Research Article

Performance of bovine high density SNPs genotyping array in indigenous Pakistani cattle breeds

Hamid Mustafa[1*], Waqas Ahmad Khan[2], Huson J. Heather[3], Zulfiqar Hussain Kuthu[4], Kim EuiSoo[5], Adeela Ajmal[1], Noor Ul Ain[6] and Tad S. Sonstegard[5]

1. Department of Livestock Production, University of Veterinary and Animal Sciences, Ravi Campus, Pattoki–Pakistan
2. Department of Biotechnology, Faculty of Science, University of Sargodha, Sargodha-Pakistan
3. Department of Animal Sciences, Cornell University, Ithaca, NY-USA
4. Faculty of Veterinary and Animal Sciences, University of Poonch Rawalakot, Azad Kashmir-Pakistan
5. Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland 20705-USA
6. The University of Agriculture, Peshawar-Pakistan
Corresponding author's email: hamidmustafapasha@gmail.com

Abstract

The bovine high density SNPs genotyping array has a wide range of applications including detection of selection signatures, identification of copy number variants (CNVs), genome wide association studies, and dendrogram relationships. This array can also be used to improve the accuracy of genomic predictions for low heritable traits viz. reproductive traits. The effectiveness of this array for genomic selection mainly depends on polymorphism level. In this study, we used 136 individuals from ten different Pakistani cattle breeds, include Achi (18), Bhagnari (14), Cholistani (13), Dajal (10), Dhanni (10), Kankraj (12), Lohani (19), Red Sindhi (13), Sahiwal (14), and Tharparkar (13) using high density SNPs genotyping and this array contained approximately 777, 962 SNPs. The results of this study revealed that approximately 500, 939 SNPs were found polymorphic in these breeds. The results indicate that high density SNPs bead chip would offer an informative genotyping platform for quantitative trait loci (QTL) mapping in indigenous Pakistani cattle breeds.

Keywords: Copy number variants; Genotyping; Polymorphism; Quantitative trait loci; SNPs

Introduction

The bovine high density SNPs is the most comprehensive genome wide genotyping array. This chip contained more than 777, 962 SNPs markers that are equally distributed across entire bovine genome [1]. This array was first time introduced and became available in 2010 [2]. Applications of this array include genome wide association studies, identification of quantitative traits loci (QTLs), genetic merits prediction, and extent of linkage disequilibrium (LD). The

potential of this array has been proved in several studies and identified genomic regions, which effects on phenotypic variations [3]. These regions are mainly associated with milk traits [4]; feed intake, growth traits, and carcass traits [5]. In addition, the prediction of genomic values in an effective breeding program is based on reliable genomic data, which have been extensively used for cattle selection [6]. The genomic selection is based on the existence of linkage disequilibrium (LD), SNPs associations, and quantitative traits loci (QTLs) that effect on traits of interest [7]. In advanced countries (United States and Canada) genomic information's have widely used for the estimation of dairy and beef animals genetic worth [8]. This array would also be used to identify copy number variations (CNVs), which are commonly used for QTLs association with both normal and disease phenotype and explore dendrogram relationships among different cattle populations [9]. Currently, this array is very effective to identify genome wide selection signatures in different dairy and beef breeds [4, 8, 10, 11].

In Pakistan, all genetic improvement programmes for cattle is still based on conventional quantitative genetic methods. There is also limited availability of phenotypic and pedigree data information for estimation of breeding values. However, it has been found that genetic selection has a high potential to increase genetic gain in cattle and also permits more accurate genetic predictions for low heritability traits in farm animals than conventional phenotypic selection [6]. The indigenous cattle breeds of Pakistan are still lacking the opportunity for high throughput genomic evaluation for better understanding of domestication process and breed improvement programs. To date, no indigenous Pakistani cattle breed has been included either in training or a validation population using bovine high density SNPs Bead Chip. The evaluation of this high-throughput technique may help to improve the cattle farming. Therefore, the aim of this study was to find the level of informativeness of the bovine high density SNPs bead chip by measuring loci polymorphism in indigenous Pakistani cattle population.

Materials and methods

Animal's selection and DNA extraction

In this study, we used 136 individuals from ten different Pakistani cattle breeds; include Achi (18), Bhagnari (14), Cholistani (13), Dajal (10), Dhanni (10), Kankraj (12), Lohani (19), Red Sindhi (13), Sahiwal (14), and Tharparkar (13) from different livestock experiment stations (Table 1). A 10 ml jugular blood was collected in falcon tubes containing EDTA and genomic DNA was extracted using modified organic method. Genomic DNA (gDNA) was qualitative and quantitatively checked through gel electrophoresis and Nanodrop spectrophotometer (Nanodrop 1000).

Genotyping was performed at the BFGL, ARS/USDA, USA, platform with the Illumina bovine high density SNPs bead chip v2 containing 777, 962 SNPs disseminated across the bovine genome. Approximately, 200 ng genomic DNA was used to performed genotyping for each sample. Genomic DNA was processed according to manufacturer protocol (the Illumina, Inc. San Diego, CA, USA). Amplified sample was incubated at 37^0C for 20 h. The hybridization of samples was performed on the bovine high density SNPs bead chip at 48^0C for 20h. Following the hybridization, nonspecifically hybridized loci were stained and imaged on an Illumina iScan Reader after processing for single base extension reactions.

Data analysis

Genotypic data were generated from the iScan system. Primary data analysis including genotyping calling, clustering and data normalization was performed by using

Illumina genome studio version 1.9.0 software. Ped and map file was created for downstream analyses from the genome studio using plink input report 2.1.1. Quality assurance module were used form SVS, version 7; Golden Helix Inc., USA (www.goldenhelix.com) for genotypic statistics calculation for each marker including Hardy-Weinberg equilibrium (HWE), minor allele frequency (MAF), call rate (CL), allele and genotype count. Quality control (QC) criteria for further analysis were < 95% call rate and <0.05 minor allele frequency (MAF). Hardy Weinberg equilibrium (P<0.001) was tested to help identify genotyping errors.

Table 1. Sampling detail

Population	N	Purpose	Province
Achi	18	Milk and Meat	Khyber Phaktunpkhua
Bhagnari	14	Work	Balochistan
Cholistani	13	Milk and Meat	Punjab
Dhanni	10	Work and Milk	Punjab
Dajal	10	Work and Meat	Punjab
Kankraj	12	Work and Meat	Sindh
Lohanni	19	Work and Milk	Balochistan
Red Sindhi	13	Milk	Sindh
Sahiwal	14	Milk	Punjab
Tharparkar	13	Work and Milk	Sindh

Results and discussion
Call rate (CL)
All samples were genotyped successfully with 64 % call rate (CL). In Illumina Infinium assay, poor performance of probes is indicated by low call rate, due to the result of poorly separated clusters genotyping or due to mismatched of cluster positions and rejected for further analysis [2]. The call rate for a sample is mainly depends on the genomic DNA quality [12]. In this study, the call rate for Pakistani cattle breeds were ranged from 0.80 to 0.90 with a mean 0.85±0.0029. This indicates that on average 661, 267 markers were genotyped for each studied individual. The result of this study is comparable with previous reported call rates in different taurine and indicine breeds [1, 13].

Polymorphic and fixed loci number
Fixed loci are considered as non-informative or non-polymorphic loci and therefore, reduced the SNPs markers value [14]. The polymorphic loci value for the bovine high density SNPs Bead Chip significantly differed between different cattle populations and ranged from 538,517 for indicine to 651,994 for taurine (http://www.illumina.com/datasheet_bovine_HD_Snp.pdf) (Table 2). The number of total non-polymorphic (fixed) autosomal loci for Pakistani indigenous cattle breeds were 63, 629 (8.2% of all autosomal probes) giving a final number of 500,939 polymorphic loci. The polymorphic markers numbers in this study is in close agreement with previous reported markers numbers in different zebu breeds and confirmed that no any indigenous Pakistani cattle breed has been included either in training or a validation population using the Bovine high density SNPs Bead Chip [13].

Minor allele frequency (MAF)
A rare variant of minor allele frequency (MAF) estimation is difficult due to the limitation of markers in genome wide association studies. The minor allele frequency (MAF) used in a typical genome wide association studies (GWAS) is below 1%. In this study, we used minor allele

frequency for all autosomes in each breed ranged from 0.18 to 0.23 with average value 0.21±0.1552, which was slightly lower than previously reported MAF in Brahman, Gir, and Nellore cattle [1]. Simultaneously, the observed minor allele frequency in this study was higher than African zebu [15]. The average minor allele frequency (MAF) ranged from 0.18 (BTA 22) to 0.23 (BTA 25). Elimination of low polymorphic loci and monomorphic loci (MAF<0.01), the remaining average minor allele frequency was 0.21±0.1462.

Table 2. The Bovine HD SNP Bead Chip content validation in Pakistani cattle breeds

Breed	N	Polymorphic Loci	Mean MAF*	Median MAF
Achi	18	493, 872	0.18	0.17
Bhagnari	14	508,995	0.20	0.19
Cholistani	13	510,898	0.22	0.17
Dajal	10	533,996	0.22	0.20
Dhanni	10	498,898	0.20	0.19
Kankraj	12	499,399	0.18	0.15
Lohani	13	521,999	0.20	0.17
Red Sindhi	19	499,798	0.22	0.21
Sahiwal	14	500,968	0.21	0.22
Tharparkar	13	502,538	0.18	0.16
	136	500,939	0.21	0.18

*Minor Allele Frequency (MAF) >0.05

Hardy-Weinberg Equilibrium (HWE)
The identification of potentially loci, which is under selection or predisposed to genotyping error, is measure with Hardy-Weinberg equilibrium (HWE) [16]. The present study found 392 loci, which deviated from hardy-Weinberg equilibrium (HWE) and distributed equally across all autosomes, which is in close agreement with the study of Edea et al [15].

SNPs on sex chromosomes
We studied sex chromosomes separately, due to their impaired performance in males. 2130 SNPs contained on chromosome X and Y, 915 had minor allele frequency higher than 1 %. First 71 SNPs were deviated from Hardy-Weinberg equilibrium (p<0.001) and only 22 SNPs were performed poor, which ultimately removed from data set. After filtration only 921 SNPs performance were found satisfactory and could be used for further application regarding sex chromosomes analysis similarly reported in indicine cattle [1].

Linkage disequilibrium (LD) Pruning
The linkage disequilibrium method is used to identify linked SNPs to defined genetic variation in genome wide association studies (GWAS) [14]. SNPs linkage or in high linkage disequilibrium (LD) genetic information is same in runs of homozygosity, therefore LD pruning is recommended [17]. In this study, we used r^2 threshold to detect linked SNPs (r^2= 0.45). This method prune 446, 715 SNPs distributed across all autosomes. The average distance between the SNPs markers for the Bovine high density SNPs Bead Chip is 3.43 kb (http//www/illumina.com).

Conclusion
In conclusion, the analysis of this study showed that 42, 669 markers in the bovine high density SNPs Bead Chip are limited for use of population genetics or association studies in indigenous Pakistani cattle breeds. The monomorphic loci, which is less than 1% has no genomic coordinates. This study provides 500, 939 informative SNPs, which

is 64% of the all autosomes and can be potentially used for further genomic applications. The results of this study is comparable with previously then reported studies in worldwide cattle populations [1]. The results of this study encourage that the bovine high density SNPs bead chip is very informative and can be used for further genomic studies in Pakistani indigenous cattle breeds based on different agro-ecological zones. The use of this array can also improve the understanding of evolutionary processes of Pakistani indicus cattle breeds and future synthetic breed development for improved milk and meat production.

Authors' contributions

Conceived and designed the experiments: H Mustafa & TS Sonstegard, Performed the experiments: H Mustafa, HJ Heather & K Eui-soo, Analyzed the data: H Mustafa, WA Khan, HJ Heather & K Eui-soo, Contributed materials/ analysis/ tools: Hamid Mustafa and TS Sonstegard, Wrote the paper: H Mustafa, ZH Kuthu, K Eui-soo, A Ajmal & NU Ain.

References

1. Matukumalli, LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS & Van Tassell, CP (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4: e5350-5063.
2. Mai MD, Sahana G, Chistiansen FB & 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.
3. Fan H, Wu Y, Qi X, Zhang J, Li J, Gao X, Zhang L, Li J & Gao H (2014). Genome-wide detection of selective signatures in Simmental cattle. *J Appl Genet* 55: 343-351.
4. Mustafa H, Nisar A, Huson JH, Kim ES, Waqas AK, Adeela A, Khalid J, Talat NP, Afzal A, Jim JK & Tad SS (2018). Whole genome study of linkage disequilibrium in Sahiwal cattle. *South African J Anim Sci* 48: 353-360.
5. Fan B, Zhi-Qiang D, Danielle MG & Max FR (2010). Development and Application of High-density SNP Arrays in Genomic Studies of Domestic animals. *Asian-Aust. J. Anim. Sci* 23: 833 – 847.
6. Meuwissen THE, Hayes BJ & Goddard ME (2001). Prediction of total genetic value using genome wide dense marker maps. *Genetics* 157: 1819-1829.
7. Hayes BJ, Bowman PJ, Chamberlain AJ & Goddard ME (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92: 433-43.
8. Makina SO, Taylor JF, van Marle-Koster E, Muchadeyi FC, Makgahlela ML, MacNeil MD & Maiwashe A (2015). The extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. *Front Genet* 6: 337.
9. Hou Y, Liu GE, Bickhart DM, Matukumalli LK, Li C, Song J, Gasbarre LC, Van Tassell CP & Sonstegard TS (2012). Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. *Funct Integr Genomics* 12: 81-92.
10. Stella A, Ajmone-Marsan P, Lazzari B & Boettcher P (2010). Identification of selection signatures in cattle breeds

selected for dairy production. *Genetics* 185: 1451-1461.

11. Randhawa IA, Khatkar MS, Thomson PC & Raadsma HW (2016). A Meta-Assembly of Selection Signatures in Cattle. *PLoS One* 11: e0153013.

12. Gurgul A, Kacper U, Klaudia P, Tomasz Z, Ewelina S & Monika B. 2013. The Evaluation of Bovine SNP50 Bead Chip Assay Performance in Polish Red Cattle Breed. *Folia Biologica (Kraków)* 61: 3-4.

13. Porto-Neto RL, Sonstegard TS, George E L, Derek M B, Marcos VB Da S, Marco A M, Yuri TU, Jose F G, Cedric G & Curtis P Van T (2013). Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14: 876.

14. Rincon G, Weberk L,Vaneenennaama L, Goldenb L & Medrano JF (2011). Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. *J Dairy Sci* 94: 6116-6121.

15. Edea Z, Dadi H, Kim SW, Dessie T & Kim KS (2012).Comparison of SNP variation and distribution in indigenous Ethiopian and Korean cattle (Hanwoo) populations. *Genomics Inform* 10: 200-205.

16. Leal SM. (2005). Detection of genotyping errors and pseudo-SNPs via deviations from Hardy-Weinberg equilibrium. *Genet Epidemiol* 29: 204-214.

17. Purcell S1, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3): 559-575.