Indian Journal of Animal Sciences **89**(8): 861–865, August 2019/Article
https://doi.org/10.56093/ijans.v89i8.93019

Genome wide association study of test days and 305 days milk yield in crossbred cattle

AAMIR BASHIR WARA[1], AMIT KUMAR[2], AKANSHA SINGH[3], A K ARTHIKEYAN[4], TRIVENI DUTT[5] and B P MISHRA[6]

ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh 243 122 India

ABSTRACT

The present study was conducted to identify SNPs associated with test days and 305 days milk production at genome level after correcting for the non-genetic factors affecting these traits in crossbred (*Vrindavani*) cattle. Crossbred cattle (96) were genotyped using double digestion genotyping-by-sequencing technique for genome wide association study (GWAS) with first lactation milk production traits. The effect of season was significant on TD36, TD66 and TD96. Initial quality control for genotyping call rate, Hardy-Weinberg equilibrium and minor allele frequency were achieved by using PLINK tool. SNPs (9638) were retained for ascertaining GWAS with first lactation milk production traits and was accomplished by regressing SNPs on first lactation milk traits using PLINK. Total 23, 28, 112, 3, 13 and 5 SNPs were found to be significantly associated with AVDY, PY, TD36, TD66, TD96 and FL305MY, respectively. Most of the SNPs were located within *KIRREL3* or near to it on chromosome 29, followed by *LRRC3* and *TSPEAR* on chromosome 1. Three SNPs (NC_007299.6_145850854, NC_007328.5_26544467 and NC_007328.5_26544511) were significantly associated with all milk production traits. Our findings revealed majority of significant SNPs for milk traits were located within non-coding genomic regions.

Key words: Crossbred cattle, Milk production traits, Genome wide association study, Genotyping-by-sequencing

In livestock genomics, the current accessibility to large panels of SNPs by application of high-throughput genotyping techniques has given new impetus towards search for the underlying variation in complex traits through the use of genome-wide association studies (GWAS).The assumption in GWAS is that significant associations arise because SNP in linkage disequilibrium with, and therefore close to a causative mutation affecting the trait of interest. In dairy cattle, the GWAS of production traits, such as milk yield, fat yield, and protein yield, (Maxa *et al.* 2012, Minozzi *et al.* 2013, Nayeri *et al.* 2016, Van den Berg *et al.* 2016); functional or fertility traits, such as calving to the first service interval, days open, non-return rate, and conception rate (Olsen *et al.* 2011, Minozzi *et al.* 2013, Nayeri *et al.* 2016, Gaddis *et al.* 2016); health traits, such as clinical mastitis and udder type (Sahana *et al.* 2011, Flury *et al.* 2014); and longevity (Zhang *et al.* 2016), have been reported.

However, genome-wide association studies in Indian crossbred cattle with milk production traits is not available. Accordingly, GWAS was accomplished in 96 crossbred (Vrindavani) cattle for first lactation milk traits including average first lactation daily yield (AVDY), peak yield (PY), test day milk yield (TD36, TD66 and TD96) and first lactation 305 days milk yield (FL305DMY).

MATERIALS AND METHODS

Data: Phenotype data pertaining to first lactation records for AVDY, PY, TD36, TD66, TD96 and FL305MYwere recorded from 96 Vrindavani cows in their first lactation maintained under similar management regime from Cattle and Buffalo Farm, ICAR-Indian Veterinary Research Institute (IVRI), Izatnagar. Blood samples were collected from all 96 animals under optimum aseptic condition and further DNA was isolated and stored until downstream process for genotyping. The use and care of animal for this study purpose were in accordance with ethical standards laid by Institutional Animal Ethics Committee, IVRI.

*Adjustment for non-genetic factors:*Phenotypic data distributed over different seasons and year of first lactation was subsequently classified into three seasons and two periods. The records were adjusted for significant different non-genetic factors (period and season) affecting the traits.

Genotyping: Paired end genotyping was done using double digestion genotyping by sequencing (ddGBS) technique. Two restriction enzymes ApekI/PstI were used

Present address:[1]PhD Scholar (amirwar10@gmail.com), ICAR-National Dairy Research Institute, Karnal, Haryana. [2]Senior Scientist (vetamitchandan07@gmail.com), [3,4]PhD Scholar (vetakki10@gmail.com, karthikeyan0318@gmail.com), Animal Genetics Division; [5]Principal Scientist (triveniduttivri @gmail.com), Livestock Production and Management Division; [6]Principal Scientist (bpmishra_1@hotmail.com), Animal Biotechnology.

to digest the whole genome into several smaller fragments and subsequently these fragments were sequenced by parallel high throughput sequencing technology (Illumina NextSeq 500) and followed by execution of pipeline for SNP calling which broadly included filtering the high quality reads from raw data, assembly of high quality filtered reads and SNP mining using dDOCENT pipeline and finally mining out polymorphic SNP markers using GATK toolkit. SNP calling pipeline was followed after quality check of all individual samples which included quality filtering of all individual samples followed by mapping to *Bos taurus* 5.0.1 reference genome using Burrows-Wheeler Aligner algorithm. SNPs were called from read mapping using FreeBayes (version v0.9.10).

Genome wide association study: The statistical analysis of each SNP for association with the average yield, peak yield, test day and 305 days milk yield was accomplished using linear models in which the trait score is the y-variable and genotype is one of the explanatory variables.

$$y = x\beta_1 + \beta_0 + \varepsilon$$

y, a continuous valued phenotype; x: SNP genotype at a given locus; β_1, regression coefficient or the parameter that represents the strength of association between the SNP x and the phenotype y; β_0, intercept term; ε, noise or the part of y that is not explained by the SNP x (e.g., environmental effect).

The basic association analysis was accomplished using PLINK v1.9 for all the first lactation traits. The genomic-control corrected p-values were used for checking the possible association of SNPs with all the selected traits.The Manhattan plot were generated for each association reports through package "GWASTools" in R environment. The top ten significant ($P<0.001$) SNPs associated with each traits were listed out and a region around the variation spanning 20 kb was screened using UCSC genome browser for presence of any RefSeq genes.

RESULTS AND DISCUSSION

Milk yield is an economically important trait of dairy production systems, in which up to 40% of total variation is due to genetic nature. However, the milk yield per lactation is also affected by different non-genetic factors and it is associated with other lactation parameters such as initial yield, peak yield and persistency (Seangjun *et al.* 2009). In present study, data were analysed for the effect of different non-genetic factors *viz.* period and season. The Least squares means of overall population mean ± SE (kg) for AVDY, PY, TD36, TD66, TD96 and FL305MY were 9.23±0.23, 15.66±0.11, 11.56±0.42, 11.72±0.37, 11.13±0.34 and 2825.86±114.86 kg, respectively (Table 1).

The effect of season of calving (summer, rainy and winter) was significant on TD36, TD66 and TD96 hence were adjusted using least-square constants, whereas it was non-significant for FL305MY, AVDY and PY. However, no significant effect was observed for any of the trait for period under the study.

The high throughput Double Digestion Genotyping by Sequencing (ddGBS) data generated 1717311 single nucleotide polymorphism (SNPs) called from the original sequenced data using FreeBayes (v0.9.10). SNPs with minor allele frequency less than 0.01 and genotyping rate <80% were omitted for GWAS study. Only bi-allelic loci were retained in the study and all other forms were excluded to ensure an unbiased association. 9638 SNPs were retained for ascertaining GWAS with first lactation milk production traits. At $P<0.001$, total of 23, 28, 112, 3, 13 and 5 SNPs were found to be significantly associated with AVDY, PY, TD36, TD66, TD96 and FL305MY, respectively. Most of the significant SNPs were present on chromosome 1, 3, 6, 11, 13, 27, 29 and non-autosomes. The genomic-control corrected p-values were used instead of unadjusted p-values to depict the true association of SNPs with first lactation

Table 1. Least square means for 305 days milk yield (FL305DMY), average daily yield (AVDY), peak yield (PY) and test day (TD36, TD66 and TD96)

Effect	305DMY		AVDY		PY		TD36		TD66		TD96	
	N	Mean±SE (kg)	N	Mean±SE (kg)	N	Mean±SE (kg)	N	Mean±SE (kg)	N	Mean±SE (kg)	N	Mean±SE (kg)
Overall	96	2825.86± 114.86	96	9.23± 0.23	96	15.66± 0.11	96	11.56± 0.42	96	11.72± 0.37	96	11.13± 0.34
Season												
Rainy	17	2697.41^{a}± 261.04	17	8.89± 0.46	17	15.11± 0.69	17	10.05^{a}± 0.84	17	10.52^{a}± 0.74	17	9.60^{a}± 0.68
Winter	18	3119.40^{b}± 150.97	18	9.88± 0.43	18	16.60± 0.64	18	11.75^{b}± 0.17	18	13.57^{b}± 0.69	18	13.30^{b}± 0.68
Summer	61	2877.90^{ab}± 0.38	61	8.91± 55.31	61	15.18± 0.37	61	11.47^{c}± 0.45	61	11.08^{c}± 0.39	61	10.42^{c}± 0.36
Period												
2006-2011	29	2839.12± 176.27	29	9.32± 0.36	29	15.21± 0.54	29	11.45± 0.66	29	11.94± 0.57	29	11.15± 0.53
2012-2017	67	2956.83± 120.01	67	9.13± 0.25	67	16.12± 0.38	67	11.66± 0.46	67	11.50± 0.40	67	11.12± 0.37

Different superscripts in the same column across each effect indicate a statistically significant difference ($P<0.05$).

milk yield traits and the results are depicted under Manhattan plot in Fig. 1, 2, PY and FL305MY, respectively. The chromosomal coordinates of top ten significant SNPs along with RefSeq genes and their biological function, those were present within 20 kb of these SNPs were analysed and the search for QTL near the SNPs associated with various traits revealed that most of the significant SNPs were in close proximity with milk yield, protein yield and fat yield QTLs.

Association of SNPs for Average Daily Yield: A total of 139 significant SNPs (P<0.01) were found scattered across different chromosome with maximum number of SNPs on chromosome 1. When genomic-control applied the number of SNPs associated were found to remained same P<0.01. Top ten SNPs base pair position were enlisted in table 2. The 20 kb region around top ten significant SNPs showed presence of *KIRREL3* on chromosome 29*,* and *TSPEAR, KRTAP10-8,* and *KRTAP12-2* on chromosome 1.

Association of SNPs for Peak Yield: At P<0.01, 97 significant SNPs were found to be associated with highest number of SNPs on chromosome number 1. With genomic-control corrected p-value the number of SNPs associated at 0.01 remained same (Fig.1). The genome scan for genes within 20kb revealed presence of *ADCY3*, *CENPO*, and *PTRHD1* on chromosome 11, *KIRREL3* on chromosome 29*, LOC617218, KRTAP12-2, TSPEAR* and *LRRC3* on chromosome 1 (Table 2).

Association of SNPs for Test Day 36, 66 and 96: For TD36, TD66 and TD 96,the total SNPs associated at P<0.01 were 199, 101 and 134 SNPs, respectively. Maximum number of associated SNPs were found to be on BTA27, BTA 29 and BTA 29 for TD36, TD66 and TD 96, respectively.

When P-value was corrected for genomic control the number of SNPs associated at P<0.01 remained same for TD36, dropped to 30 from 101 SNPs for TD66 and only 122 SNPs were found to be associated with TD 96. *TSPEAR* and *KRTAP12-2* on chromosome 1, *DCAF6* on chromosome 3, *KIRREL3* on chromosome 29*, OTOR* on chromosome 13 were found to be present within 20kb region of top 10 significant SNPs associated with TD 36. For TD66, genome scan for genes within 20kb revealed presence of *RNF122* on chromosome 27, *KIRREL3* on chromosome 29, *LOC780781* and *LRRC3* on chromosome 1. For TD 96, the 20kb region around top ten significant SNPs showed presence of *KIRREL3* on chromosome 29.

Association of SNPs for First Lactation 305 Milk Yield: The SNPs (P<0.01) associated with FL305MY was 174, but with genomic-control corrected p-value the number of SNPs associated at 0.01 dropped to 55 (Fig. 2). The genome scan for genes within 20kb revealed presence *KRTAP10-8* and *LRRC3* on chromosome 1 (Table 2).

Distribution of the top 100 most significant SNP effects for production, health and reproduction traits showed that milk yield associated SNPs were mostly present on chromosome 10, 13 and X (Cole *et al.* 2011) which was in concurrence with our finding. GWAS performed by Li (2012), in Dutch Holstein-Friesian for two lactations, found 128 SNPs significantly associated (P<0.0001) with milk yield, located on BTA 2, 4, 6, 7, 10, 11, 14, 21 and 28, which was in accordance with present results. Similarly GWAS accomplished for milk production traits in Chinese Holstein Population (Jiang, 2010) showed 20 significant SNPs (P<0.05) on BTA 1, 3, 5, 9, 14 and 26 were associated with milk yield. On the similar lines Cole *et al.* (2011) performed GWAS for thirty one traits including milk yield in Holstein cows and revealed that milk, fat, and protein yields had tendency of sharing common SNP effects which

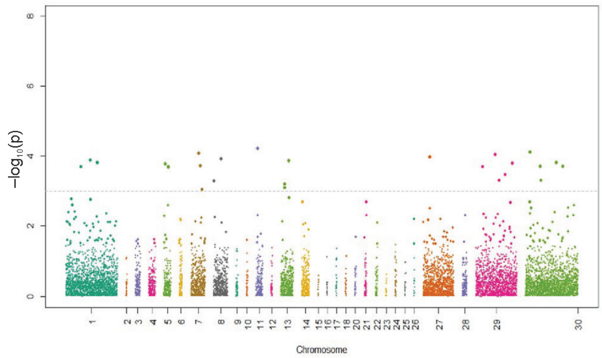

Fig. 1. Manhattan plot depicting associated loci with peak yield.

Table 2. Top ten SNPs associated with AVDY, PY, FL305MY, TD36, TD66 and TD96

AVDY		PY		FL305MY		TD36		TD66		TD96	
BTA	Base pair position	BTA	Base pair position	BTA	Base pair position	BTA	Base pair position	BTA	Base pair position	BTA	Base pair position
27	26544511	11	74471034	1	145846847	1	146002710	27	26544467	29	30463646
29	30574052	30	883	27	26544467	3	706410	3	27243821	1	145846774
27	26544467	7	56824723	1	145867706	30	750	6	19994346	27	26544467
1	145958213	7	56824773	4	29092872	27	26544482	6	19994304	1	145850854
1	145958259	29	30522196	6	19994478	29	30548371	27	28813613	29	30538438
5	53934973	27	26544467	27	29011190	30	816	29	30463646	29	30564710
1	145941640	8	74988152	27	37799336	27	26544467	1	145846774	3	27243821
1	145995375	1	146008313	1	145939378	14	40788582	1	145899623	21	1978663
1	145850854	13	83412347	6	19994404	13	10778387	29	30548371	11	107548897
14	300232	1	145866000	1	145890340	27	29087954	1	145899645	27	26544482

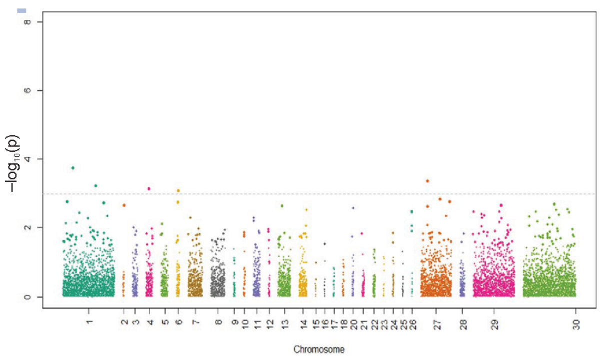

Fig. 2. Manhattan plot depicting associated loci with first lactation 305 days milk yield

was in accord with this study through the finding that base pair positions of most significant SNPs were close to QTLs affecting milk yield, protein yield and fat yield. In other GWAS for lactation characteristics in Thai multibreed dairy cattle (Yodklaew *et al.* 2017) found most significant ($P<10^{-8}$) SNPs associated with 305days milk yield were annotated on chromosome number 9, 14, 29, 13, 20, 29, 8, 26, 7 and 17 while for peak yield 145 SNPs were found in significant association, located on 13, 16, 24, 11, 3, 29, 10. Saowaphak *et al.* (2017) investigated the putative QTL associated to LPL, DO, and FM305 using GWAS in 36 animals genotyped by using the Illumina Bovine SNP50 Bead Chip and found that five QTLs associated with FM305 contained two genes on chromosomes BTA8 (*GNA14*) and BTA15 (*LRRC4C*) and some QTLs were also found on chromosomes 3, 4, 5, 6, 7, 12, 13, 14, 16, 18, 19, 23, and 29. Genome wide association for milk production and female fertility traits in Canadian Holestein cattle demonstrated that in total 292 SNPs were found to be significant for milk production at genome wide level localized on BTA 5, 6, 14, 15, 20 (Nayeri *et al.* 2016). Only limited research on the X chromosome is available, compared to the extensive literature on autosomes. Sermyagin *et al.* (2017) performed the GWAS for milk production traits in Russian Holstein and black-and-white cattle population for traits like milk yield for 305 days of lactation, milk fat and protein yield and found the highest number of highly significant ($P<10^{-6}$) SNPs for MY on chromosomes 1, 2, 3, 11, 17, and 23. Among all chromosomes, BTA14 had received a wide attention by above investigators.

Apart from a large number of QTL reported on BAT14, the well-known *DGAT1* gene located at <0.44Mb is generally accepted as a major gene affecting milk production. Under the study, an apparent feature of finding is that a large proportion of significant SNPs are located on BTA14. In particular, all segments on BTA14 which harbour multiple SNPs for the milk traits also harbour the *DGAT1* gene. When 20 kb region around the top ten significant SNPs were searched for presence of any RefSeq genes, it revealed the presence of *LRRC3, KRTAP10-8, KIRREL3, TSPEAR, KRTAP12-2, ADCY3, CENPO, PTRHD1, LOC617218, DCAF6, OTOR, RNF122,* and *LOC780781* gene locus within the specified region. Genome–wide association studies (Yue *et al.* 2017) suggested some new candidate genes for milk production traits in Chinese Holstein cattle viz. *EEF2K* for 305 day milk yield and *SLC22A8* for lactation persistency which we were not able to find in our study. The SNPs under the study were either present within the coding or non-coding regions of genes or near proximity to these genes. Majority of the SNPs were located within/nearby *KIRREL3* on chromosome 29, followed by *LRRC3* and *TSPEAR* on chromosome 1. Three SNPs namely NC_007299.6_ 145850854, NC_007328.5_ 26544467 and NC_007328.5_ 26544511 were significantly associated ($P<0.01$) with all milk production traits under present investigation. However, our findings showed that majority of significant SNPs for milk traits are found within non-coding regions of the genome could be confounded to genotyping technique which was also reported by Ibeagha-Awemu *et al*. (2016) while accomplishing GWAS in 1246 Canadian Holstein Cows using GBS.

The scanty number of SNPs associated with traits under current study may be ascribed to low depth of sequencing which were not uniform across the animals. Thepresence of majority of SNPs significantly associated with the traits were found within the non-coding region of the genome. A larger number of SNPs markers are necessary to explain genetic variation in these complex milk production traits. However, with proper depth of sequencing, the ddGBS could be an efficient way for search of novel and relevant SNPs markers associated with production traits both among breeds and within various breeds of Indian origin. The gains from the incorporation of significant genome wide SNPs for estimation of genomic breeding value will increase the accuracy of selection in crossbreds at the early age in dairy cattle breeding thus it will pave a way for genomic selection in the country.

ACKNOWLEDGEMENTS

This study was supported by the ICAR-IVRI with the necessary facilities. The authors would also wish to acknowledge the Director and Joint Director of Academic and Research, IVRI for their kind support to carry out this work.

REFERENCES

Cole J B, Wiggans G R, Ma L, Sonstegard T S, Lawlor T J, Crooker B A, Van Tassel C P, Yang J, Wang S, Matukumalli L K and Da Y. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics* **12**(1): 408.

Flury C, Boschung C, Denzler M and Bapst B. 2014. Genome-wide association study for 13 udder traits from linear type classification in cattle. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17–22 August 2014. Vancouver, BC, Canada.

Gaddis K P, Null D J and Cole J B. 2016. Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. *Journal of Dairy Science* **99**(8): 6420–35.

Harvey W R. 1990. *User's guide for LSMLMW mixed model least* *square and maximum likelihood computer program (PC-2 version)*. Ohio State University Press, Columbus.

Ibeagha-Awemu E M, Peters S O, Akwanji K A, Imumorin I G and Zhao X. 2016. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Scientific Reports* **6**: 31109.

Jiang L, Liu J, Sun D, Ma P, DingX, Yu Y and Zhang Q. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS ONE* **5**(10): 13661.

Li F. 2012. Genome-wide association analysis for 305-day milk production traits in dairy cattle. University of Natural Resources and Life Sciences, Vienna.

Maxa J, NeuditschkoM, Russ I, Förster M and Medugorac I. 2012. Genome-wide association mapping of milk production traits in Braunvieh cattle. *Journal of Dairy Science* **95**(9): 5357–64.

Minozzi G, Nicolazzi E L, Stella A, Biffani S, Negrini R, Lazzari B, Ajmone-Marsan P and Williams J L. 2013. Genome wide analysis of fertility and production traits in Italian Holstein cattle. *PLoS ONE* **8**(11): 80219.

Nayeri S, Sargolzaei M, Abo-Ismail M K, May N, Miller S P, Schenkel F, Moore S S and Stothard P. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genetics* **17**(1): 75.

Olsen H G, Hayes B J, Kent M P, Nome T, Svendsen M, Larsgard A G and Lien S. 2011. Genomewide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. *Animal Genetics* **42**(5): 466–74.

Sahana G, Guldbrandtsen B and Lund M S. 2011. Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. *Journal of Dairy Science* **94**: 479–86.

Saowaphak P, Duangjinda M, Plaengkaeo S, Suwannasing R and Boonkum W. 2017. Genetic correlation and genome-wide association study (GWAS) of the length of productive life, days open, and 305-days milk yield in crossbred Holstein dairy cattle. *Genetics and Molecular Research* **16**(2).

Seangjun A, Koonawootrittriron S and Elzo M A. 2009. Characterization of lactation patterns and milk yield in a multibreed dairy cattle population in the Central Thailand. *Kasetsart Journal (Natural Science)* **43**: 74–82.

Sermyagin , Gladyr , Plemyashov K V, Kudinov A, Dotsev A V, Deniskova and Zinovieva N A. 2017. Genome-Wide Association Studies for Milk Production Traits in Russian Population of Holstein and Black-and-White Cattle. Proceedings of the conference "Research and Development-2016". pp. 591–99. Springer International Publishing. Russia.

Van den Berg I, Boichard D and Lund M S. 2016. Comparing power and precision of within-breed and multibreed genome-wide association studies of production traits using whole-genome sequence data for 5 French and Danish dairy cattle breeds. *Journal of Dairy Science* **99**(11): 8932–45.

Yodklaew P, Koonawootrittriron S, Elzo M A, Suwanasopee T and Laodim T. 2017. Genome-wide association study for lactation characteristics, milk yield and age at first calving in a Thai multibreed dairy cattle population. *Agriculture and Natural Resources* **51**(3): 223–30.

Yue S J, Zhao Y Q, Gu X R, Yin B, Jiang Y L, Wang Z H and Shi K R. 2017. A genome-wide association study suggests new candidate genes for milk production traits in Chinese Holstein cattle. *Animal Genetics* **48**(6): 677–81.

Zhang Q, Guldbrandtsen B, Thomasen J R, Lund M S and Sahana G. 2016. Genome-wide association study for longevity with whole-genome sequencing in 3 cattle breeds. *Journal of Dairy Science* **99**(9): 7289–98.