GENOMIC EVALUATIONS: PAST, PRESENT, FUTURE

George R. Wiggans and Tabatha A. Cooper
Animal Improvement Programs Laboratory, Agricultural Research Service, USDA

An expanded version of this report has been submitted to the Journal of Dairy Science.

ABSTRACT

The implementation of genomic evaluations for dairy cattle has caused profound changes in dairy cattle breeding. All young bulls purchased by major artificial insemination (**AI**) organizations are selected based on genomic evaluations. The reliability of those evaluations can reach around 75 % for yield traits, which is adequate for wide marketing of semen for 2-yr-old bulls. The shortened generation interval from using genomic evaluations is the most important factor in increasing the rate of genetic improvement. Genomic evaluations are based on 43,382 single nucleotide polymorphisms (**SNP**) generated from the Illumina BovineSNP50 BeadChip. This technology, which resulted from an international consortium of government, university, and industry cooperators, became available in December 2007, and the first unofficial USDA evaluations based on SNP genotypes were released in April 2008. Genomic evaluations became official for Holsteins and Jerseys in January 2009 and for Brown Swiss in August 2009. A steady increase in evaluation accuracy has resulted from including additional bulls with genotypes and traditional evaluations. Much of the increase occurs automatically as bulls genotyped as young bulls receive a traditional evaluation at 5 yr of age. Cow evaluations also contribute to evaluation accuracy, and that contribution is increased by adjusting their evaluations to the same mean and variance as bull evaluations. However, that adjustment reduces the evaluations of most cows. Full integration of U.S. and Canadian genotype databases provided a critical source of genotypes to achieve acceptable accuracy initially and continued benefits for both countries. Exchange of genotypes with other countries has added predictor bulls for Brown Swiss, and collaboration for other breeds is expected. In July 2010, Illumina released two new genotyping chips: the 3K chip with 2,900 SNP and the high density (**HD**) chip with 777,962 SNP. The 3K chip is expected to increase greatly the number of animals genotyped and replace microsatellites in parentage verification. The HD chip can provide more accurate genomic evaluations by better tracking of the loci responsible for genetic differences. To integrate multiple chips, a method to impute missing genotypes was developed. That method is based on splitting each genotype into its maternal and paternal haplotypes and tracing their inheritance. The same method is used to impute genotypes of nongenotyped dams based on their genotyped progeny and mates. The reliability of the resulting evaluations is appropriately discounted to reflect errors inherent in the process. Increases in evaluation accuracy are expected to continue because of added predictor animals and more SNP. The large population of existing genotypes can be used for the evaluation of new traits; however, the challenge is to measure the new traits for enough animals to allow estimation of SNP effects with sufficient accuracy for application to the general population.

INTRODUCTION

Genetic evaluation of dairy cattle has provided the means for steady genetic improvement in production, fitness, and conformation traits. The evaluations have been based on milk recording and breed association programs for type traits. Widespread use of superior bulls through AI has been the primary vehicle for progress. Identification of superior bulls has been expensive and time consuming because of the need to wait for milking daughters and the cost of collecting their data to achieve an evaluation of adequate accuracy. The great promise of DNA analysis has recently become a reality with the advent of low cost genotyping of large numbers of SNP markers.

PAST

For many years and many species, DNA markers have been used for research and as a tool. Programs for parentage verification began with analysis of blood groups and transitioned to microsatellite markers. Those markers were investigated for use in selection programs but were, at best, only of modest value. For dairy cattle, traits of economic importance are controlled by many genes, each of small effect; therefore, many markers are required to track inheritance. The search for major genes in dairy cattle was similarly disappointing as only a small portion of genetic variation could be explained by the collection of markers investigated.

The critical change was the development of assays that can be used to genotype large numbers of SNP at low cost. Although SNP are only biallelic (2 states), the large number available allows tracking the inheritance of short chromosomal segments. An international consortium led by USDA's Dr. Curtis Van Tassell worked with Illumina (San Diego, CA) to develop a set of SNP to be included on a chip (Van Tassell et al., 2008). From over 60,000 SNP, a commercial set of 54,001 was included in original release of the BovineSNP50 BeadChip. Consortium members had access to the new chip in the fall of 2007, and it became publicly available in late December 2007.

Genotypes from chips processed at the Beltsville Agriculture Research Center, University of Missouri, and University of Alberta were used in initial research to determine which SNP should be used in genomic evaluation. Some SNP were excluded because of low call rate, poor calling properties, or high correlation with other SNP (Wiggans et al., 2009). Procedures were developed to check for parent-progeny conflicts and other inconsistencies (Wiggans et al., 2010b). Extensive simulation work (VanRaden, 2008) enabled development of genomic evaluation methods, which were applied once genotypes became available. The first unofficial USDA evaluations based on SNP genotypes were released in April 2008. Genomic evaluations became official for Holsteins and Jerseys in January 2009 and for Brown Swiss in August 2009.

The cost of genotyping thousands of animals was covered by research grants and contributions from AI and breed organizations. In return for their support, the AI organizations received the exclusive right to have males genomically evaluated until May 2013. The commercial laboratories of GeneSeek (Lincoln, NE) and Genetics & IVF Institute (Fairfax, VA) participated in the initial genotyping and processing of commercial samples. They were joined later by DNA LandMarks (Quebec, Canada) and Genetic Visions (Middleton, WI).

PRESENT

Genomic Evaluation

Nomination

Over 2,000 new genotypes are received by USDA each month. Efficient procedures have been developed to accommodate such a large amount of data. The AI and breed organizations that arrange for genotyping are designated as requesters. They arrange for a sample (usually hair follicles) to be collected and attached to a bar-coded mailer. That mailer is usually sent to the requester, but may be sent directly to the genotyping laboratory. The bar code facilitates sample processing at the laboratory. The requester is expected to nominate each animal by making an entry in a database maintained by USDA's Animal Improvement Programs Laboratory (**AIPL**) before the sample reaches the genotyping laboratory. The nomination is either through a web interface or pedigree records containing the bar code (also known as sample identification). The nomination process insures that the pedigree for the animal is in the AIPL database before the genotype arrives at AIPL and simplifies matching the identification associated with the genotype with the animal's information in the AIPL database.

Genotyping

At the genotyping laboratories, DNA is extracted from the sample (blood, hair, semen, or nasal swab). The process of amplification of the DNA, fragmentation of the DNA, hybridization to the chip, labeling, and detection of the genotype takes 3 d. The data in the resulting intensity files (data generated from the laser reader) are then clustered to determine the SNP genotypes (Illumina, 2010). Those genotypes and the corresponding identification information are then transferred to AIPL.

Genotype Storage and Validation

The AIPL database can store multiple genotypes for an animal and relies on the chip identification and sample location on the chip to identify a sample uniquely. As samples are loaded, they are checked on an animal basis for call rate and parent-progeny conflicts. In addition to conflicts with reported parents, a conflict also is designated if comparison with all other genotypes indicates that an animal has a parent-progeny relationship that is not found in the pedigree (usually the genomically correct parent). A report of SNP with less than a 90 % call rate, SNP with a departure from Hardy-Weinberg equilibrium (difference between number of expected and actual heterozygous SNP), or SNP with over 1 % parent-progeny conflicts is returned to the submitting laboratory. Those checks serve as a measure of the quality of the genotype calls. The database allows for storage of genotypes from chips with differing numbers of SNP. Currently, a low density (3K) chip with 2,900 SNP, the BovineSNP50 BeadChip, and an HD chip with 778,962 SNP are supported. Comparisons of SNP genotypes from different chips are supported, but limited to the SNP in common.

The 3K chip, which was introduced by Illumina in July 2010, is expected to increase greatly the number of animals genotyped and replace microsatellites in parentage verification.

Many conflicts can be resolved. For most cases of sire conflict, an alternative sire is suggested. Identical genotypes often are the result of embryo splits or identical twins. Because bulls have only one X chromosome, their genotypes for X-specific SNP appear to be homozygous, and that characteristic is used in sex validation. Some cows inherit both of their X chromosomes from the same male ancestor and, therefore, appear to be males. If a common male ancestor can be found, genotypes for such cows are accepted. The usability of genotypes is evaluated whenever the pedigree of a genotyped animal changes.

Genotype Preparation

Genotypes for 43,382 SNP are extracted from the database. During extraction, multiple genotype calls for an individual animal are merged. Identical twins and animals from split embryos have their genotypes harmonized. Genotypes are imputed (constructed from relatives) for dams with a sufficient number of genotyped progeny and mates to reach a 90 % call rate on an allele basis. Since April 2010, dams with imputed genotypes have been included in genomic evaluations (Table 1). Imputation also is used to add SNP that are on the SNP50 chip but not on the 3K chip. Imputation involves splitting the genotype into paternally and maternally contributed chromosomes (haplotypes). Haplotype inheritance is traced and used to fill in missing genotypes. Table 1 shows the number of usable genotypes by breed for most of the genomic evaluations released since April 2009.

Table 1. Numbers of genotyped dairy cattle by breed and U.S. evaluation release date.

Breed	Evaluation release date	Predictor animals[a]		Young animals[b]		Animals with imputed genotypes	All animals
		Bulls	Cows	Bulls	Cows		
Holstein	April 2009	7,600	2,711	9,690	1,943	0	21,944
	June 2009	7,883	3,049	11,459	2,974	0	25,365
	August 2009	8,512	3,728	12,137	3,670	0	28,047
	October 2009	8,568	3,965	13,288	4,797	0	30,618
	January 2010	8,974	4,348	14,061	6,031	0	33,414
	February 2010	9,378	5,086	15,328	7,620	0	37,412
	April 2010	9,770	7,415	16,007	8,630	1,471	41,822
	May 2010	9,958	7,940	16,594	9,772	1,955	44,264
	June 2010	9,958	8,122	17,507	10,713	2,004	46,300
	July 2010	9,963	8,186	18,187	11,309	2,035	47,645
	August 2010	10,430	9,372	18,652	11,021	2,029	49,475
	September 2010	10,611	9,453	19,389	13,333	1,990	**52,786**
Jersey	February 2010	1,977	479	1,172	197	0	38,25
	April 2010	2,072	637	1,250	202	97	4,161
	May 2010	2,079	702	1,308	231	150	4,320
	June 2010	2,088	740	1,391	259	148	4,478
	July 2010	2,088	753	1,461	273	153	4,575
	August 2010	2,145	792	1,476	258	152	4,671
	September 2010	2,153	793	1,590	282	148	**4,818**
Brown Swiss	February 2010	1,168	54	179	15	0	1,416
	April 2010	1,185	98	188	31	47	1,502
	May 2010	1,188	114	199	34	63	1,535
	June 2010	1,215	116	215	38	66	1,584
	July 2010	1,227	116	223	38	66	1,604
	August 2010	1,248	124	228	35	69	1,635
	September 2010	1,250	121	239	32	64	**1,643**

[a]Animals with traditional evaluations.
[b]Animals without traditional evaluations.

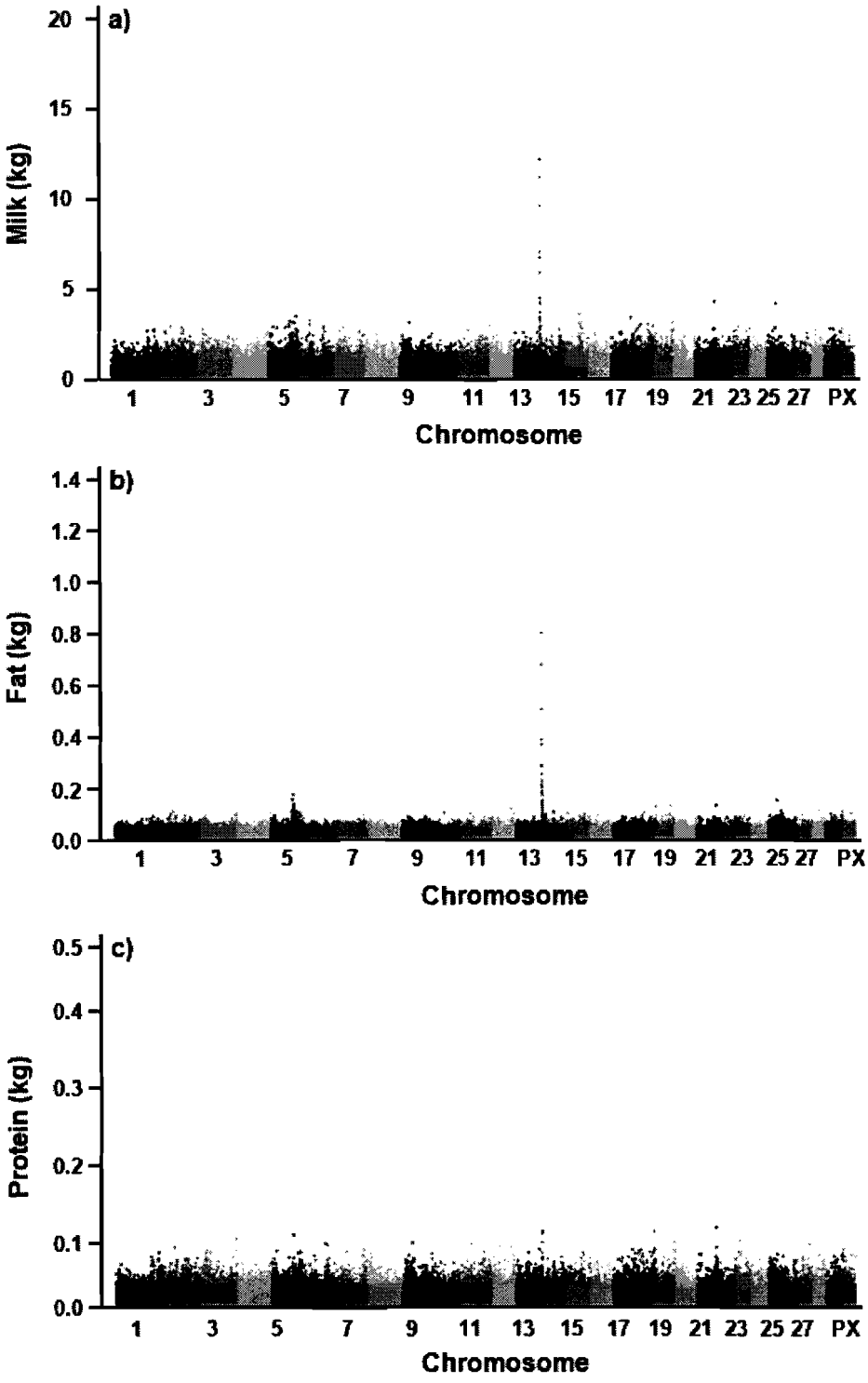

Figure 1. Absolute value of July 2010 single nucleotide polymorphism effects for a) milk, b) fat, and c) protein yields of Holstein predictor animals by chromosome (P = pseudo-autosomal region of X chromosome).

Traditional Evaluation

Traditional evaluations for over 30 traits are the basis for genomic evaluations. Evaluations are deregressed, which means that the shrinkage based on amount of information inherent in estimation of random effects is undone to make the data more like individual records. Cow evaluations for milk, fat, and protein yields and component percentages are adjusted to give them a mean and variance similar to bull evaluations with similar accuracy. That step is necessary to remove the overestimation usually associated with cow evaluations for yield traits compared with bull evaluations. Their evaluations must be comparable, because both cow and bull evaluations are used to estimate the SNP effects. Making that adjustment increased the gain in reliability from genomics by about 3 percentage units for yield traits (Wiggans et al., 2010a).

Estimation of SNP Effects

The effect on each trait from replacing one allele in the genotype with the other allele is estimated for each of the 43,382 SNP (VanRaden, 2008). Typically 100 iterations of computation are done. In addition to the SNP estimates, a polygenic effect is estimated to capture the genetic variation not accounted for by the SNP. An animal's genomic evaluation combines the SNP effect estimates, the polygenic effect, and information from traditional evaluations not already included in the genomic information. Absolute values of SNP effects for all 43,382 SNP used in evaluation of Holstein milk, fat, and protein yields are shown in Figure 1. Most SNP have small effects, which are distributed evenly across all chromosomes. Methods for the visualization of SNP effects were described by Cole and VanRaden (2010).

Measurement of Accuracy

Reliability measures how much information contributes to the evaluation. For genomic evaluations, the genomic relationship with predictor animals and their evaluation reliability are the primary determinants of accuracy. Thus, the genomic contribution is similar across animals; i.e., it is lower for less related animals such as those with foreign ancestors or subpopulations that contributed little to the current population (Wiggans and VanRaden, 2010). For milk, fat, and protein yields, reliabilities for most Holstein young bulls ranged from 73 to 79 % (Figure 2), which was adequate to support wide marketing of 2-yr-old bulls. Reliabilities are appropriately discounted to reflect errors inherent in the imputation process, which primarily affects evaluations based on 3K genotypes.

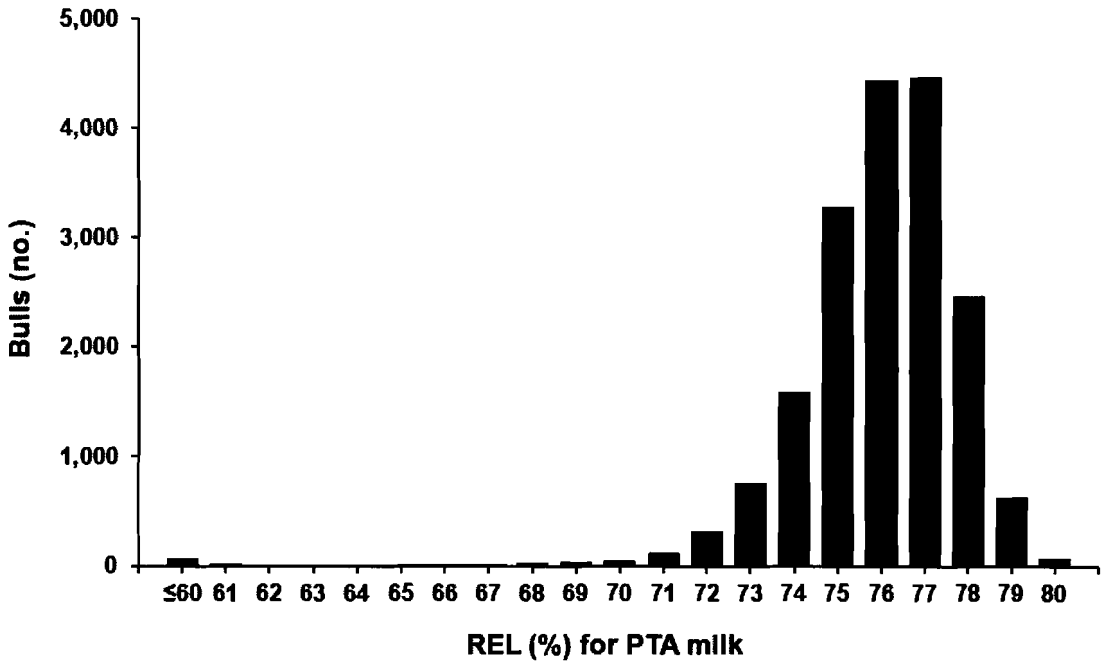

Figure 2. Distribution of July 2010 genomic reliabilities (**REL**) of predicted transmitting abilities (**PTA**) for milk yield for young Holstein bulls.

Table 2. Comparison between parent average and genomic reliabilities and bias in genomic evaluations when using August 2006 data to predict June 2010 bull evaluations by trait and breed.

		Reliability (%)			
Breed	Trait	Parent average	Genomic	Gain[a]	Bias[b]
Holstein	Milk (kg)	38.1	67.5	29.4	−4.0
	Fat (kg)	38.1	73.1	35.0	−0.9
	Protein (kg)	38.1	63.7	25.6	0.6
	Fat (%)	38.1	85.7	47.6	0.0
	Protein (%)	38.1	77.9	39.8	0.0
	Productive life (mo)	31.0	64.2	33.2	−1.5
	Somatic cell score	33.9	60.4	26.5	0.0
	Daughter pregnancy rate (%)	29.8	46.8	17.0	−0.2
	Sire calving ease	27.1	40.9	13.8	1.0
	Daughter calving ease	26.2	44.3	18.1	−1.0
	Sire stillbirth	22.7	29.8	7.2	2.1
	Daughter stillbirth	26.6	29.3	2.7	0.3
Jersey	Milk (kg)	39.5	53.9	14.3	89.8
	Fat (kg)	39.5	49.9	10.4	5.8
	Protein (kg)	39.5	49.1	9.6	3.4
	Fat (%)	39.5	64.9	25.3	0.0
	Protein (%)	39.5	61.4	21.8	0.0
	Productive life (mo)	24.2	50.8	19.1	−0.4
	Somatic cell score	18.7	48.9	13.8	0.1
	Daughter pregnancy rate (%)	24.1	60.0	29.2	−0.1
Brown Swiss	Milk (kg)	37.2	53.8	16.7	−163.0
	Fat (kg)	37.2	53.1	16.0	−6.3
	Protein (kg)	37.2	53.0	15.9	−4.1
	Fat (%)	37.2	59.1	22.0	0.0
	Protein (%)	37.2	57.8	20.6	0.0
	Productive life (mo)	28.3	54.2	25.8	−1.2
	Somatic cell score	32.2	53.4	21.2	0.0
	Daughter pregnancy rate (%)	24.9	28.1	3.0	0.0

[a]Genomic reliability minus parent average reliability.
[b]2010 deregressed value minus 2006 genomic evaluation.

Across all yield, health, and fertility traits (where applicable), the average genomic reliability was 57 % for Holsteins, 55 % for Jerseys, and 52 % for Brown Swiss when predicting June 2010 evaluations from August 2006 data (Table 2). Gains in reliability above parent average ranged from 2.7 to 47.6 percentage units for Holsteins, 9.6 to 29.2 percentage units for Jerseys, and 3.0 to 25.8 percentage units for Brown Swiss. Bias in genomic predictions also is shown in Table 2; a negative value indicates that the genomic prediction is higher than the deregressed value.

Distribution

Genomic evaluations are calculated monthly. At the 3 traditional releases, all genomic evaluations are released. Between those releases, only genomic evaluations of new animals or bulls that are not being marketed are released. Genomic evaluations of bulls that are being marketed are not released to avoid changes between traditional evaluations. Evaluations of bulls that are < 2 yr old and not enrolled in the cross-reference program of the National Association of Animal Breeders are distributed only to the owner and requesting AI organization.

FUTURE

Genomic evaluations are expected to continue to increase in accuracy. The largest contribution will be from additional predictor animals. Table 1 shows the natural increase in the U.S. predictor population at each traditional evaluation from bulls with a first progeny-test result at approximately 5 yr of age. The U.S. predictor population also increases the month following evaluation release when newly evaluated Canadian bulls can contribute. In addition, the

Cooperative Dairy DNA Repository (Ashwell and Van Tassell, 1999) has semen straws for over 10,000 bulls with evaluations, but that have not been genotyped. A way may be found to pay for genotyping some of those bulls. Work continues on how to make information from cows more useful. New Zealand has genotyped all their bulls, and are looking to cows as a way to add to their predictor population.

In July 2010, Illumina released an HD chip with 777,963 SNP. Another company is expected to offer an HD chip in late 2010. Although an HD chip can provide more accurate genomic evaluations by better tracking of the loci responsible for genetic differences, the gains are not expected to be large (VanRaden and Tooker, 2010). As with the 3K chip, HD SNP would be imputed from current genotypes. The first step is to collect enough HD genotypes so that most haplotypes are represented. Several thousand genotyped animals may be required. The ultimate density is full sequencing, and its cost has been dropping. With full sequencing for a substantial number of animals, SNP that are the causative mutation or are closely linked to it may be identified. Identification of those SNP may enable an increase in evaluation accuracy and a reduced number of SNP needed for evaluation. In addition, such a SNP set might be useful across breeds.

A method to calculate traditional and genomic evaluations in a single procedure is being developed (Misztal et al., 2010). Holstein Association USA plans to adopt that system for conformation evaluations in 2011.

The dairy industry has expressed interest in extending genomic evaluations beyond the Holstein, Jersey, and Brown Swiss breeds (Olson and VanRaden, 2010). To achieve acceptable evaluation accuracy, the use of information across breeds may be necessary, which may require the use of HD chips.

Collaboration

Collaboration is the least expensive way to increase the predictor population. Collaboration between the United States and Canada was quite successful in initially increasing the size of the predictor population and continues to add to it. Genotypes from the United States were traded with Switzerland to increase the number of predictor bulls for Brown Swiss. EuroGenomics is a collaboration among several European countries, which have a combined predictor population larger than in North America. That collaboration has increased interest in North America to increase the predictor population.

Research collaboration has helped to improve evaluation methodology, and coordination across country has aided with producer acceptance by minimizing differences and explaining existing differences. Discussions are underway for various levels of U.S. collaboration with other countries.

New Traits

A genotype can be used to evaluate all traits, which suggests that new traits can be easily added to the selection objective. However, genomic evaluations require accurate estimates of SNP effect, which in turn need accurate traditional evaluations. As milk production becomes increasingly specialized, the selection objective will need to account for more of the traits that affect profitability. Once sufficient data have been accumulated to evaluate a trait genomically, selection can be made at birth (or before) so that the benefit of shortened generation interval and resulting increase in rate of genetic gain can be realized.

CONCLUSIONS

Genomic evaluations have revolutionized dairy cattle breeding. Those evaluations are based on genotypes that are extensively checked for quality, and conflicts are resolved. They are becoming more accurate as animals are added to the predictor population. All young bulls purchased by major AI organizations now are selected based on genomic evaluations. The development, implementation, and acceptance of genomic evaluations have allowed the marketing of 2-yr-old bulls, which is expected to increase the rate of genetic improvement greatly.

ACKNOWLEDGMENTS

Dr. John Cole is thanked for his assistance in preparing the Manhattan plots. The AIPL staff is thanked for conducting the evaluations being reported and the research to support them.

LITERATURE CITED

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

Cole, J.B., and P.M. VanRaden. 2010. Visualization of results from genomic evaluations. J. Dairy Sci. 93:2727–2740.

Illumina. 2010. BovineSNP50 Genotyping BeadChip. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf.

Misztal, I., I. Aguilar, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 92:743–752.

Olson, K.M., and P.M. VanRaden. Multibreed genomic evaluation of dairy cattle. J. Dairy Sci. 93(E-Suppl. 1):471. (Abstr.)

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, 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., and M.E. Tooker. 2010. Gains in reliability from combining subsets of 500, 5,000, 50,000 or 500,000 genetic markers. J. Dairy Sci. 93(E-Suppl. 1):534. (Abstr.)

Wiggans, G.R., T.A. Cooper, and P.M. VanRaden. 2010a. Cow adjustments for genomic predictions of Holsteins and Jersey bulls. J. Dairy Sci. 93: 533–534. (Abstr.)

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. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431–3436.

Wiggans, G.R., and P.M. VanRaden. 2010. Improved reliability approximation for genomic evaluations in the United States. J. Dairy Sci. 93(E-Suppl. 1):533. (Abstr.)

Wiggans, G.R., P.M. VanRaden, L.R Bacheller, M.E. Tooker, J.L. Hutchison, T.A. Cooper, and T.S. Sonstegard. 2010b. Selection and management of DNA markers for use in genomic evaluation. J. Dairy Sci. 93:2287–2292.