Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Increasing the Number of Single Nucleotide Polymorphisms Used in Genomic Evaluations of Dairy Cattle

G. R. Wiggans, T. A. Cooper, D. J. Null and P. M. VanRaden
Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland, USA

ABSTRACT: A small increase in the accuracy of genomic evaluations of dairy cattle was achieved by increasing the number of SNP used to 61,013. All the 45,195 SNP used previously were retained, and 15,818 SNP were selected from higher density genotyping chips if the magnitude of the SNP effect was among the top 1,000 effects of the added SNP for at least 1 trait. The imputation of the additional SNP was based on 4,037 GeneSeek Genomic Profiler HD and 2,262 Illumina BovineHD genotypes available as of March 2013. Markers for 7 mutation tests based on recessive codes provided by breed associations were included in the SNP set. For Holsteins, the mean gain in reliability across all traits from using the additional SNP was 0.5 percentage points. Correlations of new with previous predictions were close to 0.99 for all traits.
Keywords: single nucleotide polymorphism; genomic evaluation; dairy cattle; genetics

Introduction

Genomic evaluation began in the United States in April 2008 (Wiggans et al., 2011) based on genotypes from the Illumina (San Diego, CA) BovineSNP50 BeadChip (Illumina, 2012a). From the over 54,000 SNP on the chip, around 45,000 were selected for evaluation of Holstein, Jersey, and Brown Swiss breeds. Selection was based on minor allele frequency, parent-progeny error rate, and deviation of heterozygosity from expected (Wiggans et al., 2010). In 2010, the Illumina BovineHD BeadChip became available with 777,962 SNP (Illumina, 2012b). Investigation by Harris et al. (2011) found only a small benefit from using all those SNP in genomic evaluation. More than half of the SNP were found to have high correlations with other SNP (VanRaden et al., 2013). The Illumina Bovine3K BeadChip (2,900 SNP) began being used in 2010 (Illumina, 2011) and in 2011 was replaced by the Illumina BovineLD chip (Boichard et al., 2012) with 6,909 SNP and the capability for additional customized content. GeneSeek (Lincoln, NE) has developed several chips with varying numbers of SNP for dairy cattle (Neogen Corporation, 2013a). The GeneSeek Genomic Profiler HD (GHD) has 76,999 SNP (Neogen Corporation, 2013b) that were selected to capture most of the benefit of SNP on the BovineHD chip by choosing those that were most informative in dairy genomic evaluations along with SNP for other objectives. The objective of this research was to determine if accuracy of genomic evaluations could be increased by using additional SNP from the GHD chip.

Materials and Methods

Data. Data collected for genomic evaluations were used in a cutoff study (VanRaden et al., 2009) to determine the accuracy of data from 4 yr earlier (August 2009 training set) in predicting performance of a validation set of bulls that did not have data 4 yr earlier. For Holsteins, April 2013 performance was predicted; the training set included 24,356 bulls, and the validation set included 810 to 4,395 bulls depending on the trait. For Jerseys, December 2013 performance was predicted; 5,678 bulls were in the training set, and 470 to 699 bulls were in the validation set. Several sets of SNP were compared with the 45,195 (45K) SNP that had been previously used for genomic evaluation. All SNP sets included markers for 7 mutation tests based on recessive codes provided by breed associations. The first tested SNP set included 61,013 (61K) SNP (45,195 SNP from the 45K set plus 15,818 GHD SNP). The added GHD SNP were chosen because the magnitude of their SNP effects was among the top 1,000 effects of the added SNP for at least 1 trait. The second tested SNP set was the 73,721 (74K) unique usable GHD SNP, which had 28,788 SNP in common with the 45K set. The final tested SNP set included 91,176 (91K) SNP (45,195 SNP from the 45K set plus 45,980 additional GHD chromosomal SNP, with Y-specific SNP excluded).

Gains in reliability due to genomics were determined by comparing reliabilities of parent average and genomic evaluation from August 2009 in predicting April 2013 daughter performance for Holstein bulls in the validation set and December 2013 performance for Jersey bulls.

Imputation of the 61K, 74K, and 91K SNP sets for Holsteins was based on 4,037 GHD and 2,262 HD genotypes available as of March 2013. To determine if prediction accuracy for the 61K set was improved by using the 2,416 BovineHD and 17,997 GHD genotypes available in February 2014 for Holsteins, individual SNP effects were computed using a training set of 25,148 Holsteins with traditional evaluations as of August 2009. December 2013 performance was predicted for validation sets of 1,122 to 5,140 bulls depending on trait. Investigation is underway to determine if the number of SNP in the 61K set should be reduced for routine evaluation by removing SNP that fail to meet requirements for call rate, parent-progeny discrepancies, or heterozygosity frequency for all chips (Wiggans et al., 2009).

Table 1. Holstein genomic evaluation reliabilities by trait and number of included SNP[1]

	Reliability, %, for SNP set[2]							
Trait	45K	61K		74K		91K	
Milk yield	69.2	69.3	(0.1)	68.9	(–0.3)	69.2	(0.0)
Fat yield	68.4	68.7	(0.3)	68.6	(0.2)	68.4	(0.0)
Protein yield	60.9	60.8	(–0.1)	60.6	(–0.3)	60.8	(–0.1)
Fat percentage	93.7	94.4	(0.7)	93.9	(0.2)	93.5	(–0.2)
Protein percentage	86.3	87.1	(0.8)	86.3	(0.0)	86.1	(–0.2)
Net merit	51.6	51.7	(0.1)	51.6	(0.0)	51.3	(–0.3)
Productive life	73.7	74.0	(0.3)	73.1	(–0.6)	73.8	(0.1)
Somatic cell score	64.9	65.8	(0.9)	65.6	(0.7)	65.6	(0.7)
Daughter pregnancy rate	53.4	54.1	(0.7)	53.6	(0.2)	53.8	(0.4)
Service-sire calving ease	45.8	45.7	(–0.1)	45.1	(–0.7)	46.2	(0.4)
Daughter calving ease	44.2	45.8	(1.6)	44.9	(0.7)	44.9	(0.7)
Service-sire stillbirth rate	28.2	28.3	(0.1)	28.7	(0.5)	29.9	(1.7)
Daughter stillbirth rate	27.6	37.8	(0.2)	37.1	(–0.5)	39.2	(1.6)
Final score	58.8	58.7	(–0.1)	58.4	(–0.4)	58.7	(–0.1)
Strength	70.2	70.7	(0.5)	70.1	(–.01)	70.5	(0.3)
Stature	68.5	69.0	(0.5)	68.8	(0.3)	69.1	(0.6)
Dairy form	71.8	72.2	(0.4)	71.9	(0.1)	72.0	(0.2)
Body depth	70.7	71.4	(0.7)	70.9	(0.2)	71.1	(0.4)
Foot angle	56.9	57.5	(0.6)	57.4	(0.5)	57.7	(0.8)
Rear legs (side view)	59.9	60.8	(0.9)	61.0	(1.1)	61.1	(1.2)
Rear legs (rear view)	54.3	54.7	(0.4)	54.5	(0.2)	54.6	(0.3)
Rump angle	70.2	70.9	(0.7)	70.7	(0.5)	70.9	(0.7)
Rump width	65.0	65.4	(0.4)	65.0	(0.0)	65.2	(0.2)
Feet and legs	44.0	45.1	(1.1)	45.1	(1.1)	45.1	(1.1)
Fore udder attachment	70.4	70.6	(0.2)	70.0	(–0.4)	70.4	(0.0)
Rear udder height	59.4	59.9	(0.5)	59.6	(0.2)	59.8	(0.4)
Udder depth	75.3	76.2	(0.9)	76.0	(0.7)	76.1	(0.8)
Udder cleft	62.1	62.2	(0.1)	62.0	(–0.1)	62.2	(0.1)
Front teat placement	69.9	70.1	(0.2)	70.2	(0.3)	70.4	(0.5)
Rear teat placement	59.3	59.6	(0.3)	59.5	(1.1)	45.1	(1.1)
Teat length	66.7	67.2	(0.5)	66.6	(–0.1)	66.9	(0.2)
All traits			(0.5)		(0.1)		(0.4)

[1]45K = 45,195 SNP; 61K = 61,013 SNP; 74K = 73,721 SNP; 91K = 91,176 SNP.
[2]Differences from evaluation reliability based on 45K SNP are in parentheses.

Results and Discussion

Correlations of new with previous predictions were close to 0.99 for all traits and were >0.996 for milk, fat, protein, daughter pregnancy rate, productive life, SCS, and final score. However, use of additional markers did change the carrier status predicted from haplotypes in some cases because each haplotype now had a different set of markers included and might have been shifted slightly to the left or right relative to the location of the true genetic defect contained within the haplotype.

Table 1 shows Holstein genomic reliability and gains over parent average from the 45K SNP set for the 61K, 74K, and 91K SNP sets by trait. The 61K set performed best with a mean gain of 0.5 percentage points in reliability and no traits had a loss of >0.1. The 75K had the poorest performance, which indicates that some of the SNP from the 45K set that were not included in the GHD chip make a contribution to accuracy even in the presence of SNP with higher minor allele frequency. One factor may be that a larger portion of SNP must be imputed; therefore, imputation errors may be reducing the accuracy gain. Results for the 91K SNP set were similar to those for the 61K set but more variable; for traits with losses, the losses were larger. For the 61K set with more high-density animals included, overall gain in reliability across all traits wwas 0.7 percentage points (compared with 0.5 for April 2013 performance), with increases of 0.5 for yield, 0.8 for health, 0.8 for calving, and 0.7 for type traits.

Contrary to expectation, additional SNP for some traits actually caused a loss in reliability. Although this could be a consequence of the random nature of the system, ideally any increase in the number of SNP included in the evaluation should improve reliability for nearly all traits by enough to justify any re-rankings that it caused. The most likely reason for reduced accuracy from including more SNP is that they were missing from most existing genotypes and, therefore, had to be imputed from the relatively few genotypes that included them. With the approximately 1,000 GHD and a few BovineHD genotypes

received each month, imputation accuracy will improve as demonstrated by the increase from using the genotypes available as of February 2014. The animals that had evaluations with the greatest changes between the previous and 61K sets were those with Bovine3K genotypes and those without parent genotypes; imputation is least accurate for such animals. Mean reliability gains for Jerseys (not shown) were similar to those for Holsteins but more variable; mean gain was 0.3 percentage points across all traits (1.0 for yield, 0.2 for health, and 0.1 for type traits).

Conclusion

For Holsteins, the mean gain in reliability across all traits was 0.5 percentage points for predictions using the 61K SNP set compared with the previously used 45K set; gains were smaller with 74K or 90K SNP sets. Gains were slightly larger for the 61 K set after adding many animals with GHD genotypes. The 61K SNP set improved the value of genomic testing and was adopted for official use in US genomic evaluations in December 2013.

Literature Cited

Boichard, D., Chung, H., Dassonneville, R. et al. (2012). PLoS ONE 7:e34130.

Harris, B. L., Creagh, F. E., Winkelman, A. M. et al. (2011). Interbull Bull. 44:3–7.

Illumina Inc. (2011). http://res.illumina.com/documents/products/datasheets/datasheet_bovine3k.pdf. Accessed on January 14, 1024.

Illumina Inc. (2012a). http://res.illumina.com/documents/products/datasheets/datasheet_bovine_snp5o.pdf. Accessed on January 14, 2014.

Illumina Inc. (2012b). http://res.illumina.com/documents/products/datasheets/datasheet_bovinehd.pdf. Accessed on January 14, 2014.

Neogen Corporation (2013a). http://www.neogen.com/Agrigenomics/Dairy.html. Accessed on January 14, 2014.

Neogen Corporation (2013b). http://www.neogen.com/Agrigenomics/pdf/Slicks/GGP_HD_Dairy.pdf. Accessed on January 14, 2014.

VanRaden, P. M., Null, D. J., Sargolzaei, M. et al. (2013). J. Dairy Sci. 96:668–678.

VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R. et al. (2009). J. Dairy Sci. 92:16–24.

Wiggans, G. R., Sonstegard, T. S., VanRaden, P. M. et al. (2009). J. Dairy Sci. 92:3431–3436.

Wiggans, G. R., VanRaden, P. M., Bacheller, L. R. et al. (2010). J. Dairy Sci. 93:2287–2292.

Wiggans, G. R., VanRaden, P. M., and Cooper, T. A. (2011). J. Dairy Sci. 94:3202–3211.