Technical note: Characteristics and use of the Illumina BovineLD and GeneSeek Genomic Profiler low-density bead chips for genomic evaluation[1]

G. R. Wiggans,*[2] T. A. Cooper,* C. P. Van Tassell,† T. S. Sonstegard,† and E. B. Simpson‡
*Animal Improvement Programs Laboratory, and
†Bovine Functional Genomics Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705
‡GeneSeek, Lincoln, NE 68521

ABSTRACT

The GeneSeek Genomic Profiler (GGP) BeadChip (GeneSeek, Lincoln, NE), which became available commercially in February 2012, is based on the Illumina BovineLD Genotyping BeadChip (Illumina Inc., San Diego, CA), with 1,745 additional single nucleotide polymorphisms (SNP) for genomic evaluation and SNP for proprietary single-gene tests. The BovineLD chip with 6,909 SNP, which replaced the Illumina Golden-Gate Bovine3K Genotyping BeadChip, has been available since October 2011. The GGP's additional SNP for genomic evaluation were selected to improve imputation by filling SNP gaps on chromosomes and including more Bovine3K SNP than were on the BovineLD chip and to impute microsatellite alleles to facilitate parentage validation. The SNP for single-gene tests were included to minimize the number of separate tests required for those genes, particularly for bulls. The September 2012 US national genomic evaluation included genotypes from BovineLD and GGP chips for 82,510 animals. For those data, BovineLD and GGP performance was similar. The call rate for SNP on these chips that were used in genomic evaluation was 99.6%. The 9 Y-chromosome SNP in common on the BovineLD and GGP chips were highly effective in sex validation (call rate of 99% for males and 0.01% for females). For both chips, the rate of parent–progeny conflicts on a SNP basis (≤0.004%) was similar to that for SNP on the Illumina BovineSNP50 Genotyping BeadChip. Imputation accuracy for 45,187 BovineSNP50 SNP averaged 99.4% for Holsteins. Imputation accuracy was slightly higher for the GGP chip compared with the BovineLD chip because of its additional SNP. Reliability for genomic evaluations using BovineLD and GGP genotypes was 3 percentage points higher than that for Bovine3K genotypes.
Key words: bead chip, BovineLD, GeneSeek Genomic Profiler, genomic evaluation

Technical Note

Genomic evaluation of dairy cattle began in the United States in 2008 (Wiggans et al., 2011). For the first 2 yr, only the BovineSNP50 Genotyping BeadChip (Illumina Inc., 2011b; Illumina Inc., San Diego, CA) was available. The BovineHD Genotyping BeadChip (Illumina Inc., 2010) was introduced in January 2010 but has not been used routinely because of its cost and the small increases in evaluation accuracy that it provides (Harris et al., 2011; VanRaden et al., 2013). The Bovine3K BeadChip (Illumina Inc., 2011c) was introduced in September 2010 (Wiggans et al., 2012) to increase the adoption of genomic testing by reducing cost. The Bovine3K chip used Golden Gate chemistry instead of the Infinium chemistry used in the BovineSNP50 chip. This resulted in slightly less accurate genotyping calls and a higher frequency of genotyping failure. To address those issues, the BovineLD Genotyping BeadChip (Illumina Inc., 2011a) was released in September 2011 with 6,909 SNP and using Infinium chemistry (Boichard et al., 2012). This chip can be customized by adding SNP. In February 2012, GeneSeek (Lincoln, NE) released the GeneSeek Genomic Profiler (**GGP**) BeadChip for Dairy Cattle (Neogen Corporation, 2012). The GGP chip uses the add-on capability to include proprietary single-gene tests as well as 1,745 additional SNP. The additional SNP were included to improve imputation by filling SNP gaps on chromosomes and increasing overlap with Bovine3K SNP, as well as to facilitate imputation of microsatellite alleles for parentage validation (McClure et al., 2012). The SNP for single-gene tests were included to minimize the number of separate tests required for those genes, particularly for bulls. Results for single-gene tests are not included with genotypes provided for genomic

Table 1. Numbers of animals genotyped with the BovineLD (Illumina Inc., San Diego, CA) or GeneSeek Genomic Profiler (GGP; GeneSeek, Lincoln, NE) chips with genotypes usable for genomic evaluation in September 2012 by animal sex, genotyping chip, and breed

Category	Genotypes (no.)		
	Female	Male	All animals
Chip			
BovineLD	54,105	1,115	55,220
GGP	22,470	4,820	27,290
Breed			
Ayrshire	12	10	22
Brown Swiss	414	138	552
Guernsey	0	2	2
Holstein	66,364	5,338	71,702
Jersey	9,785	447	10,232
All animals	76,575	5,935	82,510

evaluation; therefore, those SNP were not included in this study. Characteristics of the GGP chip were examined, and experience with the BovineLD and GGP chips and their performance in genomic evaluation was documented.

In the United States and Canada, monthly genomic evaluations are based on genotypes from GeneSeek, DNA LandMarks (Saint-Jean-sur-Richelieu, QC, Canada), Genetic Visions (Middleton, WI), and Pfizer Animal Genetics (Kalamazoo, MI). From January to August 2012, an average of 11,400 genotypes was added to the genotype database for dairy cattle at the US Department of Agriculture (**USDA**) each month. Usable BovineLD and GGP genotypes for 82,510 animals were available for the September 2012 genomic evaluation (Table 1). Nearly 93% of those genotypes were for females, but the proportion of males was higher for GGP genotypes (18%) than for BovineLD genotypes (2%). The majority of animals genotyped with these chips were Holsteins (87%) followed by Jerseys (12%), Brown Swiss (<1%), Ayrshires (<1%), and Guernseys (<1%). Although the Ayrshire, Guernsey, and Milking Shorthorn breeds are not currently genomically evaluated, their genotypes can be used for parentage validation.

Table 2 shows the distribution of SNP by chromosome group for the BovineLD and GGP chips. The GGP chip includes all BovineLD SNP. The 108 duplicate SNP on the GGP chip resulted from including multiple copies of SNP for single-gene tests and microsatellite imputation to ensure that at least one would survive losses during manufacture. Table 3 shows the increase in overlap with Bovine3K SNP achieved with the GGP chip and the inclusion of SNP from the BovineHD chip to reduce the size of SNP gaps on chromosomes. The GGP chip has 1,195 more SNP used in genomic evaluation than the BovineLD chip. The 486 GGP SNP that were unique to the BovineHD chip are not currently used for genomic evaluation but will likely be included when the set of SNP used is modified.

Call Rate

Call rate on a SNP basis was assessed for the BovineLD and GGP chips. The Y SNP are used for sex validation and are expected to have no calls for females; therefore, call rates for Y SNP were calculated separately by animal sex. Only genotypes from laboratories that genotype both males and females were included in

Table 2. Distribution of SNP by chromosome by genotyping chip

SNP group	Illumina BovineLD[1]	GeneSeek Genomic Profiler[2]
Autosomal	6,651	8,291
X pseudoautosomal region	17	17
X-specific	219	268
Y	9	19
Mitochondrial	13	13
New[3]	—	46
Duplicates	—	108
All SNP	6,909	8,762

[1]Illumina Inc. (San Diego, CA).

[2]GeneSeek (Lincoln, NE).

[3]Not on any other Illumina genotyping chip.

Table 3. Numbers of SNP that overlap for the BovineLD (Illumina Inc., San Diego, CA) or GeneSeek Genomic Profiler (GGP; GeneSeek, Lincoln, NE) chips with other Illumina chips and with SNP used for genomic evaluation

Chip	BovineLD	GGP
Illumina BovineSNP50 v2	6,854	8,029
Illumina Bovine3K	2,159	2,677
Illumina BovineHD unique SNP	17	486
All Illumina BovineHD	6,844	8,334
All Illumina arrays	6,909	8,608
SNP used for genomic evaluation	6,836	8,031

the call rates for Y SNP to ensure that call rates for females were reliable. Although the GGP chip includes 19 Y SNP, only the 9 in common with the BovineLD chip were studied.

Mean call rate of SNP on autosomal and X chromosomes for each chip was >99.5% (Table 4). A few SNP had call rates below the 90% threshold used to select SNP originally for genomic evaluation (Wiggans et al., 2011). With each submission, the laboratories are returned a list of SNP with low call rates and have an opportunity to adjust clustering. If the call rate for those few SNP does not improve, they probably will be excluded the next time that SNP are selected for genomic evaluation. For Y SNP, the high call rate for males and nearly zero call rate for females indicate that those SNP provide excellent accuracy for sex validation and confirm the results of Boichard et al. (2012).

Conflicts

Conflicts between sire and progeny SNP genotypes were analyzed for 71,299 Holsteins with genotypes that were received by USDA through August 2012 and with a confirmed sire–progeny relationship. Confirmation limits for SNP conflicts were those used for official USDA genomic evaluations (Wiggans et al., 2012). The category of unreliable genotype, which includes genotypes with more SNP conflicts than acceptable for a true parent–progeny relationship but fewer than needed to declare a parent–progeny conflict (Wiggans

et al., 2012), was retained but rarely assigned with the BovineLD or GGP chips. For both chips, the mean rate of sire–progeny conflicts (Table 5) was very low ($\leq0.004\%$). The SNP with the most conflicts ($<1\%$) was from the GGP chip. Fewer than 1.6% of SNP had $>0.05\%$ conflicts (data not shown). Conflicting SNP genotypes were designated as missing before imputation.

Minor Allele Frequency

The Holstein minor allele frequency (**MAF**) distribution for SNP used in genomic evaluation was determined for SNP in common on the BovineLD and GGP chips and for those unique to the GGP chip (Figure 1). For SNP in common, the percentage of SNP increased with MAF. That trend also was found for SNP unique to the GGP chip. However, the percentage of SNP with an MAF of ≥0.32 was lower for GGP-only SNP than for SNP in common because of the varying objectives considered in selecting the GGP-only SNP.

Imputation Accuracy

Imputation accuracy was determined by comparison of genotypes imputed from BovineLD or GGP chips with actual BovineSNP50 genotypes for 1,314 Holsteins that had been genotyped at both densities and had either both parents or just sire genotyped. With the widespread use of the Bovine3K and BovineLD chips

Table 4. Means, SD, minimums, and maximums for call rate for SNP by chromosome type, animal sex, and genotyping chip

Chromosome type	Animal sex	Chip	SNP (no.)	Animals (no.)	Call rate (%)			
					Mean	SD	Minimum	Maximum
Autosomal and X	All animals	Illumina BovineLD[1]	6,836	55,220	99.57	1.09	69.33	99.99
		GeneSeek Genomic Profiler[2]	8,031	27,290	99.71	0.55	92.46	100.00
Y	Male	Illumina BovineLD	9	1,126	99.00	—	—	—
		GeneSeek Genomic Profiler	9	4,820	99.50	—	—	—
	Female	Illumina BovineLD	9	18,103	0.01	—	—	—
		GeneSeek Genomic Profiler	9	22,453	0.01	—	—	—

[1]Illumina Inc. (San Diego, CA).
[2]GeneSeek (Lincoln, NE).

Table 5. Mean percentage of parent-progeny conflicts for Holsteins on a SNP basis for genotypes received by the US Department of Agriculture through August 2012 by genotyping chip

Chip	SNP (no.)	Animals (no.)	SNP conflicts (%)	
			Mean	Maximum
Illumina BovineLD[1]	6,836	50,164	0.004	0.55
GeneSeek Genomic Profiler[2]	8,031	21,135	0.003	0.87

[1]Illumina Inc. (San Diego, CA).

[2]GeneSeek (Lincoln, NE).

for females, dam genotype was often from one of those low-density chips. For animals with just the sire genotyped, the maternal grandsire had also been genotyped in most cases. Imputation was done using version 2 of the Fortran program findhap.f90 (VanRaden, 2011). Boichard et al. (2012) reported that the BovineLD chip increased imputation accuracy compared with the Bovine3K chip. Overall imputation accuracy for 45,187 BovineSNP50 SNP (Table 6) averaged 99.4%. Mean imputation accuracy was slightly higher for the GGP chip than for the BovineLD chip (0.2 percentage points with both parents genotyped and 0.4 percentage points with only sire genotyped). However, different animals were genotyped with the 2 chips; therefore, the advantage of the GGP chip for imputation accuracy compared with the BovineLD chip was not proven.

Age Distribution

The distribution of animal age on the date that the genotype was received was determined by sex for animals that were <3 yr old. Most animals were genotyped when they were 1 or 2 mo old (Figure 2). The concentration at early ages was greatest for males, probably reflecting the competition among AI organizations to obtain the best genetics and the desire of bull owners to make early culling decisions on calves that will not be marketable.

Reliability

Reliabilities for genomic evaluations based on BovineLD and GGP genotypes were calculated as described by VanRaden et al. (2011b), with a modification to account for differing reliabilities from various chips (VanRaden et al., 2011a). That modification automatically accounts for the effects of number of called SNP and number of highest density genotypes (i.e., those including nearly all SNP used in the analysis) on imputation accuracy and evaluation reliability. Before the BovineLD and GGP chips were available, the fraction of correctly imputed SNP was previously assumed to be constant for all animals with Bovine3K genotypes. The percentage of SNP that are expected to have been correctly imputed (imputation accuracy) for an animal is now calculated as

$$[n_{call}/(n_{call} + 120)] \times [n_{BovineSNP50}/(n_{BovineSNP50} + 50)],$$

where n_{call} is the number of called SNP and $n_{BovineSNP50}$ is the number of BovineSNP50 genotypes from which to impute for an animal's breed. The first part of the formula represents the reliability of imputation for the chip density used to genotype the animal, and the last part represents the maximum imputation accuracy possible based on the number of BovineSNP50 genotypes. Both reliabilities are analogous to the $n/(n + k)$ formula in the USDA animal model for traditional bull reliability, where n is the number of daughter equivalents and k is a variance ratio (VanRaden and Wiggans, 1991). The constants 120 and 50 were selected to make the expected imputation accuracy agree with the imputation accuracy observed for chips of various marker densities and various reference populations. A

Table 6. Means, SD, minimums, and maximums for accuracy of imputation to a BovineSNP50 basis (45,187 markers) by genotyping chip and availability of parent information

Chip	Parent information available	Imputation accuracy (%)			
		Mean	SD	Minimum	Maximum
Illumina BovineLD[1]	Sire and dam (n = 414)	99.3	0.30	95.8	99.7
	Sire only (n = 48)	98.9	0.60	96.1	99.5
GeneSeek Genomic Profiler[2]	Sire and dam (n = 773)	99.5	0.18	98.2	99.8
	Sire only (n = 79)	99.3	0.42	96.8	99.8
Both	All (n = 1,314)	99.4	0.29	95.8	99.8

[1]Illumina Inc. (San Diego, CA).
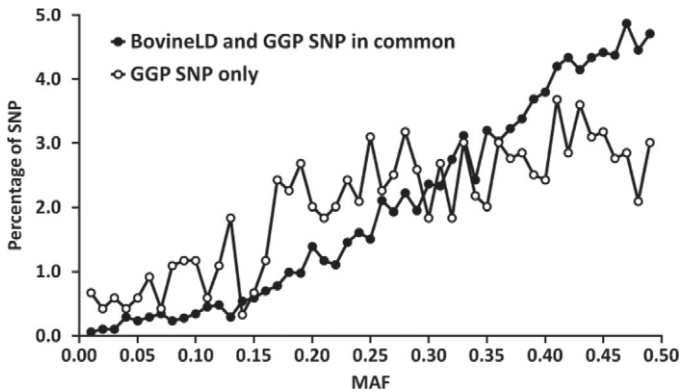
[2]GeneSeek (Lincoln, NE).

Figure 1. Minor allele frequency (MAF) distribution of SNP usable for Holstein genomic evaluations for SNP common to the Illumina BovineLD (Illumina Inc., San Diego, CA) and GeneSeek Genomic Profiler (GGP; GeneSeek, Lincoln, NE) BeadChips and for SNP unique to the GGP chip.
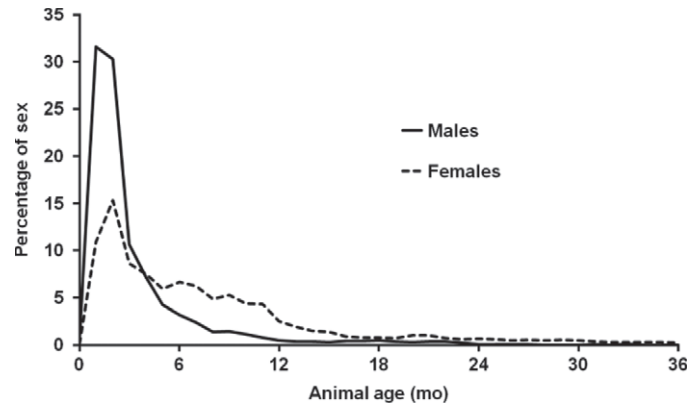

Figure 2. Percentages of males and females with low-density (BovineLD, Illumina Inc., San Diego, CA; GeneSeek Genomic Profiler, GeneSeek, Lincoln, NE) genotypes by animal age when the genotype was received by the US Department of Agriculture for animals that were <3 yr old.

possible improvement in the formula would be to account for whether the parents had been genotyped. For example, with 6,800 called markers for a Jersey from a population with 6,000 Jersey BovineSNP50 genotypes, the percentage of SNP that are expected to have been correctly imputed is 97.5%. When observed SNP are also included, overall accuracy is the mean of correctly imputed SNP and observed SNP weighted by number of each type of SNP, which for the example is [38,387 (0.975) + 6,800]/45,187, or 97.9%.

Mean reliabilities of USDA September 2012 genomic evaluations were calculated for Holsteins born from September 2010 through August 2012. The effect of chip type on reliability of genomic evaluations is shown in Table 7. Reliability was highest (75.1%) for the BovineSNP50 chip for both males and females. Reliabilities for the BovineLD and GGP chips were intermediate (71.4 to 73.1%) between the Bovine3K (69.2 to 70.1%) and BovineSNP50 chips. Although the GGP chip was

expected to provide slightly higher reliability than the BovineLD chip, the lower reliability of parent average for GGP bulls (25.5%) compared with all other chips (30.8 to 35.0%) caused genomic reliability to be lower for the GGP chip than for the BovineLD chip.

The BovineLD and GGP chips have received widespread use and deliver improved accuracy over the Bovine3K chip. The GGP chip has been used to a greater extent than the BovineLD chip for bulls because it provides single-gene tests. The BovineSNP50 chip remains the industry standard, but genotypes from the BovineLD and GGP chips result in genomic evaluations of only slightly lower accuracy at approximately half the genotyping cost.

ACKNOWLEDGMENTS

Table 7. Reliabilities of US Department of Agriculture September 2012 parent average and genomic PTA for Holsteins born from September 2010 through August 2012 by genotyping chip and animal sex

			Reliability (%)	
Animal sex	Chip	Animals (no.)	Parent average	Genomic PTA
Female	Illumina Bovine3K[1]	20,388	31.7	69.2
	Illumina BovineLD[1]	42,075	26.7	71.4
	GeneSeek Genomic Profiler[2]	14,461	28.6	72.6
	Illumina BovineSNP50[1]	14,927	30.9	75.1
Male	Illumina Bovine3K	1,476	35.0	70.1
	Illumina BovineLD	724	31.3	73.1
	GeneSeek Genomic Profiler	3,508	25.5	72.1
	Illumina BovineSNP50	21,352	30.8	75.1

[1]Illumina Inc. (San Diego, CA).

[2]GeneSeek (Lincoln, NE).

vice, USDA, Beltsville, MD) for technical review and assistance in manuscript preparation.

REFERENCES

Boichard, D., H. Chung, R. Cassonneville, X. David, A. Eggen, S. Fritz, K. J. Gietzen, B. J. Hayes, C. T. Lawley, T. S. Sonstegard, C. P. Van Tassell, P. M. VanRaden, K. Viaud-Martinez, and G. R. Wiggans. 2012. Design of a bovine low-density SNP array optimized for imputation. PLoS ONE 7:e34130.

Harris, B. L., A. M. Creagh, A. M. Winkelman, and D. L. Johnson. 2011. Experiences with the Illumina high density bovine beadchip. Interbull Bull. 44:3–7.

Illumina Inc. 2010. BovineHD Genotyping BeadChip. Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf.

Illumina Inc. 2011a. BovineLD Genotyping BeadChip. Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovineLD.pdf.

Illumina Inc. 2011b. BovineSNP50 Genotyping BeadChip. Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp5O.pdf.

Illumina Inc. 2011c. GoldenGate Bovine3K Genotyping BeadChip. Accessed Aug. 31, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine3K.pdf.

McClure, M., T. Sonstegard, G. Wiggans, and C. P. Van Tassell. 2012. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. Front. Genet. 3:140.

Neogen Corporation. 2012. GeneSeek Genomic Profiler for Dairy Cattle. Accessed Aug. 31, 2012. http://www.neogen.com/GeneSeek/pdf/Catalogs/DairyGenomicProfiler.pdf.

VanRaden, P. 2011. findhap.f90. Accessed Sept. 17, 2012. http://aipl.arsusda.gov/software/findhap/.

VanRaden, P., G. Wiggans, T. Cooper, C. Van Tassell, and T. Sonstegard. 2011a. Genomic evaluations using 6,909 markers. Changes to Evaluation System (December 2011). Accessed Sept. 18, 2012. http://aipl.arsusda.gov/reference/changes/eval1112.htm.

VanRaden, P. M., D. J. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A. Faust, and G. A. Doak. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. J. Dairy Sci. 96:668–678.

VanRaden, P. M., J. R. O'Connell, G. R. Wiggans, and K. A. Weigel. 2011b. Genomic evaluations with many more genotypes. Genet. Sel. Evol. 43:10.

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737–2746.

Wiggans, G. R., T. A. Cooper, P. M. VanRaden, K. M. Olson, and M. E. Tooker. 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. J. Dairy Sci. 95:1552–1558.

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94:3202–3211.