Short communication

Evaluation of the BovineSNP50 genotyping array in four South African cattle populations

S.O. Qwabe[1,2#], E. vanMarle-Köster[2], A. Maiwashe[1] & F.C. Muchadeyi[3]
[1] Agricultural Research Council, Private Bag X2, Irene 0062, South Africa
[2] Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa
[3] Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort, 0110, South Africa

(Received 27 September 2012; Accepted 14 January 2013; First published online 20 March 2013)

Abstract

The BovineSNP50 genotyping array is a product with a wide range of applications in cattle such as genome-wide association studies, identification of copy number variation and investigation of genetic relationships among cattle breeds. It also holds potential for genomic selection, especially for traits that are expensive and difficult to measure. The successfulness of this chip for any of these applications depends on the degree of polymorphisms in the cattle breeds. The SNP50 array has not been validated in any South African cattle population and this could lead to overestimating the number of polymorphic SNPs available for application in it. This study is a first attempt to evaluate the Bovine64SNP50 genotyping array in the South African cattle population. Ninety six bovine samples, consisting of 45 Holstein, 29 Nguni, 12 Angus and 10 Nguni x Angus crossbred animals, were genotyped with the BovineSNP50 infinium assay. The results of this study demonstrated that 40 555 SNPs were polymorphic (MAF >0.05) in these breeds and indicate potential for application in South African cattle populations. Genomic information generated from the BovineSNP50 can now be applied in genetic prediction, genetic characterization and genome-wide association studies.

Keywords: Call rate, minor allele frequency
[#]Corresponding author: qwabes@arc.agric.za

The BovineSNP50 genotyping array (Illumina Inc., San Diego, CA), which features approximately 54 609 highly informative SNP probes uniformly distributed across the genome of the bovine, first became available in 2007 (Matukumalli *et al*., 2009). Since then the array has gained wide acceptance for use in genome-wide association studies to identify genomic regions contributing to natural variation in phenotypic traits in cattle. Several genomic regions associated with traits of economic interest have been identified. These include genomic regions associated with feed intake and feed efficiency traits (Sherman *et al*., 2009); milk production traits (Mai *et al*., 2010); growth and feedlot traits (Bolormaa *et al*., 2011) and carcass traits (Lee *et al*., 2010). Furthermore, the BovineSNP50 genotyping array has been used to demonstrate the value of genomic selection where genomic data has been used to supplement extensive sets of performance data to predict genetic merit for application in selection programmes (Meuwissen *et al*., 2001; Schaefer, 2006; Hayes *et al*., 2009; VanRaden *et al*., 2009). The success of genomic selection is based on exploitation of the linkage disequilibrium (LD) relationship between SNPs and quantitative traits loci (QTL) affecting a trait (Hayes *et al*., 2009). Genomic information has already been incorporated into the genetic evaluation of dairy and beef cattle in the United States and Canada (VanRaden *et al*., 2009; MacNeil *et al*., 2010, respectively). In addition, the BovineSNP50 array has been used to detect copy number variants (CNV), which have been implicated in both disease phenotype and normal phenotypic variation associated with QTL (Hou *et al*.,

URL: http://www.sasas.co.za
ISSN 0375-1589 (print), ISSN 222-4062 (online)
Publisher: South African Society for Animal Science
http://dx.doi.org/10.4314/sajas.v43i.7

2011). Furthermore, the BovineSNP array has been used to investigate genetic relationships among cattle breeds (Decker *et al*., 2009). The information has been applied in studies to detect selection signatures in Holstein cattle (Qanbari *et al*., 2010) and other dairy and beef breeds found on various continents (Stella *et al*., 2010).

Genetic improvement of cattle for traits of economic importance has primarily been performed using conventional selection based on quantitative genetics for many decades. Phenotypic and pedigree data collected by stud breeders participating in animal recording schemes have been used in genetic evaluations to provide breeders with estimated breeding values. However, the genetic architecture of the trait was treated as a black box, with no knowledge of the number of genes that affect the trait, let alone the effects of each gene or their locations in the genome (Dekkers, 2007). Genomic selection has the potential to increase genetic progress for traits that are difficult and expensive to measure (Meuwissen *et al*., 2001). To date, no local or indigenous South African cattle breeds have been included in either the production or the validation of the BovineSNP50 array. It is therefore necessary to investigate the usefulness of the BovineSNP50 in local breeds and also to evaluate the application in cattle breeds that are widely farmed in South Africa to establish the necessary reference populations. Thus, the aim of the study was to determine the level of polymorphism and allele frequency distribution of the BovineSNP50 in South African populations of Nguni, Angus and Holstein cattle.

Blood samples were obtained from 29 Nguni and 12 Angus animals, and 10 Nguni x Angus F1 cross steers from the beef cattle management and system development project of the Northern Cape Department of Agriculture and Land Reform (DALR) and Agricultural Research Council–Animal Production Institute (ARC-API). Blood samples (10 mL) were collected using EDTA VACUETTE® tubes. The blood samples were transported on ice to the ARC-API laboratory where they were refrigerated at – 20 ºC until extraction of DNA was performed. For Holstein, 45 semen samples were obtained from an artificial insemination company (Taurus, South Africa). Genomic DNA was extracted at the ARC-Biotechnology Platform laboratory from the whole blood and semen samples using the Qiagen DNeasy extraction kit (Qiagen, South Africa) according to the manufacturer's protocol. The protocol was adapted for the semen samples where Dithiothreitol (DTT) was added with proteinase K in the first step. Genomic DNA for all samples was quantified using a Qubit® 2.0 fluorometer and the Nano drop spectrophotometer (Nanodrop ND-1000). In addition, gel electrophoresis was performed to quantify the DNA. DNA samples extracted from blood had a 260 : 280 ratio ranging from 1.8 to 1.94, while eleven of the 45 DNA samples from semen extraction had a 260 : 280 ratio of less than 1.8. DNA concentrations for all samples ranged from 55 to 80 ng/μL.

Genotyping was conducted at the ARC-Biotechnology Platform with the Illumina BovineSNP50 BeadChip v2 which features 54 609 SNP probes distributed across the whole bovine genome with an average spacing of 49.9 kb (Matukumalli *et al*., 2009). Approximately 200 ng (12 μL of DNA loaded in each well of a BeadChip) of genomic DNA was used to genotype each sample. Samples were processed according to the Illumina Infinium–II assay protocol (Illumina, Inc. San Diego, CA 92122 USA). Briefly, each sample was whole-genome amplified for 20 hours at 37 °C. The samples were then fragmented, precipitated and re-suspended in an appropriate hybridization buffer. The samples were hybridized on the prepared BovineSNP50 BeadChip for 20 hours at 48 ºC. Following the hybridization, nonspecifically hybridized samples were removed by washing, while the remaining specifically hybridized loci were processed for the single-base extension reaction, stained and imaged on an Illumina iScan Reader.

Genotype data generated from the iScan system were loaded into Illumina Genome Studio version 1.9.0 software, which uses algorithms to perform primary data analysis, including raw data normalization, clustering and genotype calling. A final custom report was created from the genome studio using Plink Input Report 2.1.1, which created a ped (Pedigree file) and Map (SNP panel file) file to use for downstream analyses.

Basic genotype statistics for each marker, including call rate, minor allele frequency (MAF), Hardy-Weinberg Equilibrium (HWE), allele and genotype counts were calculated using the Quality Assurance Module from the SNP Variation Suite version 7 (SVS; Golden Helix Inc., Bozeman, Montana: www.goldenhelix.com). The following quality control criteria (filters) were used to remove from further analysis any SNPs with less than 95% call rate, and SNPs with less than 0.05 MAF. SNP were tested for HWE (P <0.001) to identify possible typing error. Samples with more than 10% missing genotypes were removed from the study.

For Nguni, Angus and Nguni x Angus populations all samples were successfully genotyped (>98% of SNP were genotyped). Five Holstein samples were removed because they had more than 10% missing genotype; these samples were part of the eleven samples with less than 1.8 260 : 280 ratio. Their 260 : 280 ratio ranged from 1.62 to 1.69, thus it was concluded that these failed to genotype owing to sample contamination; 77, 256 and 139 SNPs were removed in the Nguni, Angus and Holstein breeds, respectively, for violating HWE (P <0.001).

The average call rate across the four breeds was 98%. Over 95% of the SNPs had a call rate of greater than 95% in this study. This was comparable with an average call rate of 99.7% reported by the manufacture (Ilumina Inc., San Diego, CA) and also with the average call rate of greater than 97.9% reported by Makutumalli *et al*. (2009) across 21 different cattle breeds. Average call rate for individual breeds ranged from 96.9% (Holstein) to 99.7% (Angus). The results indicated that the BovineSNP50 array provides a useful tool for interrogating bovine genotype in numerous bovine breeds. Therefore, this array provides a robust resource for genome-wide association studies, genomic selection applications, investigating genetic relationship and detecting signatures of selection on South African Nguni, Angus, Nguni x Angus cross and Holstein cattle.

The successful application of the BovineSNP50 array depends largely on their degree of polymorphisms in the various cattle breeds (Fan *et al*., 2010). Therefore, SNPs with less than 5% MAF were removed from this study. Across all the populations, 40 555 or 74% of the 54 609 called SNPs were polymorphic, with an average minor allele frequency of 0.23 (Table 1). This indicated that the Bovine SNP50 array is informative among South African Nguni, Angus, Nguni x Angus and Holstein breeds to determine genetic variation underling these breeds.

The average MAF ranged from 0.17 (Nguni) to 0.22 (Holstein) (Table 1). This was in agreement with the previous observation that reveals considerable variations in MAF between breeds (Matukumalli *et al*., 2009). Holstein (41 078) and Angus (40 146), which are common breeds in the USA and Europe, had higher proportions of polymorphic SNP compared with the Nguni (35 843), an indigenous breed to South Africa. Matukumalli *et al*. (2009) observed a similar trend in studies with European and African breeds with 42 849 and 41 073 polymorphic SNPs in Holstein and Angus cattle, respectively, and only 28 869 and 35 084 SNPs in the African N'Dama and Sheko breeds. It was encouraging to confirm that the BovineSNP50 array will be equally informative for use in South African Sanga compared with the other African breeds that were included during the validation of the BovineSNP50 array.

Table 1 Number of polymorphic loci in South African breeds

Populations	Samples	Polymorphic loci*	Mean MAF***	Median MAF***
Holstein	40	41 078	0.22	0.21
Angus	12	40 146	0.21	0.20
Nguni x Angus	10	38 979	0.19	0.20
Nguni	29	35 843	0.17	0.13
All breeds combined	91	40 555	0.23	0.23

*Minor allele frequency >0.05.
***Across all 54 609 loci.

It can be concluded from this study that the BovineSNP50 array will be applicable to the South African cattle populations, provided that the quality of DNA used, meets the required quality for infinium assay. It was observed that sample contamination reduces SNP call rate for individual animals which, in this proof-of-concept study, then reduced the average call rate for Holstein. Overall, the results of this study demonstrate that the BovineSNP50 array will be useful for genomic studies across three breeds that are widely used by South African farmers for dairy and beef production.

Acknowledgments

The authors would like to thank the ARC-Biotechnology Platform for making their laboratory resources available for genotyping samples. Financial support from the ARC is greatly appreciated. Provision of semen on Holstein bulls by Taurus Co-operative is also acknowledged.

References

Anderson, C.A., Pettersson, F., Clarke, G., Cardon, L., Morris, A. & Zondervan, K., 2011. Data quality control in genetic case-control association studies. Nat Protoc. 5, 1564-1573. doi:10.1038/nprot.2010.116.

Bolormaa, S., Hayes, B.J., Savin, K., Hawken, R., Barendse, W., Authur, P.F., Herd, M.R. & Goddard, M.E., 2011. Genome-wide association studies for feedlot and growth traits in cattle. J. Anim. Sci. 89, 1684-1697.

Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilkki, J., Seabury, C.M., Caetano, A.R., Johnson, G.S., Brenneman, R.A., Hanotte, O., Eggert, L.S., Wiener, P., Kim, J.J., Kim, K.S., Sonstegard, T.S., Tassell, C.P., Neibergs, H.L., McEwan, J.C., Brauning, R., Coutinho, L.L., Babar, M.E., Wilson, G.A., McClure, M.C., Rolf, M.M., Kim, J., Schnabel, R.D. & Taylor, J.F., 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc. Natl. Acad. Sci. 106, 18644-18649.

Dekkers, J.C.M., 2007. Prediction of response from marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124, 331-341.

Fan, B., Zhi-Qiang, D., Danielle, M., Gorbach, M. & Rothschild, F., 2010. Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Aust. J. Anim. Sci. 23, 833-847.

Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K. & Goddard, M.E., 2009. Accuracy of genomic breeding values in multi-breed dairy cattle population. Genet. Sel. Evol. 41, 51-56.

Hou, Y., Liu, G.E., Bickhart, D.M., Cardone, M.F., Wang, K., Kim, E.S., Matukumalli, L.K., Ventura, M., Song, J. & VanRaden, P.M., 2011. Genomic characteristics of cattle copy number variations. BMC Genomics 12, 127-135.

Lee, Y.M., Han, C.M., Li, Y., Lee, J.J., Kim, L.H., Kim, J.H., Kim, D.I., Lee, S.S., Park, B.L., Shin, H.D., Kim, K.S., Kim, N.S. & Kim, J.J., 2010. A whole genome association study to detect single nucleotide polymorphism for carcass traits in Hanwoo populations. Asian-Aust. J. Anim. Sci. 23, 417-424.

MacNeil, M.D., Nkrumah, J.D., Woodward, B.W. & Northcutt, S.L., 2010. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. J. Anim. Sci. 88, 517-522.

Mai, M.D., Sahana, G., Chistiansen, F.B. & Guldbrandtsen, B., 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. J. Anim. Sci. 88, 352-3528.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S. & Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4: e5350-5063.

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

Qanbari S., Pimentel, E.C.G, Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. & Simianer, H., 2010. The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet. 41, 346-356.

Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 12, 218-223.

Sherman, E.L., Nkrumah, J.D., Bartusiak, C.L.R., Murdoch, B. & Moore, S.S., 2009. Fine mapping quantitative trait loci (QTL) for feed intake and efficiency in beef cattle. J. Anim. Sci. 87, 37-45.

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