Using Genomic Data to Improve Dairy Cattle Genetic Evaluations

Paul VanRaden[1], George Wiggans[1], Tad Sonstegard[2], and Curt Van Tassell[2]
USDA [1]Animal Improvement Programs and [2]Bovine Functional Genomics Laboratories
Beltsville, MD 20705

Genomic data

Genotypes for about 40,000 single nucleotide polymorphisms (SNP) now act as a third source of data for national genetic evaluations of dairy cattle, in addition to phenotypes and pedigrees that were the basis of selection for the previous 100 years. Rapid developments in genotyping tools have lowered the cost of obtaining this genomic data to just over $200 per animal. The Illumina BovineSNP50 BeadChip (Matukumalli et al., 2008), developed in cooperation with USDA's Bovine Functional Genomics Laboratory (Beltsville, MD) and the University of Missouri (Columbia, MO), has been used to genotype >15,000 dairy animals in North America already during the first year of sale. This same chip is being used worldwide by dairy cattle breeders, with rapid adoption across Europe and in Oceania.

Advances in genomic technology in recent years were driven primarily by human DNA sequencing and genotyping projects. The human sequence was completed in 2001, the cattle sequence in 2004, and the pig sequence is now being constructed. Design of the cattle SNP chip required obtaining equally spaced SNP that are highly polymorphic across many dairy and beef breeds (Van Tassell et al., 2008) so that the BovineSNP50 chip could be used for Bos taurus breeds and also for Bos indicus breeds, with possible use for buffalo, bison, yak, etc. Future advances in genomic technology may be driven more by commercial applications to animal breeding programs, with an emphasis on low cost chips applied to large populations.

Quality of genomic data is very high (Wiggans et al., 2008). Most SNP have <1% missing genotypes, and error rates are <0.05%. Missing genotypes can be filled in and errors corrected by comparing SNP of closely related animals. About 10,000 of the 50,000 SNP that were polymorphic across breeds had minor allele frequencies <2% in Holsteins and were not used in genomic evaluation. Seven different laboratories provided data, including BFGL, U. Missouri, U. Alberta, GeneSeek, Genetics and IVF Institute, Genetic Visions, and Illumina.

Changing goals

Breeders have wondered how many and what genes affect the traits we measure ever since Mendel's laws were rediscovered >100 years ago. Marker assisted selection attempted to trace large quantitative trait loci (QTL) using nearby markers linked within families. For most traits, few large QTL were discovered, and most variation could not be traced even with hundreds of markers. Fine mapping attempted to find causative mutations, but these were very difficult to find because most QTL effects were small, several candidate genes were located nearby, and too few DNA crossovers occurred to allow separation of effects.

Newly designed chips contain many thousands of markers, and genetic effects can be traced across families because recombination is rare between a QTL and very close markers. The research focus has returned to genetic evaluation rather than gene

discovery. Dairy cattle breeders are now integrating whole genome marker data directly into genetic merit predictions without attempting to track individual QTL. The methods can be viewed as purely quantitative genetic, but using genomic relationships instead of traditional pedigree relationships. Breeders understand that many genes affect most economic traits, and tools are now available to trace QTL that have very small effects.

Population structure

Breeding companies compete for market share, rate of genetic improvement, and cost effectiveness, but previously have not worked directly together on traditional selection programs. To be effective, genomic selection may require much more direct cooperation. Dairy, beef, and sheep improvement programs have been very open, with much exchange of genetic material, whereas swine and poultry improvement programs tended to be more closed and isolated within large companies. Dairy breeds are now very global, with breeders across Europe, North and South America, Asia, and Oceania all purchasing semen and embryos from elite males and females on other continents. Local breeders then sell these improved combinations of genes back to the breeding companies.

Very large populations are needed for estimating the small effects of individual genes. North American AI companies were convinced of this principle already in the early 1990's and have worked together since then to establish a DNA repository (Ashwell and Van Tassell, 1999). Semen from all bulls being progeny tested and from important ancestor bulls was contributed by each of the 7 largest companies in the United States and Canada, as documented in Table 1.

Table 1. Contributors to the Cooperative Dairy DNA Repository (CDDR).

Organization	Address	Country
ABS Global	DeForest, WI	USA, British owned
Accelerated Genetics	Baraboo, WI	USA
Alta Genetics	Balzac, AB	Canada
Genex Cooperative	Shawano, WI	USA
New Generation Genetics	Fort Atkinson, WI	USA
Select Sires	Plain City, OH	USA
Semex Alliance	Guelph, ON	Canada
Taurus-Service	Mehoopany, PA	USA

Several of these companies have foreign subsidiaries and / or global distribution systems and combined semen sales of perhaps 250 million US dollars. Interaction and cooperation among the companies is facilitated by the National Association of Animal Breeders (NAAB, Columbia, MO). NAAB has a 5-year exclusive license on genomic predictions for males computed using genotypes obtained 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

Genotypes must be matched to phenotypes to estimate SNP effects. Dairy breeders have a highly advanced, global system to compare breeding values for all bulls worldwide. National computing centers use BLUP animal models for normally distributed traits or nonlinear mixed models for non-normal traits and forward the resulting national EBVs to the Interbull Centre in Uppsala, Sweden. Interbull estimates genetic correlations of performance for the same trait across different countries and uses multi-trait across country evaluation (MACE) to obtain an EBV for each bull for expected performance within each environment. About 25 traits currently are evaluated, and then genomic evaluations are computed using EBV rather than raw data as inputs. The global system allows the US and Canadian bulls in our project to be fairly compared and also allows phenotypic data from other countries to be incorporated easily. International cooperation is even more important for smaller breeds such as Jersey, Brown Swiss, or Ayrshire than for the very large Holstein breed.

Increasing reliability

Dairy breeders report reliability as the squared correlation of EBV with true BV in an unselected, randomly mating population (accuracy squared). For most traits, the parent averages used for selection of young animals have reliability of about 35%. Progeny tested bulls often have 100 daughters with records and reliabilities of 65-90%, depending on heritability of the trait. Genomic predictions can 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. Reliability for progeny tested bulls also increases significantly for all traits, with information gains averaging about 15 daughter equivalents. As a result, genomic evaluations are scheduled to replace traditional USDA evaluations in January 2009. Dairy breeding programs with more rapid turnover of generations could result in >50% faster progress.

Gains from genomic evaluation increase as more animals and more SNP are genotyped. Evaluation methods were described and tested on simulated data by VanRaden (2008). To test gains on real data, genotypes and EBVs from 3,576 older bulls born before 1999 were combined to predict daughter deviations of 1,759 younger bulls born after 1999. Subsets of those bulls were also examined as genotypes became available. Reliability increased almost linearly with the number of bulls used for prediction (VanRaden et al., 2008). Table 2 provides those results.

Table 2. Coefficients of determination ($R^2 \times 100$) for parent average and genomic prediction of net merit for bull subsets

Bull subset		Parent	Genomic	Gain from genomic prediction compared
Predictor bulls, n	Predicted bulls, n	average	prediction	with parent average
1,151	251	8	12	4
2,130	261	8	17	9
2,609	510	8	21	13
3,576	1,759	11	28	17

Value of additional SNP was examined by including only every other or every fourth SNP sequentially across the chromosomes. Value of SNP on the X chromosome was also examined by including or excluding the 605 SNP on X. Inclusion of only one quarter or one half of the SNP produced gains about 80% or 90% as large as those from the full set. Inclusion of SNP on the X chromosome also produced moderate gains in predictive ability. The trait with largest genomic gain was fat percentage because a major QTL (DGAT1) exists. Table 3 provides those results from VanRaden et al. (2008).

Table 3. Coefficients of determination ($R^2 \times 100$) for parent average and for genomic predictions with differing numbers of markers

		Number of markers			Without X
Trait	Parent average	9,604	19,208	38,416	37,811
Net merit	11	25	26	28	27
Milk yield	28	45	47	49	47
Fat yield	15	41	43	44	43
Protein yield	27	45	46	47	46
Fat percentage	25	59	61	63	62
Protein percentage	28	48	53	58	53
Productive life	17	24	25	27	26
Somatic cell score	23	34	36	38	36
Daughter pregnancy rate	20	27	28	29	29

Results for Jersey and Brown Swiss populations have also been tested but not yet published. The Jersey breed has sufficient bulls to achieve significant gains for most traits, whereas the US Brown Swiss breed obtained less favorable results because of fewer bulls. Large populations of Brown Swiss exist in some European countries, and negotiations have begun on exchange of genomic data with them. Prediction equations developed within one breed (such as Holstein) are not accurate when applied to genotypes from another breed because generations of recombination and drift change marker-QTL associations. However, research from New Zealand (Bevin Harris, pers. comm. 2008) indicates that joint evaluation of genotypes from multiple breeds can successfully predict crossbred performance.

Numbers of genotypes have grown and could continue to grow very quickly. Prediction equations have been tested on larger simulated data sets than the current actual data, but approximate methods of processing may be needed for larger data sets in the near future. Even though cost of genotyping currently limits use of the BovineSNP50 chip to only elite animals, breeders of Holsteins and Jerseys in North America are making effective use of genomic selection.

Lower cost methods are being developed to genotype only subsets of SNP with the largest effects or SNP that provide most information. Subsets can be targeted to overall merit or to a particular trait, and haplotyping can be used to predict merit for any trait if parent genotypes are available (Habier et al., 2008). A low cost chip can also be used for parentage checking and animal traceback. Denser chips will likely become available in future years and include the current 50,000 SNP as a subset.

Conclusions

Genomic data greatly increase reliability of predicted merit when genotypes are matched to phenotypes for large numbers of animals. Reliability of dairy cattle predictions were tested using about 5,000 genotyped bulls averaging 2,000 daughters each, representing a total of >10 million phenotyped animals. Genomic analyses were not attempted within companies or even within countries but rather across all major AI companies in North America. Large gains in reliability require large families and large numbers of SNP because most traits are influenced by many genes with small effects. Dairy cattle breeders have rapidly adopted genomic technology, and genomic evaluations will replace traditional animal model evaluations in the US and Canada in 2009.

Funding

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, by the National Association of Animal Breeders, Holstein USA, 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. Both laboratories allocated substantial funds to make this project possible.

References

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

Habier, D., J.C.M. Dekkers, and R.L. Fernando. 2008. Genomic selection using low-density SNPs. J. Dairy Sci. 91 (E-Suppl. 1):506 (abstract 521).

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

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

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

VanRaden, P.M., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F. S. Schenkel. 2008. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. (In press).

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 quality of genotypes used in genomic evaluation of dairy cattle. J. Dairy Sci. (submitted).