

Benefits from Cooperation in Genomics

P.M. VanRaden¹, G.R. Wiggins¹, C.P. Van Tassell², T.S. Sonstegard² and F. Schenkel³

¹Animal Improvement Programs and ²Bovine Functional Genomics Laboratories, USDA Agricultural Research Service, Beltsville, MD, and ³University of Guelph, ON, Canada.

Abstract

Accurate estimation of small genetic effects requires very large data sets. Therefore, benefits from combining data across countries are much larger for genotypes than for phenotypes. Methods of data exchange are compared. Gains in reliability from North American data are much larger using 5,369 Holsteins than 1,361 Jerseys or 228 Brown Swiss. Reliability for young animals increased to >80% using genotypes of 40,000 proven bulls in simulation. Use of 100,000 instead of 50,000 markers improved reliability only 2%. Computer programs may apply very widely because marker inheritance is the same across many species.

Cooperation

Global cooperation among animal breeders increases genetic progress because animals with superior genetic merit in one population can quickly contribute to all populations. Genetic evaluations have been computed primarily within countries, and results exchanged through Interbull. Direct evaluation of global phenotype files could be possible, but accuracy increases only a little as compared to combining the results of national evaluations (Fikse, 2004; Maltecca *et al.*, 2004). Advantages of global phenotypic evaluations would be more uniform methods, ability to model genetic-environmental interactions by management or climate rather than country, and potential large reductions in labor and duplication. National evaluations have been preferred because local researchers know more about collecting data, modeling population-specific environmental and genetic effects, explaining results in local languages, and may respond more quickly to changes requested by users.

Gains in reliability are much greater when combining genotype than phenotype files. This economy of scale makes genomic exchange across countries much more rewarding than phenotypic exchange. Researchers in genomic evaluation should try to think outside the borders.

Genomic Data

DNA sequencing and genotyping techniques have quickly become affordable due primarily to advances with human DNA. The human genome was sequenced in 2000, followed by cattle and chicken sequencing in 2004, and pig sequencing should be completed in 2009. The cattle genome cost \$53 million to sequence, with funds provided by organizations in the United States, Canada, Australia, and New Zealand. The U.S. National Human Genome Research Institute contributed about half the funds because they expect the research with cattle to help in understanding human health.

Data sharing is required by funding organizations for most research projects with human DNA, and data must be shared even before the research laboratory that generated the data has any articles written or accepted. "The highest priority of the International Human Genome Sequencing Consortium is ensuring that sequencing data from the human genome is available to the world's scientists rapidly, freely and without restriction." (National Human Genome Research Institute, 2008). "The principle of rapid pre-publication release should apply to other types of data from other large-scale production centers" (Wellcome Trust, 2003). Researchers and breeders currently do not have free access to cattle genotypes because genotyping is often done by for-profit companies rather than research grants.

Dairy cattle breeders across North America, Europe, and Oceania have rapidly adopted Illumina's BovineSNP50 BeadChip (Matukumalli *et al.*, 2008), developed in cooperation with researchers from USDA (Beltsville, MD, and Clay Center, NE), the University of Missouri (Columbia, MO), and the University of Alberta (Edmonton, Canada). This SNP chip is available worldwide to genotype both dairy and beef cattle for >50,000 markers costing about \$250 / animal. Sale of the chip began in December 2007, and more than 19,000 dairy animals in North America were genotyped by December 2008. Seven different laboratories have provided data to the North American database, including BFGL, U. Missouri, U. Alberta, GeneSeek, Genetics and IVF Institute, Genetic Visions, and Illumina. Database tools were constructed to automate cooperation in North America, to manage genotyping requests for proven bulls from stored DNA, to avoid duplication while genotyping young bulls and females on farms, and to store and edit genomic data (Wiggans *et al.*, 2008).

Access to Genotypes

Sharing of genotypes could greatly increase genetic progress. Unfortunately, free sharing or trading may lead to less profit rather than more profit. If some organizations make genotypes freely available, other organizations will have a competitive advantage by combining the free genotypes with additional privately held genotypes, eventually driving the free access organizations out of business. Contracts to trade or share genotypes can be established by pairs of organizations, but the number of organization or country pairs is potentially very large. Genomic data could be sold to or given away or stolen by a third party without agreement of each of the original contributors, and with little ability to detect the leak.

Global sharing may be seen as ideal, but requires 100% cooperation and provides little incentive for an individual or organization to genotype more ancestors to improve accuracy. Many organizations may join in a cooperative group such as those in North America or those in Europe, but contributors will still fear that competitors outside the cooperation will gain access to the shared genotypes without

providing any in return. As a result, genomic data are now tightly controlled and often not available even within the same country to qualified researchers that most of the contributors trust.

Cooperators differ in size, and this makes exchange more difficult. Suppose a large organization has 5000 genotypes and a small organization has 1000 genotypes. Three potential options would be 1) to trade no genotypes, 2) to trade an equal number of genotypes, say 1000, or 3) to let both parties use all 6000 genotypes. Genetic progress increases the most with option 3), access to all genotypes. Benefits of trades 2) and 3) are larger in proportion for the small organization, but larger in total for the large organization because the extra 1000 genotypes will affect more selection decisions and result in a larger total profit.

Proportional exchange of genotypes (number exchanged is proportional to size) would seem to be a fair system, but the bargaining power of a small organization is poor because its existence depends on the trade much more than the large organization's. Thus, the large organization may force the small organization to choose either a poor deal such as 2) or no deal at all (1). The profit motive does not result in maximum progress but instead leads to a natural monopoly in which the largest organization can dominate its smaller competitors. Dairy cattle breeders are accustomed to perpetual, long-term national and international cooperation. All breeders can benefit from genomics if organizations provide reasonably priced, fair, and open service to the public.

North American cooperators have 4 more years of exclusive access remaining for genomic predictions of young bulls using genotypes from the CDDR. After April 2013, free access may not be desirable because it provides no source of funds to re-genotype DNA with a denser chip and no protection against other organizations using the free genotypes while withholding their own data. A better long term strategy may be a service fee or tax for young animal predictions to pay for continued genotyping or re-genotyping of ancestor DNA.

Organizations

Breeding companies compete to increase market share, decrease costs, and speed rate of genetic improvement, but previously did not often work directly together on traditional selection programs. Genomic selection requires much more cooperation. Breeders understand that many genes affect most economic traits, and tools are now available to trace QTL that have very small effects. However, very large data sets are needed for estimating the small effects of individual genes. Most North American AI companies were convinced of this principle already in 1992 (Da *et al.*, 1994). During the last decade, ABS Global, Accelerated Genetics, Alta Genetics, Genex, New Generation Genetics, Select Sires, Semex Alliance, and Taurus-Service worked together to store semen for all bulls that they progeny tested in the Cooperative Dairy DNA Repository (Ashwell and Van Tassell, 1999).

Interaction and cooperation among these companies is facilitated by the National Association of Animal Breeders (NAAB, Columbia, MO). NAAB has a 5-year exclusive license on genomic predictions for young bulls computed using data from the CDDR. Breeding companies outside of North America have asked to join this consortium, but no mechanism has yet been found to extend the cooperation. The companies that invested in storing DNA, developing the SNP chip, and demonstrating its value want to ensure a return on their investment. This return was achieved by limiting access to the data rather than access to the SNP chip.

Phenotypic Data

Estimation of marker effects requires matching phenotypic data to genomic data. Dairy breeders already have a highly advanced system to compare breeding values for many phenotypes of dairy cattle worldwide. National EBVs are sent to the Interbull Centre, and an EBV for each bull within each environment is returned using multi-trait across country evaluation (MACE) and estimates of genetic correlations for the same trait expressed in different countries. This global system allows

the US and Canadian bulls to be fairly compared and also allows EBVs from bulls of other countries to be included easily. Genotyped animals in the North American database currently include 15,860 from the US, 3,082 from Canada, and 522 from 13 other countries.

Female and young bull EBVs are not exchanged by Interbull, but can be obtained with more work and cooperation. For many years, country pairs such as US and Canada have routinely exchanged and converted EBVs for requested females. Unfortunately, some breeders attempt to genotype young animals even before the pedigrees are reported to breed associations. More rapid exchange of information becomes important as generation intervals decrease using genomic selection.

Reliability

Genomic predictions now achieve 60-70% reliability for almost all traits of Holsteins using genotypes for 6,184 bulls and 1,637 cows with records to predict merit of young animals. For comparison, traditional parent averages used for selection of young animals have reliability of about 35% for most traits. Progeny tested bulls with 100 daughters records often have traditional reliabilities of 65-90%, depending on heritability of the trait. Reliability gains from genomic information average about 15 daughter equivalents for yield traits and up to 100 daughter equivalents for fertility. As a result, genomic evaluations replaced traditional USDA evaluations for males and females in January 2009.

Gains in reliability for Holstein, Jersey, and Brown Swiss populations were tested and are compared in Table 1. For Holsteins, 4,422 proven bulls and 947 cows with records in November 2004 were used to predict January 2009 daughter deviations of 2,035 younger bulls. Methods were the same as VanRaden *et al.* (2009), but an extra year of predictor animals was included as compared to the previous August 2003 cutoff study. The Jersey cutoff included 1,149 older bulls and 212 cows with records to predict 388 younger bulls. The Brown Swiss study used 228 proven bulls to predict 118 younger bulls.

Reliability averaged 29% higher than parent average reliability for the 27 traits of Holsteins. This average was 6% higher than the 23% gain using 3,569 predictor bulls in VanRaden *et al.* (2009). The average included 4 traits not tested in the other breeds, with gains of 31% for body depth, 18% for legs (rear view), 7% for sire calving ease, and 20% for daughter calving ease. For 23 traits of Jerseys, average gain in reliability was 11% above parent average reliability. For 6 of 24 traits of Brown Swiss, negative signs indicate that parent averages were more accurate than the genomic predictions, and for most other traits the reliability gains were much smaller than Holstein and Jersey gains. Clearly, accurate genomic evaluations require large data sets.

Table 1. Gains in reliability for Holstein, Jersey, and Brown Swiss breeds.

Trait	Gain in reliability		
	HOL	JER	BSW
Net Merit	24	8	3
Milk	26	6	0
Fat	32	11	5
Protein	23	0	1
Fat %	50	36	10
Pro %	35	29	5
Longevity	32	8	2
SCS	23	5	16
Fertility	28	7	-
Final score	20	1	-
Stature	34	16	3
Strength	28	12	5
Dairy form	32	13	-
Foot angle	25	8	-
Legs (side)	29	20	6
Rump angle	34	1	11
Rump width	29	9	12
Fore udder	36	9	-
Rear udd ht	24	16	-
Udder depth	37	18	3
Udder cleft	25	2	9
Teat place	33	7	1
Teat length	38	22	7
Average	29	11	N/A

International cooperation is even more important for smaller breeds such as Jersey, Brown Swiss, or Red Dairy Cattle than for the very large Holstein breed. Negotiations have begun on exchange of genomic data with

European Brown Swiss breeders, particularly Switzerland. Low genetic correlations can reduce the benefits of cooperation (Mulder and Bijma, 2006), but genomic methods can evaluate each animal's genes in multiple environments at no additional cost.

Future genomic advances may be driven by commercial applications, with an emphasis on low cost chips applied to these large populations. The transition from microsatellites for parentage testing to a SNP panel such as that of Heaton *et al.* (2007) will require international cooperation. Higher density chips can improve reliability but will require more investment. Recent advances in genomics have occurred quickly, without being slowed down by intellectual property disputes. Patenting would seem to be the exact opposite of cooperation.

Software

Software developed for human DNA research often does not apply to cattle because breeding value estimation was not the goal. Breeders of other animal and plant species can easily benefit from the same methods used by dairy cattle breeders because goals are similar. For example, Bernardo and Yu (2007) recommended applying the same methods to corn breeding. Genomic software could potentially be exchanged much more widely than the previous specialized programs for phenotypic analysis. The same computer algorithms for allele frequency estimation, genomic relationships, genomic inbreeding, haplotyping, and even raw genotype editing could be used across a wide variety of species because the biology of SNPs is constant even while phenotypic data and breeding programs differ. However, many researchers may still desire to write their own code because this could lead to further development and deeper understanding as compared to using standard, de-bugged, user-friendly, packaged programs.

Summary

Dairy cattle breeders are rapidly adopting genomic technology. Numbers of genotypes have grown and could continue to grow very

quickly. Simulation indicates that reliability for genotyped young animals could continue to increase to >80% if number of genotyped proven bulls increases to 40,000 or more. Actual Holstein predictions were tested using 5,369 genotyped bulls and cows as predictors, with gains in reliability for 27 traits averaging 29% above parent average reliability. Large gains in reliability require large families and large numbers of SNP because most traits are influenced by many genes with small effects.

Cooperation across country borders such as that of AI companies in North America helps to obtain the large numbers of genotypes needed to make genomic selection of young animals accurate. Genomic predictions for the smaller North American Jersey population and much smaller Brown Swiss population had much smaller gains in reliability than those for Holsteins. For Holsteins and Jerseys, genomic evaluations replaced traditional USDA animal model evaluations in January 2009 because breeders are convinced that the new predictions are more accurate.

Acknowledgments

This project was supported by National Research Initiative Grants 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 USA, and US Jersey. The authors thank staff of the Bovine Functional Genomics Laboratory for assistance in providing data and staff of the Animal Improvement Programs Laboratory for data analysis and manuscript improvement.

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.)
- Bernardo, R. & Yu, J. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 2007 47, 1082-1090
- Da, Y., Ron, M., Yanai, A., Band, M., Everts, R.E., Heyen, D.W., Weller, J.I., Wiggans, G.R. & Lewin, H.A. 1994. The dairy bull DNA repository: A resource for mapping quantitative trait loci. *Proc. 5th World Congr. Genet. Appl. Livest. Prod.*, Guelph, ON, Canada 21, 229–232.
- Fikse, W.F. 2004. Comparison of performance records and national breeding values as input into international genetic evaluation. *J. Dairy Sci.* 87, 2709-2719.
- Heaton, M.P., Snelling, W.M., Smith, T.P.L., Keele, J.W., Harhay, G.P., Wiedmann, R.T., Bennett, G.L., Freking, B.A., Van Tassell, C.P., Sonstegard, T.S., Gasbarre, L.C., Moore, S.S., Murdoch, B., McKay, S.D., Kalbfleisch, T. & Laegreid, W.W. 2007. A marker set for parentage-based DNA traceback in beef and dairy cattle. *Proc. Plant & Animal Genomes XV Conference*, San Diego.
- Maltecca, C., Bagnato, A. & Weigel, K.A. 2004. Comparison of international dairy sire evaluations from meta-analysis of national estimated breeding values and direct analysis of individual animal performance records. *J. Dairy Sci.* 87, 2599-2605.
- Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M., Heaton, M.P., O'Connell, J.R., 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).
- Mulder, H.A. & Bijma, P. 2006. Benefits of cooperation between breeding programs in the presence of genotype by environment interaction. *J. Dairy Sci.* 89, 1727-1739.
- National Human Genome Research Institute. 2008. Data release and access principles and policy. <http://www.genome.gov/10001801>
- 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.S. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92, 16-24.
- Wellcome Trust. 2003. Sharing data from large-scale biological research projects. <http://www.wellcome.ac.uk/About-us/Publications/Books/Biomedical-science/WTD003208.htm>

Wiggans, G.R., Sonstegard, T.S., VanRaden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Chesnais, J.P., Schenkel, F.S. & Van Tassell, C.P. 2008. Genomic evaluations in the United States and Canada: A collaboration. *Proc. Inter. Comm. Anim. Recording*, Niagara Falls, NY, June 16–20, 6 pp. 2008.