enrolled with the National Association of Animal Breeders or a Canadian AI organization or they reach 2 years of age. Genomic evaluations will be routinely available for cows later in 2008.

Methods to exchange genomically enhanced evaluations across countries need to be developed. Genomic evaluations, which can be obtained shortly after birth, are expected to revolutionize dairy cattle breeding programs by changing how bulls are selected for and culled from AI service and how cows are chosen to produce sons. Some bulls will be marketed at 2 years of age based solely on genomic evaluations, which creates a potential for a large increase in rate of genetic gain because of the much shorter generation interval possible.

***Key words**: Genomic evaluation, Single nucleotide polymorphism, Dairy cattle, Genomics*

Introduction

In 2007, researchers in quantitative and molecular genetics in the United States and Canada began to collaborate on the development and implementation of genomic evaluations and the integration of those evaluations into national genetic evaluations for dairy cattle. This collaboration is an outgrowth of the establishment of CDDR in 1999 (Ashwell and Van Tassell, 1999), which has been collecting semen from North American AI organizations for almost 10 years. Five U.S.-based and 2 Canadian-based AI organizations currently contribute to CDDR. The repository now contains semen from >17 400 bulls and is a primary source of DNA for investigating genomic evaluation of dairy cattle.

In collaboration with Illumina, Inc. (San Diego, CA), the BovineSNP50 BeadChip was developed (Van Tassell *et al.*, 2008). That effort involved selection and discovery of high-quality SNP that were uniformly distributed over the genome and that had high minor allele frequency. Over 50 000 SNP were incorporated into the chip.

For the collaborative project, DNA was extracted from semen from the CDDR as well as from independently obtained semen. Genotyping was performed primarily by the Bovine Functional Genomics Laboratory, U.S. Department of Agriculture (USDA; Beltsville, MD), the University of Missouri (Columbia, MO) and the University of Alberta (Edmonton, AB). Some genotypes were provided by GeneSeek (Lincoln, NE), the Genetics & IVF Institute (Fairfax, VA) and Illumina, Inc. The project was funded by multiple research grants from USDA as well as contributions by U.S. AI organizations through the National Association of Animal Breeders (Columbia, MO) and the Semex Alliance (Guelph, ON) and by Holstein Association USA (Brattleboro, VT) and American Jersey Cattle Association (Reynoldsburg, OH).

Data editing included removing SNP that were monomorphic or had a minor allele frequency of <5%. For routine analysis, this limit was reduced to 2%. Additionally, each SNP was compared with all others to eliminate those that were redundant because of complete linkage disequilibrium ($r^2 = 1$). Over 38 000 SNP remained for prediction of genomic evaluations (Wiggans *et al.*, 2008). To determine the accuracy of genomic prediction, evaluations calculated in 2003 were used to predict April 2008 evaluations for bulls without evaluations in 2003. That restriction generally meant that bulls born before 1999 were used to predict bulls born in 2001 and 2002 (VanRaden *et al.*, 2008b).

Field test

The prediction accuracy reported by Van Raden *et al*. (2008b) was sufficient to proceed with a field test, which began in April 2008. The first release in April was followed by releases in July and August; >5 300 animals nominated by 7 AI organizations were included. The nominated animals were primarily bull calves, and their genomic evaluations were used to select which full brother to purchase from an embryo flush. Some bulls awaiting the results of progeny testing were genotyped to determine if they should be culled. Additionally, some bulls with current evaluations were included to improve prediction accuracy. Cows were also included to assess their suitability as bull dams.

The AI organizations mainly arranged to have blood samples collected and sent to GeneSeek, Genetics & IVF Institute, or the University of Alberta for DNA extraction and genotyping. In some cases, semen, hair or previously extracted DNA was provided. Some genotyping of nominated animals also was done at the Bovine Functional Genomics Laboratory in fulfillment of a research commitment. Holstein Association USA assisted by offering a low-cost American ID plan, which enabled the AI organizations to enroll all animals in a national ID program. The variability of ID reporting required a considerable reconciliation effort. Work continues on streamlining the processing of sample ID, which now includes bar coding of samples and standardization of ID.

Accuracy of sample ID was determined by checking genotypes for inconsistencies between parent and progeny when the parent had been genotyped. If the genotype for a SNP had been determined for both animals, the number of times that each animal was homozygous for different alleles was counted. The low rate of genotyping errors for the assay meant that usually <5 inconsistencies were found of usually >15,000 SNP where both animals were homozygous when the parent-progeny assignment was correct. However, several thousand inconsistencies usually were found when the parent-progeny assignment was incorrect, which enabled easy detection of sample ID mix-ups or pedigree inconsistencies.

The genomic relationship matrix (VanRaden, 2007) was calculated as part of the genomic evaluation. Comparison of the genomic and pedigree relationships was also used to detect pedigree errors, particularly for animals with an incorrectly identified maternal grandsire. To estimate gene frequencies in the base population, allele frequency estimates were generated for all genotyped animals and their ancestors. The oldest genotyped bull was born in 1952. When the genotype for a particular SNP of a genotyped animal was missing, the estimated frequency (VanRaden *et al*., 2008a) was used to set the genotype to 0, 1 or 2 (number of the counted allele present) if the frequency was within 10 percentage units of 0, 50 or 100%, respectively.

Genomic evaluations were calculated with the system of VanRaden *et al.* (2008b) based on genotypes for 7,195 animals, including 1,263 cows. Twenty-seven traits were evaluated: 5 yield (milk, fat, and protein volumes and fat and protein percentages), 3 functional (somatic cell score, productive life, and daughter pregnancy rate), 2 calving (calving ease and stillbirth), 16 type (final score and 15 linear traits: stature, strength, body depth, dairy form, rump angle, rump width, rear legs–side view, rear legs–rear view, foot angle, fore udder attachment, rear udder height, udder cleft, udder depth, front teat placement and teat length), and net merit (a lifetime genetic-economic index). Estimates of SNP effects were combined with parent average or genetic evaluation to produce a genomic evaluation. That critical step required determining the contribution of genomic data beyond

traditional evaluations. The genomic contribution was calculated by comparing the results from an animal model evaluation for only genotyped animals with their genomic evaluation.

The addition of animals to the predictor group increased evaluation reliability substantially compared with earlier evaluations for the group being predicted. For genomic evaluations, mean expected reliability of 689 nominated young bulls ranged from 63 to 75% across traits (VanRaden *et al.*, 2008c). Those reliabilities are theoretical and may exceed what is actually realized. Because all genotyped animals with evaluations were used to estimate SNP effects, realized reliability could not be determined. Genomic evaluations of the nominated animals were provided to nominating organizations and owners in the United States, and the evaluations of the predictor bulls were made available on a web site as preliminary research results.

Routine genomic evaluation

Based on field trial results, USDA is implementing routine genomic evaluations. Estimates of SNP effects will be updated 3 times per year with each national evaluation. Newly genotyped animals will be evaluated once or more between those updates based on the volume of genotype data received. Over 12 500 Holstein, 1 250 Jersey and 350 Brown Swiss genotypes currently have been processed to check for conflicts between parent and progeny DNA and to impute some of the missing genotypes.

Animals to be genotyped will be assigned national ID and have their pedigree recorded before their DNA samples are processed to avoid sample ID and logistic issues. Commercial laboratories will perform the genotyping and report genotypes for a set of SNP that can be reliably scored. A system for confirming sample ID through checks for parent-progeny SNP genotype inconsistencies will be developed. The checks will be facilitated by a freely available database with genotypes for a subset of SNP for all animals that have been genotyped. Breed associations are expected to work on resolving conflicts as they do now with parentage testing.

Animal gender for each sample will also be confirmed by checks for heterozygous SNP not in the pseudoautosomal region of the X chromosome. In most cases, females have several hundred heterozygous SNP and males almost none. However, a female can have a very low number of heterozygous SNP if she inherited both X chromosomes from the same ancestor.

Under agreements between the requester and the genetic evaluation centers, the laboratories will send BovineSNP50 genotypes to the evaluation centers. The U.S. and Canadian centers will share genotypes and computer software. Requesters can be AI organizations, breed associations or individual owners. In recognition of their contribution, the AI organizations that provided semen and financial support to the project have exclusive rights to have genomic evaluations calculated for bulls until March 2013.

The U.S. and Canadian evaluation centers plan to exchange evaluations so that each country has evaluations for all genotyped animals with evaluations and their ancestors. Exchange of evaluations and generation of conversion equations before release are necessary to include the latest evaluations for foreign animals. The International Bull Evaluation Service (INTERBULL, Uppsala, Sweden) calculates evaluations only for bulls, and those evaluations are not available soon enough to provide needed information.

In 2009, USDA expects to make genomic evaluations the official evaluation for all genotyped animals and to allow those evaluations to affect evaluations of descendants that have not been genotyped (Gengler and VanRaden, 2008). As the official evaluation, genomic evaluations will appear in pedigrees from breed associations. Because AI organizations wish to protect their investment in genomic evaluations for bull calves, those evaluations are not official until the bull is enrolled with the National Association of Animal Breeders or a Canadian AI organization or reaches 2 years of age.

INTERBULL

INTERBULL plays an important role in international semen sales by enabling comparison of evaluated bulls across countries. If the genotype is available, the national center can calculate a local genomic evaluation for a bull with no local progeny. Semen exporters will want importing countries to use genotypes of their bulls when they provide local evaluations. Availability of genomic evaluations may help to overcome the evaluation reduction that occurs for top bulls because of correlations of <1 between daughter performance in individual countries. Genomic evaluations will be most accurate if populations in the exporting and importing countries are similar genetically. If those populations are quite different, the SNP effects estimated in the importing country may be less accurate in predicting the daughter performance for foreign bulls than for domestic bulls. For countries that do not calculate genomic evaluations, INTERBULL is the only source of evaluations for bulls from other countries that are expressed on their domestic scale. To serve those countries fully, INTERBULL will need to provide evaluations for bulls that are marketed solely on their genomic evaluations, which would necessitate that INTERBULL remove the current requirement that a bull have daughters in 10 herds.

Another concern is how to exchange and combine genomically enhanced evaluations across countries (van der Beek, 2007). If only 1 country provides such an evaluation, then current procedures used by INTERBULL could provide an appropriate result if accuracy of the genomic contribution is reflected in the measure of evaluation accuracy. New procedures are required to avoid duplicate information when 2 countries or more submit genomically enhanced evaluations. Genomic information from different countries is not independent and should only be included once, not for each country that contributes such an evaluation. Enough information will need to be provided to allow INTERBULL to apply the principles already used to avoid duplicating pedigree information. Genomic information may be more complicated to accommodate than pedigree information because different SNP may be genotyped across countries.

A system to share genotypes across countries would be most efficient but will take some time to develop. Genotypes are expensive to obtain but provide a wealth of information. They may have uses not yet envisioned and may affect the competitiveness and role of various organizations. Careful consideration should be given to how to share genotypes, particularly internationally.

Collaboration

Implementation of genomic evaluations illustrates the synergism that can be achieved when public, private and academic institutions cooperatively pursue the same objective. One of the most important factors affecting accuracy of genomic evaluations is the number of predictor bulls used. For Holsteins, genomic information provided a substantial improvement in accuracy over parent average for all traits,

and the amount of improvement increased as the number of bulls in the predictor group increased. Jersey and Brown Swiss results indicated that more predictor animals are necessary for adequate accuracy, and additional genotypes are being obtained. Full participation by the North American AI industry through the CDDR is providing the large families needed to study the small effects of individual genes.

A joint U.S.-Canadian project also helps with industry acceptance of results and promotes understanding of a profound change in the way that bulls will be selected and semen marketed. Keeping the characteristics of the fundamental evaluation system outside the realm of marketing will help the process remain unbiased and gain wider acceptance across the entire industry. The industry appreciates assurance that the system is appropriate by having both U.S. and Canadian research teams involved.

Impact

Genomic evaluations are expected to revolutionize dairy cattle breeding programs because they can be obtained shortly after birth. They will change selection and culling of bulls for AI service as well as choice of cows to produce sons. Some bulls will be marketed at 2 years of age based solely on genomic evaluations. The resulting shorter generation interval will create the potential for a large increase in rate of genetic gain.

Acknowledgements

The authors thank Drs. Bevin Harris, David Johnson and Anne Winkelman of Livestock Improvement Corporation (Hamilton, New Zealand) for their visit and sharing their experiences, which confirmed many results and gave ideas for improvements. This project was supported by National Research Initiative Grant Nos. 2006-35205-16888 and 2006-35205-16701 from the USDA Cooperative State Research, Education and Extension Service and by the National Association of Animal Breeders Holstein Association USA (Brattleboro, VT) and American Jersey Cattle Association (Reynoldsburg, OH).

List of references

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

Gengler, N. & P.M. VanRaden 2008. Strategies to incorporate genomic prediction into population-wide genetic evaluations (abstract). J. Dairy Sci. 91(E-Suppl. 1): 506.

van der Beek, S. 2007. Effect of genomic selection on national and international genetic evaluations. INTERBULL Bull. 37: 115–118.

VanRaden, P.M. 2007. Genomic measures of relationship and inbreeding. INTERBULL Bull. 37: 33–36.

VanRaden, P.M., M.E. Tooker & N. Gengler. 2008a. Effects of allele frequency estimation on genomic predictions and inbreeding coefficients (abstract). J. Dairy Sci. 91(E-Suppl. 1): 506.

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel & F. Schenkel. 2008b. Reliability of genomic predictions for North American dairy bulls (abstract). J. Dairy Sci. 91(E-Suppl. 1): 305.

VanRaden, P., G. Wiggans, C. Van Tassell, T. Sonstegard & L. Walton. 2008c. Genomic prediction. Changes to evaluation system (April 2008), <*http://aipl.arsusda.gov/ reference/changes/eval0804.html*>.

Van Tassell C.P., T.P.L. Smith, L.K. Matukumalli, J.F. Taylor, R.D. Schnabel, C.T. Lawley, C.D. Haudenschild, S.S. Moore, W.C. Warren & T.S. Sonstegard. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods 5:247–252.

Wiggans, G.R., T.S. Sonstegard, P.M. VanRaden, L.K. Matukumalli, R.D. Schnabel, J.F. Taylor, F.S. Schenkel and C.P. & Van Tassell. 2008. Selection of single nucleotide polymorphisms and genotype quality for genomic prediction of genetic merit in dairy cattle (abstract). J. Dairy Sci. 91(E-Suppl. 1): 507.