J. Dairy Sci. 94:6116–6121
doi:10.3168/jds.2011-4764


# *Hot topic:* Performance of bovine high-density genotyping platforms in Holsteins and Jerseys

**G. Rincon,* K. L. Weber,* A. L. Van Eenennaam,* B. L. Golden,† and J. F. Medrano*[1]**
*Department of Animal Science, University of California, Davis 95616
†Dairy Science Department, California Polytechnic State University, San Luis Obispo 93407

## ABSTRACT

Two high-density single nucleotide polymorphism (SNP) genotyping arrays have recently become available for bovine genomic analyses, the Illumina High-Density Bovine BeadChip Array (777,962 SNP) and the Affymetrix Axiom Genome-Wide BOS 1 Array (648,874 SNP). These products each have unique design and chemistry attributes, and the extent of marker overlap and their potential utility for quantitative trait loci fine mapping, detection of copy number variation, and multibreed genomic selection are of significant interest to the cattle community. This is the first study to compare the performance of these 2 arrays. Deoxyribonucleic acid samples from 16 dairy cattle (10 Holstein, 6 Jersey) were used for the comparison. An independent set of DNA samples taken from 46 Jersey cattle and 18 Holstein cattle were used to ascertain the amount of SNP variation accounted by the 16 experimental samples. Data were analyzed with SVS7 software (Golden Helix Inc., Bozeman, MT) to remove SNP having a call rate less than 90%, and linkage disequilibrium pruning was used to remove linked SNP ($r^2 \geq 0.9$). Maximum, average, and median gaps were calculated for each analysis based on genomic position of SNP on the bovine UMD3.1 genome assembly. All samples were successfully genotyped (≥98% SNP genotyped) with both platforms. The average number of genotyped SNP in the Illumina platform was 775,681 and 637,249 for the Affymetrix platform. Based on genomic position, a total of 107,896 SNP were shared between the 2 platforms; however, based on genotype concordance, only 96,031 SNP had complete concordance at these loci. Both Affymetrix BOS 1 and Illumina BovineHD genotyping platforms are well designed and provide high-quality genotypes and similar coverage of informative SNP. Despite fewer total SNP on BOS 1, 19% more SNP remained after linkage disequilibrium pruning, resulting in a smaller gap size (5.2 vs. 6.9 kb) in Holstein and Jersey samples relative to BovineHD. However, only 224,115 Illumina and 241,038 Affymetrix SNP remained following removal of SNP with a minor allele frequency of zero in Holstein and Jersey samples, resulting in an average gap size of 11,887 bp and 11,018 bp, respectively. Combining the 354,348 informative ($r^2 \geq 0.9$), polymorphic (minor allele frequency $\geq 0$), unique SNP data from both platforms decreased the average gap size to 7,560 bp. Genome-wide copy number variant analyses were performed using intensity files from both platforms. The BovineHD platform provided an advantage to the copy number variant data compared with the BOS 1 because of the larger number of SNP, higher intensity signals, and lower background effects. The combined use of both platforms significantly improved coverage over either platform alone and decreased the gap size between SNP, providing a valuable tool for fine mapping quantitative trait loci and multibreed animal evaluation.

**Key words:** genotyping platforms, single nucleotide polymorphism, dairy cattle

## Hot Topic

The Bovine SNP50 (>50,000 probes) genotyping array (Illumina Inc., San Diego, CA) first became available in 2007 (Matukumalli et al., 2009). Tens of thousands of animals have since been genotyped using this platform, and much of those data have been used to realize the envisioned promise of genomic selection (Meuwissen et al., 2001) in dairy cattle. This selection approach is based on the simultaneous selection of many thousands of genetic markers that cover the entire genome. The success of genomic selection is based on exploitation of linkage disequilibrium (**LD**) between the markers and the QTL affecting a trait, and both associations and linkage phase are assumed to persist across the population. The first dairy cattle evaluations, where estimates of genetic merit included SNP50 genomic information, were officially released in the United States in early 2009 (VanRaden et al., 2009). The Bovine SNP50 array data have also been successfully used to detect copy number variants (**CNV**), which have been implicated in both disease phenotypes and the normal phenotypic

Received July 26, 2011.
Accepted September 21, 2011.
[1]Corresponding author: jfmedrano@ucdavis.edu

variation associated with quantitative traits (Hou et al., 2011).

The usefulness of Bovine SNP50 data for the development of genetic merit estimates across breeds has proven to be limited (Hayes et al., 2009). In cattle, it has been estimated that markers need to be spaced less than 10 kb apart to show consistent LD phase across breeds (de Roos et al., 2008). The Bovine SNP50 genotyping array markers are spaced at an average of approximately 50-kb intervals. Two higher-density genotyping arrays have recently become available to the bovine genomics community. Illumina Inc. released its High-Density Bovine BeadChip (**BovineHD**) array product using its Infinium HD assay in January, 2010 (Matukumalli et al., 2011), and Affymetrix Inc. (Santa Clara, CA) released its Axiom Genome-Wide BOS 1 Array (**BOS 1**) in early 2011. The availability of these very high-density arrays opens up the possibility of combining data from multiple *Bos taurus* breeds to improve the accuracy of genomic predictions. The benefits of higher-density platforms will likely be advantageous for other applications as well, such as the analysis of CNV. These 2 products have unique design and chemistry attributes, and the overlap of markers and utility of these 2 products for different applications is a topic of interest to the cattle research community. The objective of this study was to compare the results from the 2 high-density chips when using DNA samples derived from Holstein and Jersey cattle.

DNA samples from 16 dairy cattle (10 Holstein and 6 Jersey) derived from five 3-generation pedigrees were used as the basis to perform a comparison between the 2 available bovine high-density genotyping platforms: BovineHD and BOS 1. High-quality DNA was extracted from blood samples using the Gentra Puregene Blood DNA Purification Kit from Qiagen Benelux BV (Venlo, Netherlands). BovineHD BeadChip genotypes from an independent set of DNA samples taken from 46 Jersey cattle and 18 Holstein cattle were used to ascertain the amount of SNP variation accounted for by the 16 samples that were genotyped using both high-density genotyping platforms. These samples were also used as a control in an association study conducted to determine differences in copy number variation between Holstein and Jersey cattle. Samples were genotyped using the BovineHD BeadChip (Illumina Inc.) and the Axiom Genome-Wide BOS 1 Array (Affymetrix Inc.) by GeneSeek (Lincoln, NE).

Genotyping files from both genotyping platforms were obtained (.idat files for Illumina and .cel files for Affymetrix) to undertake the analyses. The .idat files were analyzed with Genome Studio software (Illumina Inc.) to perform the quality control (**QC**) analysis and to extract genotype calls, error rates, P50 GenCall score, and signal intensity files for each SNP (expressed as $\log_2$ ratios). The .cel files were analyzed with Genotyping Console 4.1 (Affymetrix Inc.) to perform QC and to obtain genotype calls and A and B intensity values.

Basic genotype statistics for each marker, including call rate, minor allele frequency (**MAF**), and allele and genotype counts were calculated using the Quality Assurance Module from SNP Variation Suite version 7 (SVS7; Golden Helix Inc., Bozeman, Montana). The data were analyzed both with and without QC criteria. Quality control criteria (filters) were used to remove from further analysis any SNP having less than a 90% overall amplification and any SNP that were in LD with each other (defined as $r^2 \geq 0.9$), retaining 1 SNP with high correlation from each set. The term $r^2$ is used to describe the correlation between SNP and to define the level of LD between markers. Linkage disequilibrium pruning was performed to develop a comparative analysis including only SNP that were assorting independently in Holstein and Jersey samples. Maximum, average, and median gaps were calculated for each analysis based on genomic position of SNP provided by the companies on the Bovine UMD3.1 genome assembly (GenBank; http://www.cbcb.umd.edu/research/bos_taurus_assembly.shtml).

Analyses were performed to define regions of CNV on a genome-wide scale, using a univariate analysis on a per-sample basis with the copy number analysis module (CNAM) from SVS7. The signal intensity files for each SNP (expressed as $\log_2$ ratios) and the genetic marker map were downloaded with a custom SVS7 script from Illumina Genome Studio for the BovineHD platform. For the BOS 1 platform, the $\log_2$ ratio values were calculated in 2 steps: first, a reference was developed for each marker considering the formula T = A + B, where A and B are signal intensity values obtained from the "AxiomGT1.summary.txt" output using the APT Affymetrix software. For each SNP, a reference value M = median (T_sample1, T_sample2, . . . T_sampleN) was also determined. The second step consisted in estimating the intensities for each individual sample as the $\log_2$(T/M) ratio.

The $\log_2$ R intensity files from both genotyping platforms were then loaded into SVS7 software to perform the CNV analysis. Numeric principal component analysis was performed on the intensity data to correct for error/chip variation for each sample. Variation in hybridization intensity or waviness, which can prevent accurate CNV inference and has been shown to correlate with the percent guanine-cytosine along the genomic DNA (genomic GC content), was corrected using the method described in Diskin et al. (2008). Copy number variant segments were defined using a moving window

of 5,000 SNP, with 20 segments per window and a minimum of 1 SNP per segment. A linear regression was performed to test whether CNV segments differed between Holstein and Jersey samples.

All samples included in this study were successfully genotyped (>98% SNP genotyped) with both BovineHD and BOS 1 platforms. The average number of genotyped SNP in the BovineHD platform was 775,681 and for the BOS 1 platform was 637,249. These numbers correspond to an average call rate of 99.7 and 98.5% of the total number of SNP included in the Illumina and Affymetrix HD chips, respectively.

Genotypes from the BovineHD platform for 64 additional independent samples were used as a benchmark to determine the proportion of polymorphic SNP represented in the 16 samples genotyped using both high-density genotyping platforms. The number of polymorphic BovineHD SNP in the 64 benchmark samples was 647,273, whereas in the 16 samples genotyped on both platforms it was 611,430. These results show that 94.4% of the SNP variation observed in Holsteins and Jerseys in the benchmark sample set was represented in the subset of 16 samples used for the HD chip comparison study, suggesting that the majority of Holstein and Jersey SNP variation was included in the smaller sample set.

Table 1 shows the number of SNP genotyped in both platforms and the genomic coverage after removing SNP with less than a 90% call rate. In our samples, 99.4 and 95.1% of the SNP had a call rate greater than 0.9 in the BovineHD and BOS 1 platforms, respectively. Gap size represents the distance between adjacent SNP and is a measurement of genome coverage. Gap size was slightly larger in the BOS 1 platform due to a decreased number of total SNP. Both genotyping platforms were developed by selecting SNP that would represent the wide genetic diversity of cattle breeds, including *B. taurus* and *Bos indicus*. Therefore, a large proportion of SNP having MAF equal to zero was expected when analyzing only 2 breeds. Excluding SNP with call rate below the 90% threshold, the number of SNP that met the second criteria (MAF > 0) and the corresponding estimate of genome coverage where both criteria were met are provided in Table 1b.

To examine specific breed effects, we compared the proportion of SNP with MAF = 0 between Holsteins and Jerseys using both genotyping platforms. With the Bovine HD, the number of alleles with MAF = 0 was slightly higher (5,671 SNP) in Jerseys, whereas with the BOS 1 platforms, Jerseys had a much larger number (57,277) of SNP with MAF = 0. Among the total SNP that were monomorphic in at least 1 breed, the percentage of monomorphic SNP in both breeds was 72 and 68%, and the percentage of monomorphic SNP in one breed and polymorphic in the other breed was 28 and 32% with Bovine HD and BOS 1, respectively. Both platforms showed decreased SNP variation in Jerseys compared with Holsteins.

When 2 SNP are in LD, their genotypic information is redundant, and only 1 is necessary to describe the variation for that region of the genome. With the high genome coverage of both HD chips, it was important to ascertain the number of unlinked SNP contributing useful information for association studies. Therefore, a common approach that is used when analyzing genotyping data are to apply a function called LD pruning. This approach essentially removes redundant markers based on a defined correlation coefficient ($r^2$) LD value, and provides an estimate of the number of unlinked SNP available for use in association studies. The results of LD pruning, defined as $r^2 \geq 0.9$ across Holstein and Jersey samples, are provided in Table 2, for the loci meeting QC standards for call rate (a) and for both call rate and MAF (b).

These analyses revealed important differences between the Illumina and the Affymetrix platforms. In spite of the lower number of SNP in the Affymetrix platform, after removal of SNP that did not meet minimum call rate standards and LD pruning, 19% more SNP assorted independently in the Affymetrix marker set, providing a smaller average gap size (11,018 bp) relative to the Illumina marker set (11,887 bp). The average gap size when considering the 2 platforms simultaneously was 5,073 bp. The difference between the 2 platforms was similar once the SNP with MAF = 0 in Holstein and Jerseys were removed (Table 2b). The final number of SNP in Table 2b is the effective number of independently assorting SNP from both platforms in our Holstein and Jersey data set. Approximately 7% more informative SNP were present in the Affymetrix platform relative to the Illumina platform. The same LD criteria ($r^2 \geq 0.9$) were applied to 64 independent Holstein and Jersey samples to determine the extent of LD captured by the 16 experimental samples included in this study. Interestingly, 95.8% of the LD regions observed in the larger population were also captured with the 16 samples, providing support that the comparisons made in this study are representative of the LD observed in both breeds.

Marker position on UMD3.1 and genotype concordance were considered to determine the number of SNP that were shared between the 2 genotyping platforms. A total of 96,031 SNP were shared between platforms, including 49,345 SNP in the BovineHD and 47,741 SNP in the BOS 1 chip from the widely used Bovine Illumina SNP50 genotyping platform. The average SNP concordance in the shared SNP for Holstein and Jersey samples was 99.9%, including 74,259 SNP, which had a concordance of 1 and the remaining SNP had a concor-

**Table 1.** Number of SNP genotyped and genomic coverage obtained using both high-density (HD) genotyping platforms after removing (a) SNP with call rate less than 0.9, and (b) removing SNP with call rate less than 0.9 and minor allele frequency (MAF) = 0

| Item | Total SNP | Removed SNP | Final SNP | Max gap (kb) | Average gap (bp) | Median gap (bp) |
|---|---|---|---|---|---|---|
| (a) Removing SNP with call rate <0.9 | | | | | | |
| Infinium BovineHD[1] | 777,962 | 4,792 | 773,170 | 1,080.181 | 3,449 | 2,686 |
| Axiom Genome-Wide BOS 1[2] | 648,874 | 31,967 | 616,907 | 1,120.175 | 4,310 | 2,490 |
| Merged platforms | 1,426,836 | 36,759 | 1,390,077 | 1,080.181 | 2,066 | 1,365 |
| (b) Removing SNP with call rate <0.9 and MAF = 0 | | | | | | |
| Infinium BovineHD | 773,170 | 161,740 | 611,430 | 1,514.901 | 4,353 | 2,551 |
| Axiom Genome-Wide BOS 1 | 616,907 | 229,205 | 387,702 | 1,127.515 | 6,852 | 3,703 |
| Merged platforms | 1,390,077 | 390,945 | 999,132 | 1,080.181 | 2,908 | 1,609 |

[1]High-Density Bovine BeadChip (Illumina Inc., San Diego, CA).

[2]Affymetrix Inc., Santa Clara, CA.

**Table 2.** Number of SNP and genomic coverage using both high-density (HD) genotyping platforms after linkage disequilibrium (LD) pruning of SNP with correlation coefficient ($r^2$) $\geq$ 0.9 and removing (a) SNP with call rate less than 0.9 and (b) SNP with call rate less than 0.9 and minor allele frequency (MAF) = 0

| Item | Total SNP | Removed SNP $r^2 \geq 0.9$ | Final SNP | Max gap (kb) | Average gap (bp) | Median gap (bp) |
|---|---|---|---|---|---|---|
| (a) LD pruning and removal of SNP with call rate <0.9 | | | | | | |
| Infinium BovineHD[1] | 773,170 | 384,907 | 388,263 | 1,090.890 | 6,881 | 4,577 |
| Axiom Genome-Wide BOS 1[2] | 616,907 | 136,711 | 480,196 | 1,123.527 | 5,159 | 3,026 |
| Merged platforms | 1,390,077 | 848,199 | 541,878 | 1,090.890 | 5,073 | 3,137 |
| (b) LD pruning and removal of SNP with call rate <0.9 and MAF = 0 | | | | | | |
| Infinium BovineHD | 611,430 | 387,315 | 224,115 | 1,611.522 | 11,887 | 6,693 |
| Axiom Genome-Wide BOS 1 | 387,702 | 146,664 | 241,038 | 1,297.304 | 11,018 | 5,749 |
| Merged platforms | 999,132 | 644,738 | 354,348 | 1,297.304 | 7,560 | 3,929 |

[1]High-Density Bovine BeadChip (Illumina Inc., San Diego, CA).

[2]Affymetrix Inc., Santa Clara, CA.

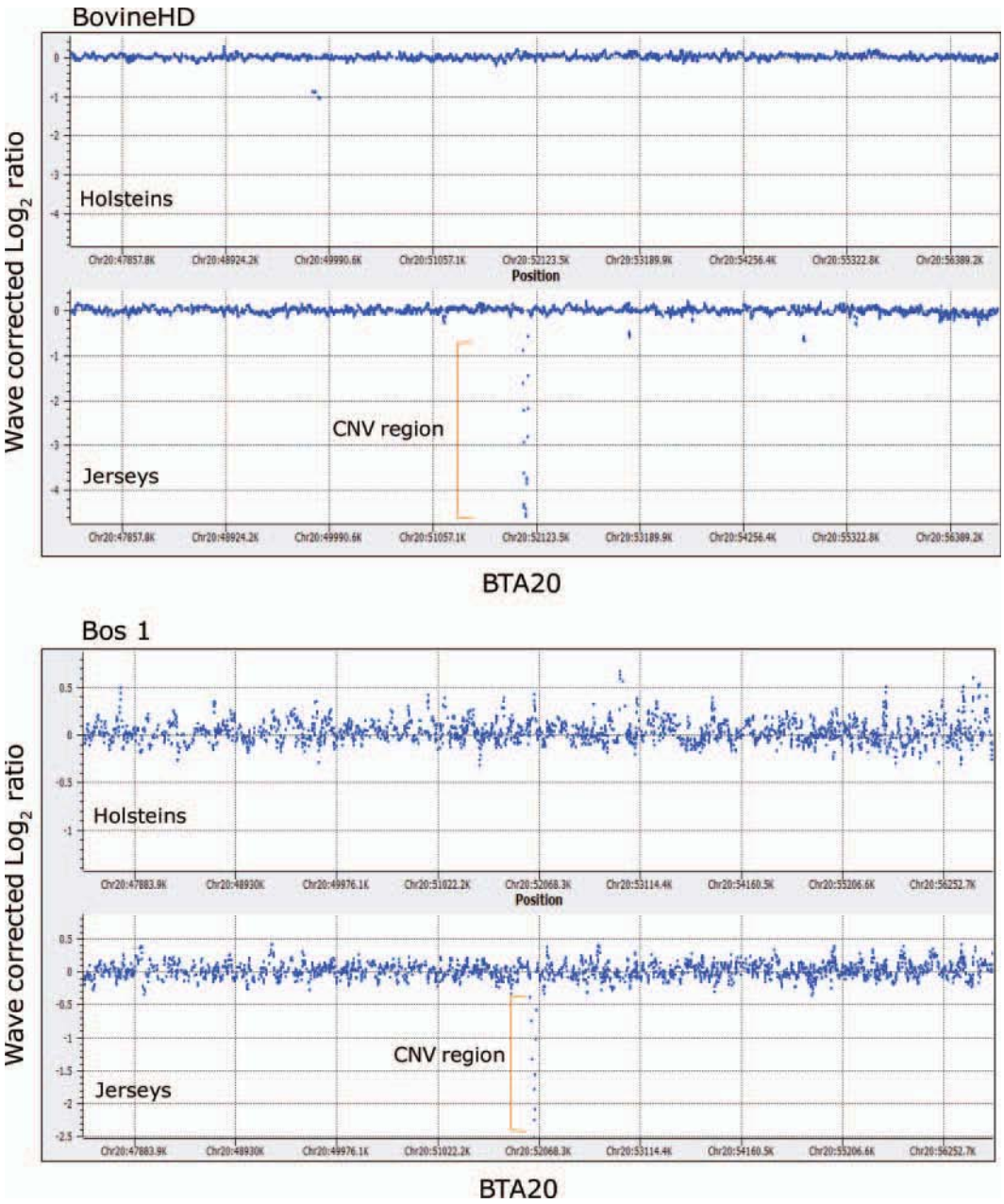


**Figure 1.** Example of a single sample showing a copy number variant (CNV) region on BTA20 with intensity differences between Holsteins and Jerseys observed in High-Density Bovine BeadChip (BovineHD; Illumina Inc., San Diego, CA) and Axiom Genome-Wide BOS 1 Array (BOS 1; Affymetrix Inc., Santa Clara, CA). Horizontal blue dots represent the wave-corrected log R ratios for each SNP along the chromosome. Note the higher dispersion in BOS 1. The lower panels show a vertical set of dots representing SNP with a lower intensity and, consequently, a CNV region divergent between Holstein and Jersey. Color version available in the online PDF.

dance ranging between 0.94 to 0.99. When considering the number of SNP shared between platforms based only on chromosomal location, a larger number of SNP (107,896) were in common; however, 11,865 of these SNP showed a very low genotype concordance. Of the 11,865 SNP, 6,713 SNP were reversed (allele in the forward strand showed one base in BovineHD and the complementary base in BOS 1) and 5,152 SNP genotypes failed in 1 of the 2 platforms as evidenced by very low call rates (<50%) and, therefore, very low SNP concordance. To further examine the reversed SNP, the SNP genotypes were recoded on the Affymetrix BOS1 platform and concordance analysis was performed for the 6,713 reversed SNP. Very high concordance (>85%) was observed for 5,237 SNP. The remaining 1,476 reversed SNP showed very low concordance (<50%), suggesting that these SNP are not correctly annotated in the BOS 1 genotyping platform.

If both platforms were combined and SNP were pruned using the call rate (MAF) and LD criteria described previously, the net number of informative SNP would total 354,348 (Table 2). The combination of SNP calls from both platforms provided a substantial decrease in the average gap size (7.56 kb) in comparison to using either platform alone (Illumina, 11.9 kb versus Affymetrix, 11.0 kb). Based on analyses of Holstein, Jersey, and Angus cattle, it was estimated that an average gap size of less than 10 kb would be required to obtain the consistent marker-QTL LD needed for effective across-breed, marker-assisted selection (de Roos et al., 2008). Our findings suggest that this threshold could not be reached between Holstein and Jersey cattle using only 1 of the high-density platforms tested here. More specifically, the application of these platforms for fine mapping and association studies in Holstein and Jersey cattle breeds may be confounded by the large proportion of SNP on each platform that were monomorphic and redundant due to LD in these 2 breeds. It has yet to be determined if a sufficient level of informative polymorphic SNP variation exists in these platforms for the analysis of more genetically distant cattle breeds, such as *B. taurus* and *B. indicus* cattle. In has been estimated that several million SNP may be required to obtain consistent linkage phase across these subspecies (Goddard and Hayes, 2009).

Copy number variation data were obtained from both BovineHD and BOS 1 platforms. The analyses revealed overlapping of CNV regions between platforms but also distinct CNV patterns in both Bovine HD and BOS 1 data. The association study, performed to detect CNV segments that were significantly different between Holsteins and Jerseys ($P < 0.01$), revealed 5 CNV regions located on BTA7, BTA10, BTA20, BTA24, and BTA27 common to both platforms; 3 regions on BTA1, BTA8, and BTA16 that were observed only in BOS 1; and 3 regions on BTA5, BTA15, and BTA23 that were observed only in BovineHD. Figure 1 shows an example of a CNV region detected with both platforms on BTA20, which was different between Holsteins and Jerseys. The regions where a different CNV pattern was observed between the 2 platforms corresponded to regions where no overlapping SNP occurred. It is important to note that for CNV analysis, the Illumina BovineHD platform provided a larger marker density with higher $\log_2$ ratio ranges and considerably less background noise, thereby making CNV analyses more definitive (Figure 1).

Both Affymetrix BOS 1 and Illumina BovineHD genotyping platforms are well designed and provide high-quality genotypes and similar coverage of informative SNP. The $\log_2$ intensity files can be extracted from both platforms for applications in CNV analysis. Despite fewer total SNP on BOS 1, 19% more SNP remained after LD pruning, resulting in a smaller gap size (5.2 vs. 6.9 kb) in Holstein and Jersey samples relative to BovineHD. For CNV analysis, the BovineHD platform provided an advantage because of the larger number of SNP, higher-intensity signals and lower background effects. The combined used of both platforms significantly improved coverage over either platform alone and this decreased the gap size between SNP, providing a valuable tool for fine mapping QTL and multibreed animal evaluation.

## ACKNOWLEDGMENTS

We thank Pfizer Animal Genetics, a business unit of Pfizer Animal Health (Kalamazoo, MI) for providing the Illumina HD genotypes of the 46 Jersey and 18 Holstein cattle used to assess SNP variation, Neogen/GeneSeek (Lincoln, NE) for genotyping, the Affymetrix development group (Affymetrix Inc., Santa Clara, CA) for their effort in developing an approach for CNV analysis with the BOS 1 chip. This work was supported, in part, by a University of California/California State University Collaborative Research Proposal to A. L. Van Eenennaam and B. L. Golden and by the University of California-W.K. Kellogg Endowment.

## REFERENCES

de Roos, A. P., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179:1503–1512.

Diskin, S. J., M. Li, C. Hou, S. Yang, J. Glessner, H. Hakonarson, M. Bucan, J. M. Maris, and K. Wang. 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 36:e126.

Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381–391.

Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009. Accuracy of genomic breeding values in multibreed dairy cattle populations. Genet. Sel. Evol. 41:51.

Hou, Y., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E. S. Kim, L. K. Matukumalli, M. Ventura, J. Song, P. M. VanRaden, T. S. Sonstegard, and C. P. Van Tassell. 2011. Genomic characteristics of cattle copy number variations. BMC Genomics 12:127.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4:e5350.

Matukumalli, L. K., S. Schroeder, S. K. DeNise, T. Sonstegard, C. T. Lawley, M. Georges, W. Coppieters, K. Gietzen, J. F. Medrano, G. Rincon, D. Lince, A. Eggen, L. Glaser, G. Cam, and C. Van Tassel. 2011. Analyzing LD blocks and CNV segments in cattle: Novel genomic features identified using the BovineHD BeadChip. Pub. No. 370-2011-002, Illumina Inc., San Diego, CA.

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.