Changes to evaluation system (December 2010)

Evaluations using 2,900 markers

By Paul VanRaden, George Wiggans, Tad Sonstegard, Curt Van Tassell, Katie Olson, Tabatha Cooper, and Lillian Bacheller

Genotypes from a 2,900 (**3K**) marker panel developed by Illumina (San Diego, CA) in cooperation with the Bovine Functional Genomics Laboratory (Beltsville, MD) were included as a data source for genomic evaluations beginning in September 2010. Animals with 3K genotypes had unofficial evaluations in monthly XML files distributed to cooperators until those evaluations became official in December 2010. Multiple marker sets are included in the same evaluation by imputing all genotypes to the highest density. Accuracy of imputation was improved by correcting the locations of several markers on the bovine map using information provided by collaborators at the Universities of Maryland, Missouri, and Guelph. Numbers of markers used are currently 42,503 from version 1 of the BovineSNP50 BeadChip (**50K**), 41,019 from version 2 of the 50K chip (**50K_V2**), and 2,614 from the 3K chip. The 3K chip also includes 14 Y-chromosome markers that are not on the 50K chips; those markers are used only for sex determination.

Markers used in the December evaluation each had less than 20% missing genotypes and less than 2% parent-progeny conflicts, with stricter limits for markers with minor allele frequencies of less than 0.5 as in Wiggans et al. (2010). After excluding genotypes with less than 90% call rate and other edits, numbers of animals used to select markers were 54,643 with the 50K chip, 2,602 with the 50K_V2 chip, and 8,305 with the 3K chip. Genotypes from the 3K chip were provided by 3 laboratories. On average, the 2,614 3K markers had 0.7% missing genotypes, almost twice as many as the 0.4% missing for the 50K chip. The average parent-progeny conflict rate for the 2,614 selected 3K markers was 0.07%, which was very good but not as low as the 0.01% for the 50K chip. Conflict rate was calculated as the total number of conflicts observed divided by animals divided by markers. Conflicts that can be detected from pedigree are set to missing before imputation. The selected markers seem very useful; however, edits and numbers of markers used may change with experience.

Edit thresholds were adjusted to use the 3K genotypes. In most cases, the adjustment was made by defining all thresholds in terms of percentage of usable single nucleotide polymorphisms (**SNP**s). A new reason for exclusion was based on parent-progeny conflicts. Clustering of homozygotes and heterozygotes often is not as distinct for 3K as 50K genotypes, which results in more genotyping errors. The 3K chip relies on GoldenGate technology in contrast to the 50K chips, which use Infinium. Some animals with 3K genotypes have more than the usual number of conflicts, but fewer than when the parent is incorrect. With the 50K chips, the distinction between correct and incorrect pedigree was very clear. With the 3K chip, some animals were found with intermediate values, which were deemed to be unreliable genotypes rather than incorrect pedigrees. The reason for rejection is similar to excluding samples with a call rate of less than 90%.

Test files in which every 5th young bull or heifer had their 50K genotypes replaced by a genotype based on the 3K marker subset were provided to industry cooperators on August 27. Those files allowed cooperators to directly compare changes that occur by genotyping with the 3K chip. Slight differences between the August test and September genomic evaluations were that the most recent Interbull data contributed to 50K marker solutions and that dams were imputed from a mixture of 3K- and 50K-genotyped progeny instead of all 50K. The test files and genomic evaluation files distributed in September included a new field labeled "chip" that indicates if an animal's evaluation is based on imputation ("imputed"), the original 50K chip ("50K"), version 2 of the 50K chip ("50K_V2"'), the 3K chip ("3K"), or a high-density chip ("HD"). The test files included 3K and 50K but not 50K_V2 or HD data. The September actual files did include 50K_V2 data.

Gain in reliability (**REL**) for young animals genotyped with 3K markers averages about 80% of that resulting from 50K markers, with slightly better results if parents are genotyped and poorer if not. For example, if genomic REL with 50K markers is 70% and parent average REL is 35%, average REL for a 3K evaluation is expected to be 0.80(70% − 35%) + 35% = 63%. Cows and bulls with 3K genotypes can add to the reference population, but they contribute about 20% less information than animals with genotypes based on 50K markers. Lower RELs for 3K animals or imputed dams are computed with the formulas of VanRaden et al. (2010). For each animal, REL_{SNP} estimates the squared correlation of imputed and true genotypes, and REL_{MAX} is the genomic REL that would result if the animal had all SNPs called correctly. The current formula for REL_{SNP} is REL_{SNP} = [0.9998(called SNPs) + C(imputed SNPs − called SNPs)]/total loci, where C is the expected percentage of imputed SNPs that have been correctly imputed. The value of C is set to 0.97 for 3K genotypes and to $n_{progeny}/(n_{progeny} + 1)$ for imputed dams because more progeny increase the probability of imputing dam haplotypes correctly.

Previously, imputed dams were evaluated as if genotyped when more than 90% of their haplotypes were imputed, but the new edit is an REL_{SNP} of more than $(0.84)^2$, which provides similar numbers evaluated. Estimates of daughter equivalents (**DE**) that would result from 50K genotyping (**DE$_{MAX}$**) are reduced for imputation (**DE$_G$**) by starting with genomic REL (**REL$_G$**) computed as $REL_{MAX} = DE_{MAX}/(DE_{MAX} + k)$, where k is the the ratio of error to sire variance. That is adjusted downward for REL_{SNP} based on the assumption that REL_{SNP} should be more than 50% to be better than parent average and results in the formula $REL_G = REL_{MAX}[2(REL_{SNP} − 0.5)]$. Finally, REL_G is converted back to DE_G using $DE_G = k[REL_G/(1 − REL_G)]$ and added to DE from traditional sources to obtain final REL. This adjustment for REL_{SNP} differs slightly from that proposed by VanRaden et al., 2010, which was $DE_G = k(REL_{MAX})REL_{SNP}/[1 − (REL_{MAX})REL_{SNP}]$.

Other new chips also require specific edits and imputation. The 50K_V2 chip is very similar to the 50K chip and includes 2,269 additional markers that did not perform well on the original chip. However, 50K_V2 also is missing 1,038 of the original markers, and those markers are imputed. As sufficient numbers of 50K_V2 genotypes accumulate, the additional 2,269 markers could be included with imputation. A high-density (**HD**) chip with 777,000 markers also is available and is being used for research. However, numbers of HD-genotyped animals of each breed are currently too few for routine evaluation. Based on 353 animals with HD genotypes, 632,665 markers met editing criteria for 3K and 50K chips, but only the 38,201 markers that match the 50K chip currently are used. Properties of the HD and 50K markers are very similar, but the new Y chromosome and mitochondrial markers on the HD chip have very low minor allele frequencies.

Actual REL gains from 3K to 50K were confirmed using 287 Holstein animals that were genotyped with the 3K chip in November and then with the 50K_V2 chip in December. Expected standard deviation (**SD**) for change in predicted transmitting ability (**PTA**) was calculated using genetic SD for the trait and the square root of the 6% difference in average reliability between 3K and 50K evaluations. Results for Holsteins were:

Trait	Expected SD for PTA change	Actual SD for PTA change	Mean PTA change (50K − 3K)	Minimum PTA change	Maximum PTA change

Milk, pounds	177	195	−59	−840	415
Fat, pounds	6.6	7.3	−2.6	−24.0	17.0
Protein, pounds	4.7	5.2	−2.3	−16.5	12.6
Productive life, months	0.61	0.62	−0.07	−1.70	2.10
Somatic cell score	0.06	0.05	0.00	−0.16	0.14
Daughter pregnancy rate, %	0.40	0.40	−0.35	−0.93	1.80
Final score	0.25	0.35	0.03	−1.41	0.97
Stature	0.25	0.29	0.05	−0.61	1.09
Strength	0.25	0.27	0.03	−0.95	1.00
Dairy form	0.25	0.32	0.00	−1.21	0.77
Udder height	0.25	0.32	0.03	−0.76	1.11
Udder depth	0.25	0.34	0.05	−0.88	1.01

The SDs may be slightly larger than between interim evaluations because traditional PTAs were different between November and December. Actual SDs for PTA change were consistent with that expected from previous research. Mean changes were also very close to 0. Results may vary for animals with no pedigree information and with number of genotyped relatives. Jersey means and SDs were very similar to those for Holsteins. However, the computed 3K REL for Jerseys was too high and nearly equal to 50K REL because average relationship to the population was overestimated by 1 to 2% for Jerseys with 3K genotypes.

Revised genomic reliabilities for Jerseys

By George Wiggans, Tabatha Cooper, and Paul VanRaden

Adjustments have been applied to yield traits for genomic evaluation of Holsteins and Jerseys since April 2010. Additional research with the Jersey adjustment indicated that reliability gains were less than reported earlier. Thus, official genomic reliabilities for Jersey yield traits in December 2010 were reduced as compared with August evaluations. Average reliability for young animals was reduced from 70 to 62%; average reliability for progeny-tested bulls was reduced from 88 to 86%. That change also affected monthly updates for new animals beginning in September 2010. Other Jersey traits were not affected except that net merit reliability for young animals was reduced from 63 to 58% as a consequence.