



Evaluating efficiency of customized medium density INDUSCHIP for genotyping of Indicine cattle breeds

SUJIT SAHA¹✉, NILESH NAYEE¹, A SUDHAKAR¹, SWAPNIL GAJJAR¹, K R TRIVEDI¹, R O GUPTA¹ and G KISHORE¹

National Dairy Development Board, Anand, Gujarat 388 001 India

Received: 19 March 2020; Accepted: 29 August 2020

ABSTRACT

To initiate genomic selection programme for indicine cattle breeds and their crosses in India, National Dairy Development Board designed a medium-density (52K) customized chip on Illumina platform named as “INDUSCHIP”. The present study was conducted to examine the efficiency of INDUSCHIP SNP panel for genotyping indicine cattle breeds. Total of 500 animals belonging to 14 different indicine breeds were genotyped with Illumina Bovine HD chip. A subset of SNPs was taken for evaluating the performance of selected SNPs in different indicine breeds. The average minor allele frequency (MAF) was found to vary between 0.20–0.29 for different indicine breeds. However, for important milk breeds like Sahiwal, Gir, Red Sindhi and Kankrej the average MAF was found to be 0.27 and above. Mean Linkage Disequilibrium (LD) at 50–60 kbp distance was found to be around 0.21. There was considerable LD decay with increasing distance between SNPs. Around 0.06% SNPs were found to be significantly deviating from Hardy-Weinberg equilibrium. From the Principal component analysis (PCA) it was found that the first three Principal Components i.e. PC1, PC2 and PC3 could separate different indicine breeds. The present study indicated that due to the presence of highly polymorphic SNPs for the breeds of indicine origin, INDUSCHIP panel was found to be effective and informative in genotyping indicine breeds.

Keywords: HWE, INDUSCHIP, LD, MAF, SNP

Systematic genetic improvement programmes for important dairy breeds of cattle and buffaloes of India was initiated by the National Dairy Development Board (NDDB) under National Dairy Plan phase I (NDP I). With the large volume of phenotypic data generated under these programmes, an opportunity has been opened up to introduce genomic selection for Indian cattle and buffalo breeds. To utilize this opportunity, the performance of commercially available medium density chips namely Illumina 50K, Geneseek 75Ki, and Illumina Bovine LD chips was evaluated in 4 major breeds of dairy cattle (i.e. Gir, Sahiwal, Kankrej and Red Sindhi) and their Taurine crosses (Nayee *et al.* 2018). These existing panels were found to be useful either only for *B. Indicus* or for *B. Taurus* but not for both. Subsequently, a medium-density customized chip named as “INDUSCHIP” was developed by NDDB (Mrode *et al.* 2019) with technical guidance from Center for Quantitative Genetics and Genomics, Foulum, Aarhus University, Denmark for genotyping the indicine breeds of cattle and their crosses. The array contains around 45K polymorphic markers for Indian cattle breeds and 8K base SNPs of Illumina Bovine LD chip. The SNPs, which were highly polymorphic across indicine breeds, uniformly

distributed across the genome and having average linkage disequilibrium (r^2) of around 0.2 at 50 kb distance, were selected. The present study was carried out to examine the performance of INDUSCHIP in various indicine breeds.

MATERIALS AND METHODS

A total of 500 individuals, belonging to 14 different indicine breeds namely Amritmahal (30), Deoni (30), Gir (86), Hallikar (30), Hariana (25), Kangayam (30), Kankrej (23), Khillar (21), Ongole (27), Rathi (39), Red Sindhi (45), Sahiwal (48), Siri (30) and Tharparkar (36) were genotyped with Illumina Bovine-HD chip at M/s Sandor Life Sciences Ltd., Hyderabad, India. Data for 490 samples having call rates > 90% were obtained in PLINK (Purcell *et al.* 2007) format for analysis. The samples having less than 90% genotyping rates and the SNPs that could be genotyped in less than 90% of individuals. A subset of 52,363 SNPs available in INDUSCHIP was taken for evaluating the performance of selected SNPs in different indicine breeds. The Minor allele frequency (MAF) and average Linkage Disequilibrium at a specific distance (measured in r^2) and Hardy Weinberg equilibrium were calculated for various breeds using PLINK. Genotyping data of 51,469 SNPs and 490 individuals were used for a Principal component analysis (PCA) using EIGENSOFT (Price *et al.* 2006) to assess the efficiency of this assay to differentiate *Bos indicus* breeds under the study.

Present address: ¹National Dairy Development Board, Anand, Gujarat 388001. ✉Corresponding author e-mail: ssaha@nddb.coop; sujitsahaabc@gmail.com

Table 1. Distribution of SNPs (%) based on minor allele frequency (MAF) in different indicine breeds

Breed	MAF range					
	<0.005	0.006–0.105	0.106–0.205	0.206–0.305	0.306–0.405	0.406–0.505
Amritmahal	16.93	12.59	13.27	16.54	20.32	20.36
Deoni	12.04	14.70	11.71	16.13	21.60	23.83
Gir	5.78	20.06	8.36	13.17	23.80	28.83
Hallikar	11.17	15.31	11.62	15.73	21.89	24.28
Hariana	9.14	16.54	10.33	15.49	22.44	26.07
Kangayam	17.48	19.55	15.45	15.90	16.32	15.30
Kankrej	9.59	14.45	10.04	14.34	20.94	30.64
Khillar	10.91	14.61	10.80	15.24	25.39	23.05
Ongole	13.65	13.33	13.35	15.47	20.28	23.91
Red Sindhi	3.88	21.82	11.77	15.38	22.29	24.85
Sahiwal	8.60	16.81	8.24	12.63	21.81	31.92
Siri	1.00	12.78	16.61	18.83	24.33	26.45
Tharparkar	15.14	14.47	12.65	15.46	22.20	20.07
Rathi	5.06	19.28	9.80	15.10	22.85	27.91

RESULTS AND DISCUSSION

Distribution of SNPs based on Minor Allele Frequency (MAF): The average Minor allele Frequency (MAF) was found to vary between 0.20 (in Kangayam) to 0.29 (Siri) in the various indicine breeds. For various milch breeds like Sahiwal, Gir, Red Sindhi and dual purpose breed like Kankrej it was found to be 0.27, 0.27, 0.26 and 0.27, respectively. Mustafa *et al.* (2018) using Illumina Bovine HD SNPs Beadchip reported the average MAF in Sahiwal as 0.21. The MAF in indicine milch breed was found to be higher than the draft purpose breeds. The distribution of SNPs based on Minor Allele Frequency (MAF) in fourteen indicine breeds is presented in Table 1 and Supplementary fig.1. It was observed that on an average 46.7% SNPs across all the breeds had MAF greater than 0.30, whereas around 26% SNPs had MAF less than or equal to 0.1. For breeds like Gir, Kankrej, Sahiwal, Rathi and Siri more than 50% of the SNPs had MAF greater than 0.3. This indicates that the SNPs selected for designing INDUSCHIP were polymorphic for these breeds and were capable of capturing

the SNP level variations to a great extent than the draft breeds, where the majority of the SNPs having lower MAF.

Bejarano *et al.* (2018) in two Colombian Creole cattle breeds Blanco Orejinegro (BON) and Romosinuano (ROMO) also observed average MAF as 0.27 ± 0.14 , with a higher proportion of SNPs (around 45%) having $MAF \geq 0.3$ using a medium-density SNP panel (BovineSNP50K_v2). On the other hand, Xu *et al.* (2019) reported a higher proportion of low MAF SNPs in Chinese cattle than the imported breeds (Simmental and Wagyu).

Linkage disequilibrium (LD) pattern: Distribution of LD (r^2) estimates over different physical map distance among the SNPs were calculated. Mean r^2 values for various distance groups are presented in Table 2.

It was observed that on an average SNPs located at 50–60 kb distance was having r^2 values at around 0.21 and there was considerable LD decay with increasing distance between SNPs. Mustafa *et al.* (2018) reported that across all the chromosomes the mean LD between SNPs ranged from 0.002 to 0.19 for r^2 with overall mean 0.18 in Sahiwal cattle and a decreasing trend in r^2 was observed as the distance between SNPs increased from 1 kb (0.35) to 100 kb (0.12). Espigolan *et al.* (2013) genotyped Nellore cattle using 444, 986 SNPs and found mean linkage disequilibrium (r^2) between adjacent markers as 0.17. Bejarano *et al.* (2018) in two Colombian Creole cattle breeds Blanco Orejinegro (BON) and Romosinuano (ROMO) observed optimal LD levels ($r^2 \geq 0.3$) in a distance of up to 70 kb and 100 kb, respectively, which is higher than the estimates observed in the current study. The lower estimate of LD in indicine breeds as observed in the present study indicates the population is not closed and having a relatively larger effective population size.

Hardy Weinberg Equilibrium (HWE): The breed wise proportion of SNPs deviating from Hardy Weinberg Equilibrium (HWE) is presented in Table 3. From the result, it was observed that only 0.06% SNPs (0.0019–0.38%) were deviating significantly from HWE ($P < 0.00001$) for various indicine breeds. However, for Gir and Sahiwal breeds, the deviation percentages were 0.38% and 0.10%, respectively,

Table 2. Mean linkage disequilibrium (r^2) values at different distance (kb) between SNPs

Distance in kb	0–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–100
Amritmahal	0.42	0.29	0.25	0.27	0.23	0.23	0.21	0.19	0.19	0.18
Deoni	0.40	0.26	0.22	0.24	0.20	0.21	0.19	0.17	0.16	0.16
Gir	0.36	0.23	0.20	0.22	0.18	0.19	0.17	0.15	0.15	0.14
Hallikar	0.38	0.26	0.21	0.23	0.18	0.19	0.18	0.15	0.15	0.15
Hariana	0.38	0.25	0.21	0.22	0.18	0.20	0.17	0.15	0.16	0.14
Kangayam	0.45	0.33	0.30	0.30	0.26	0.28	0.26	0.24	0.24	0.24
Kankrej	0.37	0.25	0.21	0.22	0.18	0.19	0.17	0.16	0.15	0.15
Khillar	0.38	0.26	0.22	0.23	0.19	0.20	0.18	0.16	0.17	0.15
Ongole	0.41	0.28	0.24	0.25	0.21	0.22	0.20	0.18	0.17	0.17
Red Sindhi	0.38	0.24	0.21	0.23	0.19	0.21	0.19	0.17	0.18	0.16
Sahiwal	0.36	0.23	0.18	0.21	0.16	0.18	0.15	0.13	0.13	0.13
Siri	0.37	0.23	0.20	0.20	0.17	0.18	0.16	0.14	0.14	0.14
Tharparkar	0.41	0.29	0.25	0.27	0.23	0.24	0.21	0.18	0.19	0.19
Rathi	0.36	0.23	0.18	0.20	0.16	0.17	0.16	0.13	0.13	0.13

Table 3. Number of SNPs deviating from HWE ($P < 0.00001$) in various indicine breeds

Breed	Deviation from HW equilibrium ($P < 0.00001$)	
	No. of SNPs	% of SNPs
Amritmahal	15	0.0283
Deoni	14	0.0264
Gir	206	0.3882
Hallikar	22	0.0415
Hariana	6	0.0113
Kangayam	24	0.0452
Kankrej	5	0.0094
Khillar	1	0.0019
Ongole	3	0.0057
Red Sindhi	46	0.0867
Sahiwal	57	0.1074
Siri	9	0.0170
Tharparkar	35	0.0660
Rathi	29	0.0546

which were higher in comparison to other indicine breeds. It may be due to the fact that being prominent dairy breeds of India, the selection was practiced for Gir and Sahiwal cattle than other draft breeds.

Principal Component Analysis (PCA): First ten principal components were obtained using EIGENSOFT software (Supplementary Table 1; supplementary Fig. 1 and Fig. 2). Principal components 1, 2 and 3 obtained from the PCA using SNPs data are plotted in Fig. 1. It was observed that separation of various indicine breeds could be possible using the first three principal components namely PC 1, PC 2 and PC 3.

The present study indicates that INDUSCHIP represents a panel with highly polymorphic SNPs for dairy breeds of Indicine origin. The SNPs selected for designing the microarray are stable in terms of Hardy Weinberg Equilibrium. Overall, only 0.06% SNPs were found to deviate significantly from HWE ($P < 0.00001$) for various indicine breeds (Supplementary Fig. 2). On an average r^2

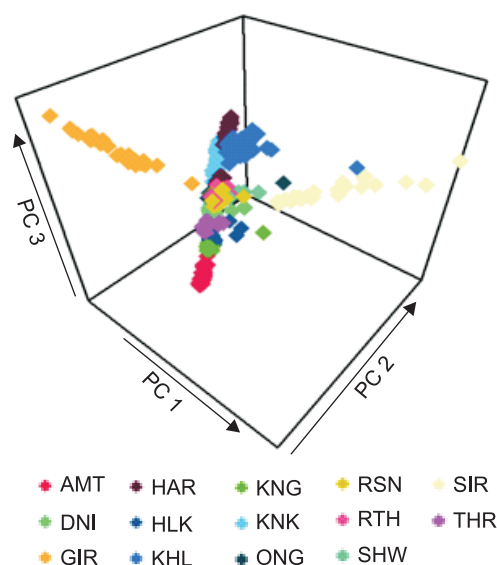


Fig. 1. PCA of SNPs of indicine breeds.

values for SNPs at 50–60 kb distance was 0.2, which is nearing the threshold of the suggested LD level (Xu *et al.* 2019). The plot obtained from PCA analysis using INDUSCHIP SNPs separated major indicine breeds. However, there is scope available for incorporating additional breed specific polymorphic SNPs for separating these indicine breeds particularly draft breeds more distinctively and improve the accuracy of genomic prediction.

ACKNOWLEDGEMENTS

Under Indo-Danish collaboration, guidance provided by Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Research Center, Foulum, Aarhus University, Denmark, made it possible to select INDUSCHIP panel. The authors humbly thank NDDB management for funding and supporting liberally for this project. The cooperation extended by the Field staff and officials from various PT, PS projects, Livestock Farms for sample collection and services extended by the staff of Genomics lab, NDDB is duly acknowledged.

REFERENCES

- Bejarano D, Martinez R, Manrique C, Parra L M, Rocha J F, Gomez Y, Abuabara Y and Gallego J. 2018. Linkage disequilibrium levels and allele frequency distribution in Blanco Orejinegro and Romosinuano Creole cattle using medium density SNP chip data. *Genetics and Molecular Biology* **41**(2): 426–33.
- Espigolan R, Fernando B, Arione A B, Fabio R P S, Daniel G M G, Rafael L T, Diércles F C, Henrique N O, Humberto T and Mehdi S. 2013. Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics* **14**: 305.
- Mustafa H, Ahmad N, Heather H J, Eui-soo K, Khan W A, Ajmal A, Javed K, Pasha T N, Ali A, Kim J J and Sonstegard T S. 2018. Whole genome study of linkage disequilibrium in Sahiwal cattle. *South African Journal of Animal Science* **48**: 353–60.
- Mrode R, Ojango J M K, Okeyo A M and Mwacharo M. 2019. Genomic Selection and use of Molecular tools in Breeding Programs for Indigenous and Crossbred Cattle in Developing Countries: Current Status and Future Prospects. *Frontiers in Genetics*. **9**: 694.
- Nayee N, Sahana G, Gajjar S, Sudhakar A, Trivedi K R, Lund M S and Guldbbrandtsen B. 2018. Suitability of existing commercial single nucleotide polymorphism chips for genomic studies in *Bos indicus* cattle breeds and their *Bos taurus* crosses. *Journal of Animal Breeding and Genetics*. doi: 10.1111/jbg.12356
- Price A L, Patterson N J, Plenge R M, Weinblatt M E, Shadick N A and Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**: 904–09.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M A, Bender D, Maller J, Sklar P, de Bakker P I, Daly M J and Sham P C. 2007. PLINK: A tool set for whole-genome association and population based linkage analysis. *American Journal of Human Genetics* **81**: 559–75.
- Xu L, Zhu B, Wang Z, Xu L, Liu Y, Chen Y, Zhang L, Gao X, Gao H, Zhang S, Xu L and Li J. 2019. Evaluation of linkage disequilibrium, effective population size and haplotype block structure in Chinese cattle. *Animals* **83**: 1–14.