



# Genome-Wide Association Studies in Indian Buffalo Revealed Genomic Regions for Lactation and Fertility

Vikas Vohra<sup>\*†</sup>, Supriya Chhotaray<sup>†</sup>, Gopal Gowane, Rani Alex, Anupama Mukherjee, Archana Verma and Sitangsu Mohan Deb

Buffalo Breeding Lab, Animal Genetics and Breeding Division, Indian Council of Agricultural Research-National Dairy Research Institute, Karnal, India

## OPEN ACCESS

### Edited by:

Mudasir Ahmad Syed,  
Sher-e-Kashmir University  
of Agricultural Sciences  
and Technology, India

### Reviewed by:

Gao Huijiang,  
Institute of Animal Sciences, Chinese  
Academy of Agricultural Sciences,  
China

Hadi Atashi,  
Shiraz University, Iran  
Alessandro Bagnato,  
University of Milan, Italy

### \*Correspondence:

Vikas Vohra  
vohravikas@gmail.com

<sup>†</sup> These authors have contributed  
equally to this work and share first  
authorship

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 16 April 2021

Accepted: 16 August 2021

Published: 20 September 2021

### Citation:

Vohra V, Chhotaray S, Gowane G,  
Alex R, Mukherjee A, Verma A and  
Deb SM (2021) Genome-Wide  
Association Studies in Indian Buffalo  
Revealed Genomic Regions  
for Lactation and Fertility.  
*Front. Genet.* 12:696109.  
doi: 10.3389/fgene.2021.696109

Murrah breed of buffalo is an excellent dairy germplasm known for its superior milk quality in terms of milk fat and solids-not-fat (SNF); however, it is often reported that Indian buffaloes had lower lactation and fertility potential compared to the non-native cattle of the country. Recent techniques, particularly the genome-wide association studies (GWAS), to identify genomic variations associated with lactation and fertility traits offer prospects for systematic improvement of buffalo. DNA samples were sequenced using the double-digestion restriction-associated DNA (RAD) tag genotyping-by-sequencing. The bioinformatics pipeline was standardized to call the variants, and single-nucleotide polymorphisms (SNPs) qualifying the stringent quality check measures were retained for GWAS. Over 38,000 SNPs were used to perform GWAS on the first two principal components of test-day records of milk yields, fat percentages, and SNF percentages, separately. GWAS was also performed on 305 days' milk yield; lactation persistency was estimated through the rate of decline after attaining the peak yield method, along with three other standard methods; and breeding efficiency, post-partum breeding interval, and age at sexual maturity were considered fertility traits. Significant association of SNPs was observed for the first principal component, explaining the maximum proportion of variation in milk yield. Furthermore, some potential genomic regions were identified to have a potential role in regulating milk yield and fertility in Murrah. Identification of such genomic regions shall help in carrying out an early selection of high-yielding persistent Murrah buffaloes and, in the long run, would be helpful in shaping their future genetic improvement programs.

**Keywords:** Murrah, buffaloes, GWAS, lactation persistency, milk yield, fertility

## INTRODUCTION

Buffalo (*Bubalus bubalis*) is an imperative livestock species and act as a key component for improving agricultural economy and supplying milk, meat, and draft power. The buffalo population across the world was recently estimated to be 194 million, 97% of which were present in Asia (FAO, 2014). Buffalo is well known for its high milk quality, with higher fat (6.4–8.0% vs. 4.1–5.0%) and protein (4.0–4.5% vs. 3.4–3.6%) contents than cow milk (Venturini et al., 2014; Khedkar et al., 2016). The concentration of these milk constituents offers a higher economic return of the buffalo milk and increases the demand of value-added products like mozzarella. Buffaloes are an

integral and crucial genetic resource in Indian dairy industry, contributing 49% to the total milk produced according to the basic animal husbandry statistics, 2019<sup>1</sup>. About 63% of global buffalo milk production and 95% of Asian buffalo milk production is contributed by Indian buffaloes (FAO, 2014). Buffaloes are more adaptable to harsh environments and often resist various bovine tropical diseases. However, the poor reproductive efficiency of buffaloes limits its potential. Buffaloes exhibit higher age at puberty and maturity, longer postpartum breeding interval, and low conception rates (Gordon, 1996; Warriach et al., 2015; Raina et al., 2016). Buffaloes in the field condition also suffer from short lactation of 252–270 days as compared to the standard 300 days (Ganguli, 1981; Griffiths, 2010). This underlines the scope for improving the milk production and fertility potential through implementation of systematic breeding programs for buffaloes in the country. However, in India, very few works have been done in order to identify the infinitesimally large number of underlying loci regulating the expression of complex traits such as milk production, lactation persistency and fertility.

Genome wide association studies (GWAS) in buffaloes for lactation traits have been mainly limited to the use of bovine single-nucleotide polymorphism (SNP) chip (Wu et al., 2013; Venturini et al., 2014) and Affymetrix's buffalo 90K SNP chip (de Camargo et al., 2015; Iamartino et al., 2017; Liu et al., 2018). There have been very few GWAS conducted in India covering all aspects of production and reproduction performance due to constraints of cost incurred and organized large-scale genotyping programs. High-density SNP panels are a prerequisite for GWAS, which have led to developments of cost-effective and efficient next-generation sequencing (NGS) technologies such as reduced representation of genomic libraries (RRLs) (Van Tassel et al., 2008). The flexibility, robustness, and low cost of double-digestion restriction-associated DNA (RAD) tag genotyping-by-sequencing (ddRAD-GBS) technique renders it suitable for identification of SNPs in any species for GWAS (Davey et al., 2011; Elshire et al., 2011).

Hence, the present study was conducted to identify novel SNPs associated with milk production, composition, lactation persistency, and fertility traits at the genomic level using the genotype-by-sequencing technique in Murrah buffalo, India's major buffalo breed and milch animal of the nation.

## MATERIALS AND METHODS

### Sampling, Data Recording, and Genotyping

A total of 672 test-day records on each trait, i.e., milk yield (TDMY), fat percentage (TDFP), and solids-not-fat (SNF) percentage (TDSNF) were collected from 96 female Murrah buffaloes reared at LRC, NDRI, Karnal, India (29.68°N and 76.99°E). Records of 96 buffaloes on 305 days' milk yield (305DMY), birth weight (bwt), age at

first calving (AFC), calving interval (CI), and age at sexual maturity (ASM) in months were collected. Other traits such as lactation persistency, postpartum breeding interval in days (PPBI), and breeding efficiency (BE) were derived from the primary phenotype records. Breeding efficiency was calculated for female buffaloes as described by Tomar (1965).

Lactation persistency was estimated by following four different methods:

- (I) Wood (1967) incomplete gamma function, where individuals were classified as persistent and non-persistent as described by Macciotta et al. (2005) considering a positive rate of incline as a favorable condition and a negative rate as an unfavorable condition for persistency
- (II) Johansson and Hansson (1940) method
- (III) Ludwick and Petersen (1943) method
- (IV) Mahadevan (1951) method

DNA was isolated from 96 Murrah buffaloes selected for the study following the standard phenol-chloroform method (Sambrook and Russell, 2006). Samples were further processed using the standard RAD protocol as described by Peterson et al. (2012). DNA double digestion was carried out with *SphI* and *MluCI* restriction enzymes. Adapters (P1 and P2) were prepared as per standard Illumina read multiplexing protocol using an inline barcode along with Illumina index for library preparation. After adapter ligation and size selection, samples were sequenced on a Illumina Hi-Seq 2000 platform.

### Variant Calling

The NGS pipeline was standardized after incorporating a few modifications in the standard mpileup variant calling pipeline (Li, 2011), to call variants present in the Murrah population. The Mediterranean buffalo genome, having accession ID GCF\_003121395.1, was retrieved from the NCBI dataset and used as a reference genome. Index and sequence dictionary files were created using the Burrows–Wheeler algorithm (BWA) (Li and Durbin, 2010) and PicardTools,<sup>2</sup> respectively. The quality of paired-end raw FASTQ files generated after sequencing was checked using FastQC (Andrews, 2010), and each report was combined through MultiQC (Ewels et al., 2016). Adapters were marked and trimmed using bbmap (Brian, 2014).<sup>3</sup> The BWA-MEM algorithm was used to align the trimmed FASTQ sequences with the reference genome. Aligned files were coordinate-sorted, and duplicate reads were removed. Read group identifiers were updated using PicardTools. The quality of aligned BAM files was checked using qualimap (García-Alcalde et al., 2012). Variants were called using bcftools-mpileup (Li, 2011).

### Quality Control of Variants

Only biallelic SNPs having more than 95% genotyping rate were retained for further GWAS. SNPs with a MAF < 0.05 and

<sup>1</sup><https://dahd.nic.in/about-us/divisions/statistics>

<sup>2</sup><https://github.com/broadinstitute/picard>

<sup>3</sup><https://sourceforge.net/projects/bbmap/>

deviating from the Hardy–Weinberg equilibrium at  $p < 0.0001$  were removed from the dataset. SNPs in LD with  $r^2 > 0.8$  were also removed. Only autosomal and X chromosome SNPs were retained for the final analysis. All the quality control operations and data preprocessing were performed using PLINK v1.9 (Chang et al., 2015).

## Statistical Model

GWAS for milk yield, fat percentage, and SNF percentage were conducted on the principal components (PCs) instead of direct traits. Principal component analysis (PCA) was performed in the R programming environment (v4.0.3) on the 672 test-day records of each trait separately. As the first two PCs cumulatively account for most of the variation in the dataset, they were selected to be GWAS traits. The traits on which GWAS were performed were PCs (PC<sub>1</sub> and PC<sub>2</sub>) of TDMYs, TDFPs, TDSNFs, and 305DMY; lactation persistency calculated by four different methods; age at sexual maturity in months; postpartum breeding interval (in days); and breeding efficiency. Genome-wide identity-by-state (IBS) for all pairs of individuals was checked. Multidimensional scaling (MDS) based on SNP information was done to check for the presence of any population stratification (**Supplementary Figure 1**) and was corrected by incorporating the first two MDS components as covariates in the model for GWAS. A genome-wide scan for significant SNPs considering only additive effects was accomplished through a simple regression model using PLINK v1.9 as described by Marees et al. (2018), where residuals were assumed to be normally and independently distributed. A linear regression model was fitted for determining the association between SNPs and continuous traits (Bush and Moore, 2012), while logistic regression was fitted for the binary trait (lactation persistent/non-persistent based on incomplete gamma function). The threshold for genome-wide significance was determined by correcting the  $p$ -values of the SNP association test with Benjamini–Hochberg’s false discovery rate (FDR) at 5 and 10% levels (Benjamini and Hochberg, 1995) using the “R” package fuzzySim v3.0 (Barbosa, 2020). A genome-wide significant threshold was set at FDR 5% and a suggestive threshold at 10% by calculating a nominal  $p$ -value for the largest index  $i$  for which  $P_{(i)} \leq (i/m) \times q$ , where  $i$  = rank of the SNPs,  $m$  = no. of individual tests performed, and  $q$  = either 0.05 or 0.1 (Glickman et al., 2014). The results were plotted as Manhattan plots and Q-Q plots using the “qqman” package of R.

### Linear regression model used for GWAS:

$$y = \beta_0 + x\beta_1 + C_1\beta_2 + C_2\beta_3 + e$$

where  $y$  = trait,  $x$  = additive effect of SNPs,  $C_1$  = first component of MDS,  $C_2$  = second component of MDS,  $\beta_0$  = intercept term,  $\beta_1$  = regression coefficient representing the strength of association between SNP  $x$  and trait  $y$ ,  $\beta_2$  = regression coefficient of  $C_1$ ,  $\beta_3$  = regression coefficient of  $C_2$ , and  $e$  = residuals or noise not explained by SNPs.

The model used for GWAS on 305DMY, however, included four covariates consisting of the first component of MDS, birth weight, age at first calving, and calving interval.

## SNP Mapping and Pathway Enrichment

SNPs were identified as genic if present within the genes or intergenic if present within a range of 20 kb from the 5′ and 3′ ends of the gene (da Costa Barros et al., 2018). ARS-UCD1.2/bosTau9 cow assembly was used as the reference genome to identify regions around significant SNPs in the UCSC genome browser (Zimin et al., 2009). Gene ontology (GO) was carried out using gProfiler<sup>4</sup> and the GO classifications significant over the Benjamini–Hochberg FDR were selected for pathway enrichment through Cytoscape v3.8.2 (Shannon et al., 2003).

## RESULTS

### PCA on Test-Day Records

PCA on test-day records proved that the first two PCs cumulatively explain 77.72, 40.77, and 48.06% of the total variation in TDMYs, TDFPs, and TDSNFs, respectively. Eigen values of PC<sub>1</sub> for TDMYs, TDFPs, and TDSNFs were found to be 4.45, 1.67, and 2.30, respectively, while for PC<sub>2</sub> they were 0.98, 1.20, and 1.05, respectively. New synthetic variable PC<sub>1</sub> and PC<sub>2</sub> for the above three traits were constructed for each buffalo utilizing original variables and variable loadings of PCA.

### Genotyping

An average of 1.3 million each of forward and reverse reads were obtained per sample. The total number of forward and reverse end reads was 252.56 million with the average read length being 151 base pairs (bp). After quality check of raw data, it was observed that the average GC content was 50.54%. The rate of duplication was quite high, i.e., 80.11%, which was later checked during variant calling. The quality score (Q) throughout the dataset varied from 35 to 40.

### Variant Calling and Quality Control

Raw files that were quality checked in FastQC were combined using MultiQC, and the report is provided as **Supplementary Document**. After alignment, BAM files were obtained, and quality was checked. Of the raw reads, 98.11% were mapped accurately to the Mediterranean buffalo reference genome. After removing the duplicate reads, the rate of duplication was reduced to 15.85%. Average mapping quality after alignment was found to be 29.05. A single VCF file was obtained, with 3,854,990 SNPs, out of which 3,792,469 SNPs which were strictly biallelic were retained for further analysis. After applying quality control constraints, 38,560 SNPs present on autosomes and X-chromosomes were retained for further downstream

<sup>4</sup><https://biit.cs.ut.ee/gprofiler/gost>

**TABLE 1** | Details of the top 10 SNPs identified through GWAS on various traits and genes identified through genome scan.

Trait	Chr <sup>†</sup> no.	Position (bp)	p-value	Within	±20 kb
PC <sub>1</sub> of milk yield	X	7,588,358*	$1.91 \times 10^{-07}$	<i>GRIA3</i>	
	9	62,699,319*	$2.39 \times 10^{-06}$	<i>ZNF292</i>	
	16	74,362,861*	$3.56 \times 10^{-06}$		
	1	182,059,836	$1.05 \times 10^{-05}$	Uncharacterized <i>LOC112444602</i>	
	6	35,648,660	$1.31 \times 10^{-05}$		<i>TIGD2</i>
	9	48,559,942	$1.38 \times 10^{-05}$	<i>GRIK2</i>	
	1	97,587,987	$2.68 \times 10^{-05}$	<i>LRRC34</i>	<i>LRRC34, ACTRT3</i>
	20	45,808,430	$2.69 \times 10^{-05}$		
	2	72,321,021	$3.65 \times 10^{-05}$		
	X	120,205,266	$4.13 \times 10^{-05}$		
PC <sub>1</sub> of fat percentage	6	34,234,112	$1.96 \times 10^{-05}$	<i>CCSER1</i>	
	20	54,742,629	$3.10 \times 10^{-05}$		
	17	15,895,247	$5.70 \times 10^{-05}$		
	4	1,579,792	$9.68 \times 10^{-05}$		
	16	66,495,537	0.0001022		
	14	27,391,536	0.0001391		
	10	45,942,651	0.0001592		<i>DAPK2</i>
	16	46,080,207	0.0001604	<i>CAMTA1</i>	
	11	30,281,532	0.0001669		
	4	39,053,011	0.0001739		
PC <sub>1</sub> of SNF percentage	14	4,664,0635	$6.39 \times 10^{-06}$		
	9	104,155,086	$6.57 \times 10^{-06}$		
	10	4,420,352	$1.03 \times 10^{-05}$		<i>TICAM2</i>
	11	88,922,443	$1.34 \times 10^{-05}$		
	12	90,113,650	$1.64 \times 10^{-05}$		
	3	146,125,164	$1.69 \times 10^{-05}$		
	8	99,790,321	$2.11 \times 10^{-05}$		<i>TXN</i>
	14	38,810,289	$2.32 \times 10^{-05}$	<i>HNF4G</i>	
	14	54,721,238	$2.68 \times 10^{-05}$	<i>SYBU</i>	
	17	42,173,707	$3.79 \times 10^{-05}$	Uncharacterized <i>LOC104974614</i>	
PC <sub>2</sub> of milk yield	1	81,353,445	$4.36 \times 10^{-06}$		
	7	57,678,016	$2.86 \times 10^{-05}$	<i>TCERG1</i>	
	4	134,421,018	$4.95 \times 10^{-05}$		
	16	17,134,516	$5.34 \times 10^{-05}$		
	1	49,359,549	$7.78 \times 10^{-05}$		
	4	26,233,428	0.0001182		
	13	60,176,598	0.0001579		<i>ANGPT4</i>
	18	63,008,459	0.0001706		
	2	85,823,516	0.0001802	<i>ANKRD44</i> , uncharacterized <i>LOC112442949</i>	
	21	42,799,270	0.0002008		<i>AKAP6</i>
PC <sub>2</sub> of fat percentage	3	81,728,404	$9.30 \times 10^{-05}$	<i>ROR1</i>	
	15	58,026,658	$9.67 \times 10^{-05}$		<i>CCDC34</i>
	14	5,954,275	0.0001596		
	6	49,443,269	0.0004445		
	23	25,106,184	0.0004533		<i>GSTA1, GSTA2</i>
	15	31,296,423	0.0004619	<i>GRIK4</i>	
	18	61,661,554	0.0004811	<i>CACNG6</i>	
	4	134,458,024	0.0004867		
	7	42,605,598	0.0005285	<i>SH3BP5L, ZNF672</i>	
	16	74,400,826	0.0005499		
PC <sub>2</sub> of SNF percentage	14	71,295,596	$3.83 \times 10^{-05}$		<i>TRIQK</i>

(Continued)

TABLE 1 | (Continued)

Trait	Chr <sup>†</sup> no.	Position (bp)	p-value	Within	±20 kb	
305 days' milk yield	12	100,102,349	$4.23 \times 10^{-05}$			
	1	58,132,034	0.0002283	<i>CFAP44</i>		
	20	53,548,085	0.0003363	<i>CDH18</i>		
	4	15,355,186	0.0003396			
	6	27,219,584	0.0003841		Uncharacterized <i>LOC782977</i>	
	22	19,933,711	0.000389			
	5	21,940,540	0.0003954			
	21	56,124,462	0.0004159			
	9	42,153,396	0.0004165	<i>SEC63</i>		
	10	5,804,150	$2.93 \times 10^{-05}$			
	2	85,890,123	$3.90 \times 10^{-05}$		<i>ANKRD44</i>	
	9	62,699,319	$4.52 \times 10^{-05}$	<i>ZNF292</i>		
	12	10,144,550	$5.91 \times 10^{-05}$			
	X	7,588,358	$6.08 \times 10^{-05}$	<i>GRIA3</i>		
	Persistency based on Wood's function as described by Macciotta	16	74,362,861	$6.59 \times 10^{-05}$		
10		5,804,363	$8.51 \times 10^{-05}$			
19		65,955,791	0.0001318			
1		97,587,987	0.0001527	<i>MYNN</i>	<i>LRRC34, ACTRT3</i>	
9		48,559,942	0.000153	<i>GRIK2</i>		
14		2,765,427	0.000693	<i>DENND3</i>		
9		72,904,775	0.000787			
3		25,896,446	0.001048			
10		98,483,390	0.001208			
21		32,145,325	0.001243	<i>PSTPIP1</i>		
5		118,217,420	0.001301			
19		66,731,726	0.001385			
6		111,661,570	0.001661			
1		101,471,188	0.00172			
Persistency (Mahadevan, 1951)		5	36,948,158	0.001727	<i>ADAMTS20</i>	
	12	61,269,055	0.0003472			
	18	46,542,221	0.0005129	<i>PRODH2, NPHS1, RREL2, 42466</i>		
	10	66,128,637	0.000711			
	5	126,657,828	0.0008095			
	14	46,873,549	0.0008564			
	22	33,269,149	0.001113	<i>FAM19A1</i>		
	12	72,874,869	0.001206			
	23	19,210,873	0.001381	<i>CLIC5</i>		
	12	32,859,855	0.001488	<i>GPR12</i>		
	13	60,835,727	0.00154	<i>DEFB125</i>		
	Persistency (Johansson and Hansson, 1940)	14	45,071,226	$1.69 \times 10^{-05}$		
		21	37,776,399	$2.76 \times 10^{-05}$		
		9	62,472,303	$3.36 \times 10^{-05}$	<i>CFAP206</i>	
		18	46,100,792	0.0001026		
18		46,100,619	0.0001026			
18		17,868,492	0.0001162	<i>C18H16orf78</i>		
2		20,133,507	0.0001273			
3		55,760,382	0.0001561			
4		49,950,395	0.0001709			

(Continued)

TABLE 1 | (Continued)

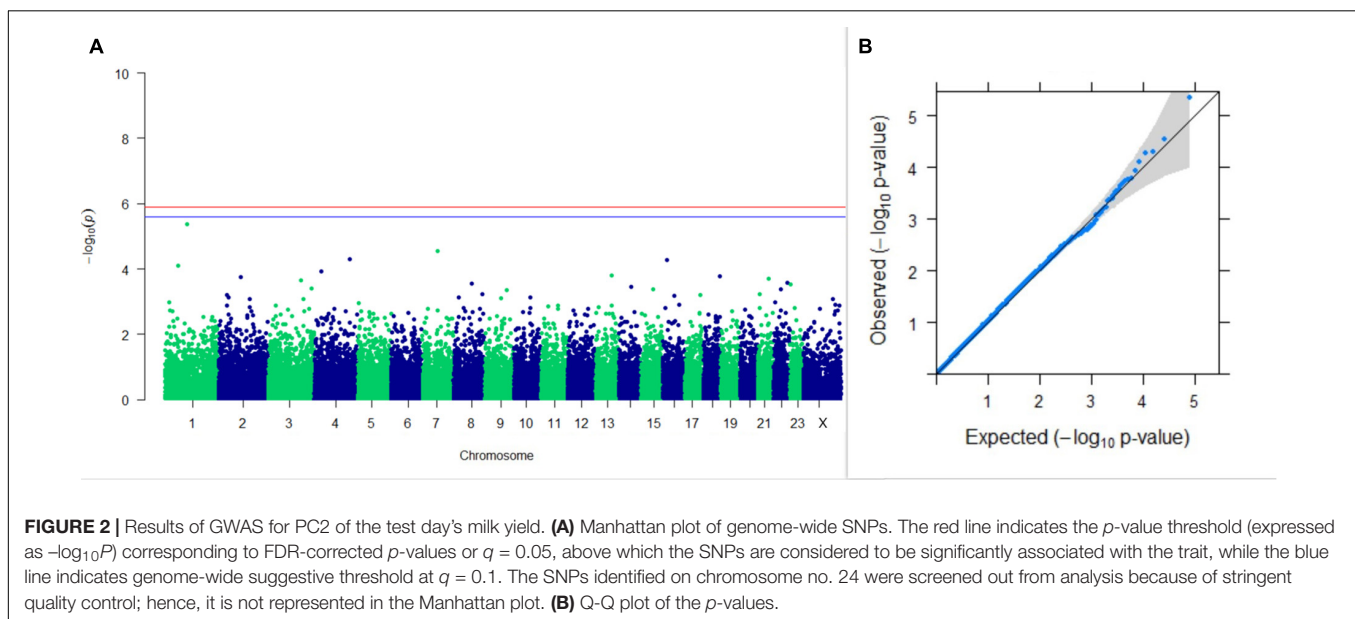
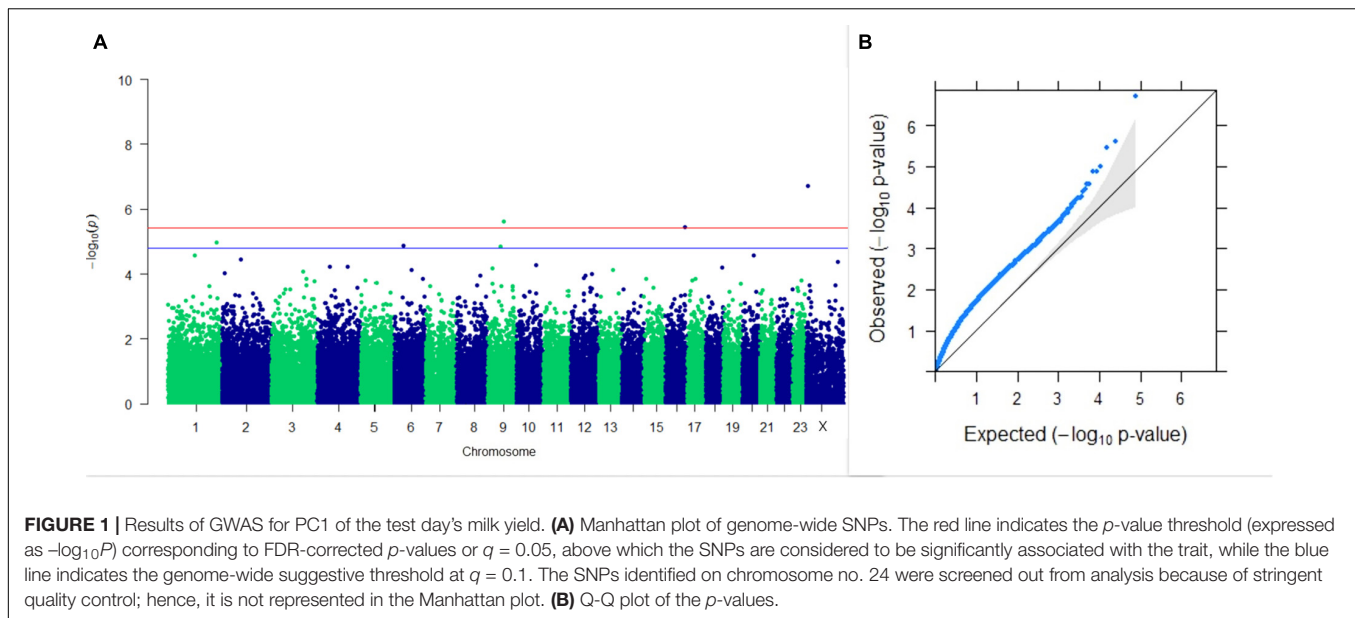
Trait	Chr <sup>†</sup> no.	Position (bp)	p-value	Within	±20 kb
Persistence (Ludwick and Petersen, 1943)	15	11,079,156	0.000183		
	6	2,687,0445	$6.07 \times 10^{-05}$	STPG2	
	9	62,472,303	$8.75 \times 10^{-05}$	CFAP206	
	18	46,100,792	0.0001368		
	18	46,100,619	0.0001368		
	4	46,668,425	0.0001908	RINT1, EFCAB10	
	10	65,940,920	0.0002065		
	7	57,678,016	0.0002898	TCERG1	
	25	40,090,282	0.0003082		
	10	35,000,168	0.0003138		
Breeding efficiency (Tomar, 1965)	18	17,868,492	0.0003109	C18H16orf78	
	10	1,310,267	$1.70 \times 10^{-05}$	APC, uncharacterized LOC112448352	
	11	83,985,064	$6.50 \times 10^{-05}$		
	12	54,602,811	0.0003442	NDFIP2	
	2	157,056,510	0.0003871		
	X	68,984,061	0.0004459		
	1	40,267,041	0.000448		
	7	117,169,472	0.0005262		
	20	5,188,573	0.0005834	Uncharacterized LOC107131404	
	1	182,894,045	0.0006526		
Age at sexual maturity (in months)	14	70,378,719	0.0009191		PDP1
	2	19,116,988	$2.31 \times 10^{-05}$	PDE11A	
	7	77,438,396	$5.80 \times 10^{-05}$		
	18	19,894,177	0.0001711		
	3	9,170,745	0.0001738	SLAMF6	
	17	58,295,542	0.0001793		uncharacterized LOC104974658
	6	32,022,004	0.0001861	GRID2	
	1	2,246,590	0.0003256	uncharacterized LOC104970778	
	15	75,434,743	0.0003302		
	1	96,574,489	0.0003995	EIF5A2, RPL22L1	
Postpartum breeding interval (in days)	10	69,569,504	0.0004181		
	2	25,677,863	$4.28 \times 10^{-05}$	ERICH2	
	4	31,943,618	$5.96 \times 10^{-05}$	IGF2BP3	MALSU1
	14	31,151,162	0.0001547	PPP1R42	TCF24
	2	69,670,013	0.0002049		CCDC93
	6	31,285,865	0.0002096	GRID2	
	11	39,293,274	0.0002278		
	7	84,151,709	0.0002888	EDIL3	
	7	84,151,714	0.0002947	EDIL3	
	13	11,218,416	0.0003107	LOC112449367	
7	17,313,196	0.0003504		Uncharacterized LOC101904981, LOC112447353	

<sup>†</sup>Chr = chromosome, \*SNPs at the positions were significantly associated with the trait at the 5% level of significance using Benjamini–Hochberg's FDR.

analysis with the final genotyping rate of 98.24%. All the SNPs present on autosome no. 24 failed to pass quality control constraints; hence, in the final analysis, the 24th autosome has been removed.

## GWAS Results

GWAS were performed with 38,560 SNPs, and the genome-wide significant threshold was set at the 5% FDR level. GWAS were performed on the two foremost



variation explaining PCs (PC<sub>1</sub> and PC<sub>2</sub>) of TDMY, TDFP, and TDSNF.

### GWAS on Lactation Traits

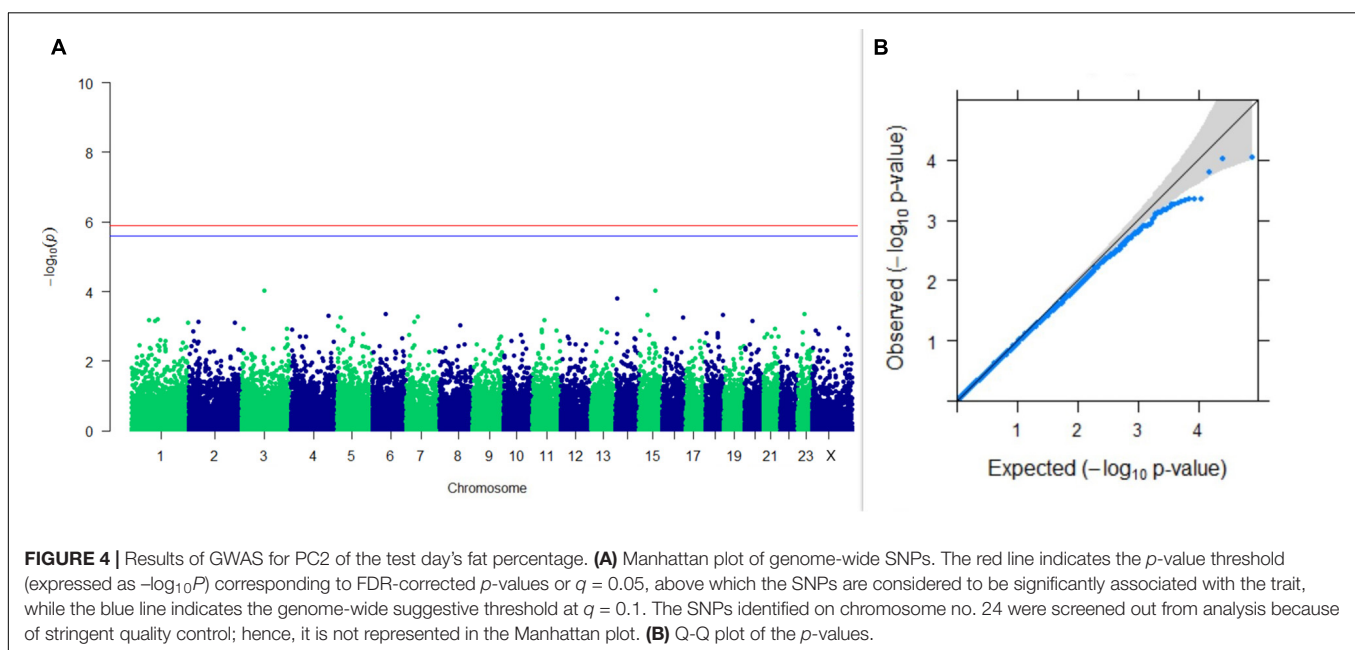
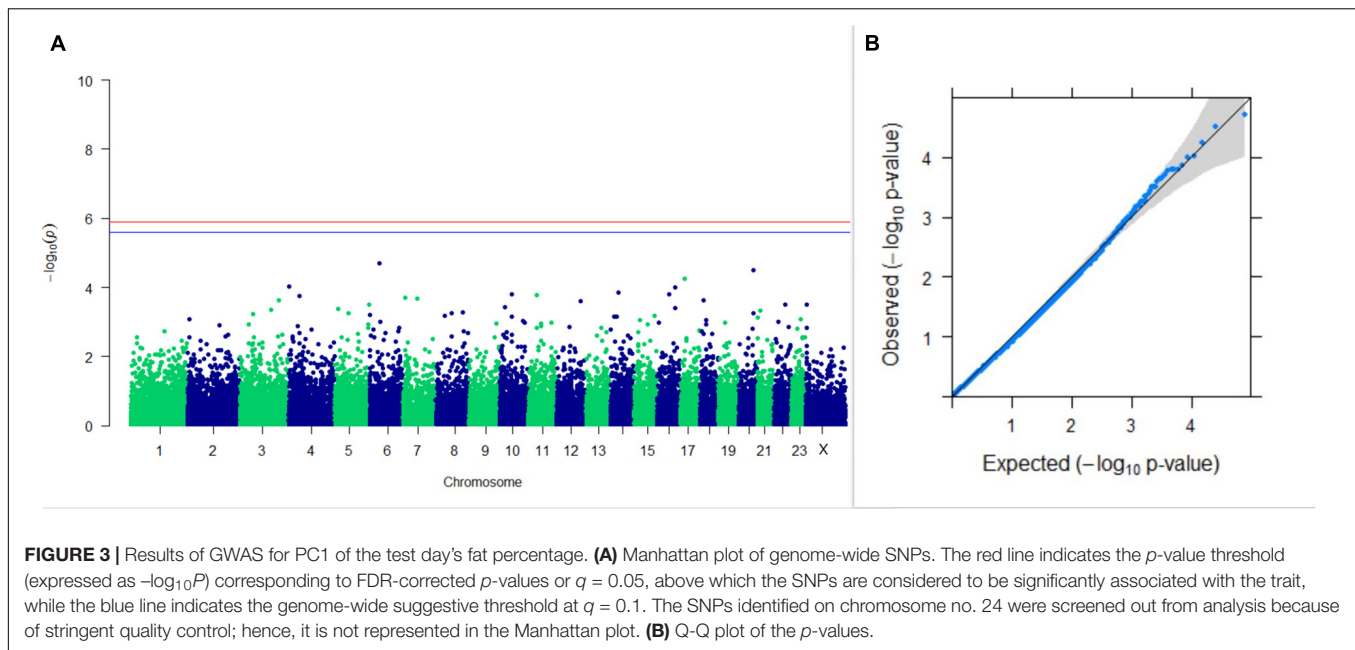
Three SNPs present on chromosomes X (7588358 bp), 9 (62699319 bp), and 16 (74362861 bp) were significantly associated with PC<sub>1</sub> of TDMYs with FDR-corrected  $p$ -values of 0.007369, 0.04579, and 0.04579, respectively. No SNPs were found to be significantly associated with PC<sub>2</sub> TDMYs. GWAS on PC<sub>1</sub> and PC<sub>2</sub> of TDFPs and TDSNFs also could not establish any significant association between the traits and SNPs. However, the top 10 SNPs having the lowest  $p$ -values in the test of association with different lactation traits, along with their position and genes within a  $\pm 20$ -kb region, are listed in **Table 1**. The

Manhattan and Q-Q plots (panels A and B, respectively) for association results of PC<sub>1</sub>-TDMYs, PC<sub>2</sub>-TDMYs, PC<sub>1</sub>-TDFPs, PC<sub>2</sub>-TDFPs, PC<sub>1</sub>-TDSNFs, and PC<sub>2</sub>-TDSNFs are given in **Figures 1–6**, respectively.

GWAS on 305DMY revealed that there was no significant association between SNPs and the trait. It was observed that five SNPs that appeared in the list with GWAS for PC<sub>1</sub> and PC<sub>2</sub> of the test day's milk yield were also in the top SNP list for 305DMY. The Manhattan and Q-Q plots for association results of 305DMY are given in **Figures 7A,B**, respectively.

### GWAS on Lactation Persistency

GWAS on lactation persistency estimated by four different methods revealed no significant association between SNPs and

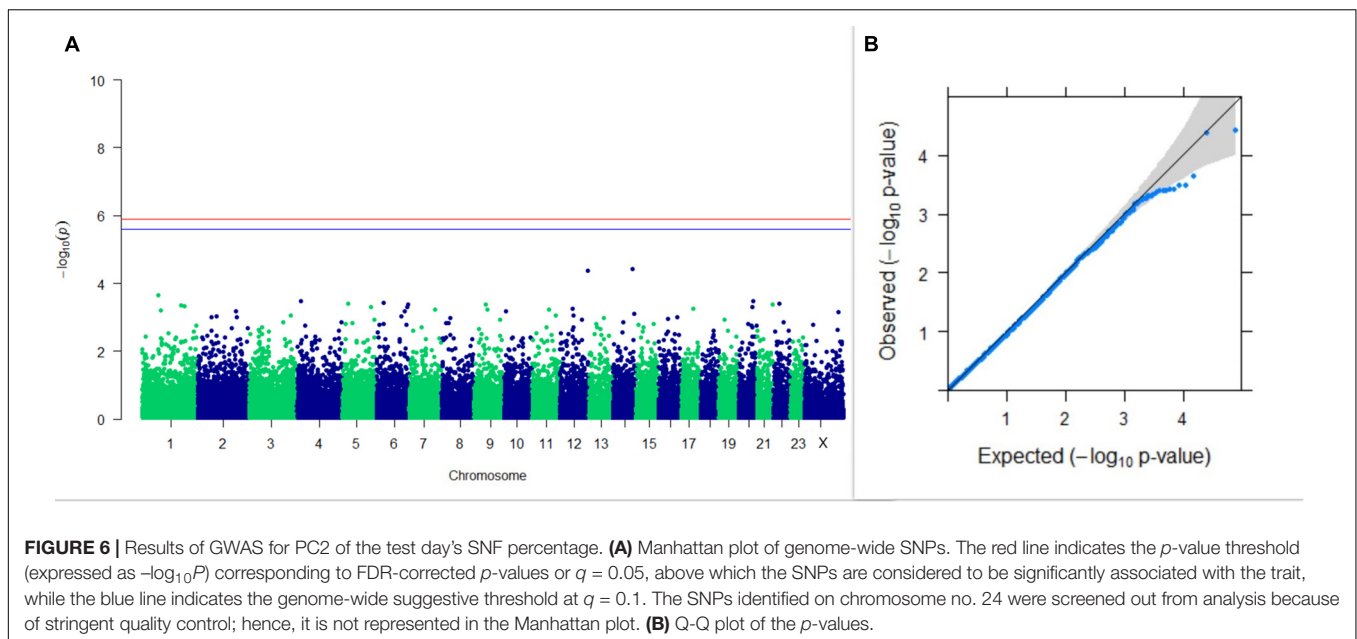
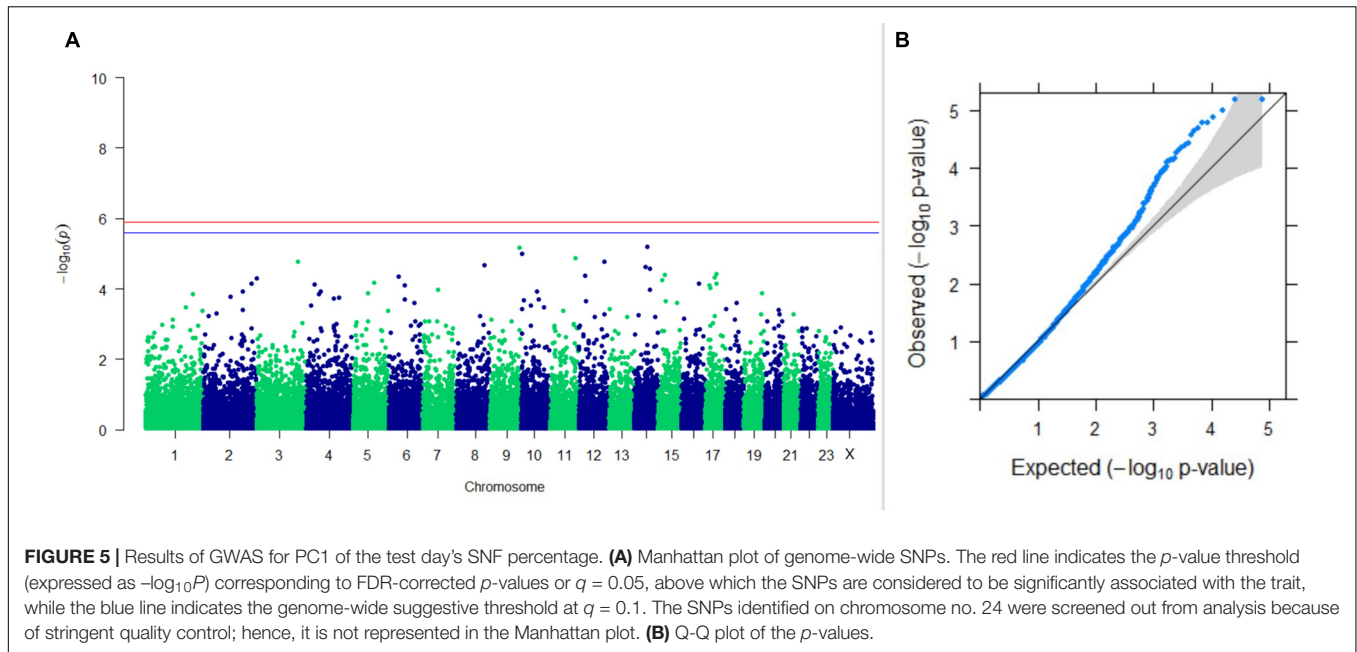


the trait generated. The top 10 SNPs having the lowest  $p$ -values along with their genomic position and genes within a  $\pm 20$ -kb region are listed in **Table 1** for all four methods. It was observed that four SNPs were found in common between the top SNPs listed for persistency estimated by methods II and III. Among the four SNPs, one was on chromosome 9 at 62472303 bp, and three were present on chromosome 18 at 46100792, 46100619, and 17868492 bp. The Manhattan and Q-Q plots (panels A and B, respectively) for association results of persistency estimated by four different methods (I, II, III, and IV) are given in **Figures 8–11**, respectively.

### GWAS on Fertility Traits

Upon performing GWAS on age at sexual maturity, postpartum breeding interval, and breeding efficiency, no significant association of any SNPs were observed for the traits; however, the top 10 SNPs with the lowest  $p$ -values and genes within a  $\pm 20$ -kb region are listed in **Table 1**. The Manhattan and Q-Q plots (panels A and B, respectively) of GWAS on age at sexual maturity, postpartum breeding interval, and breeding efficiency are given in **Figures 12–14**, respectively.



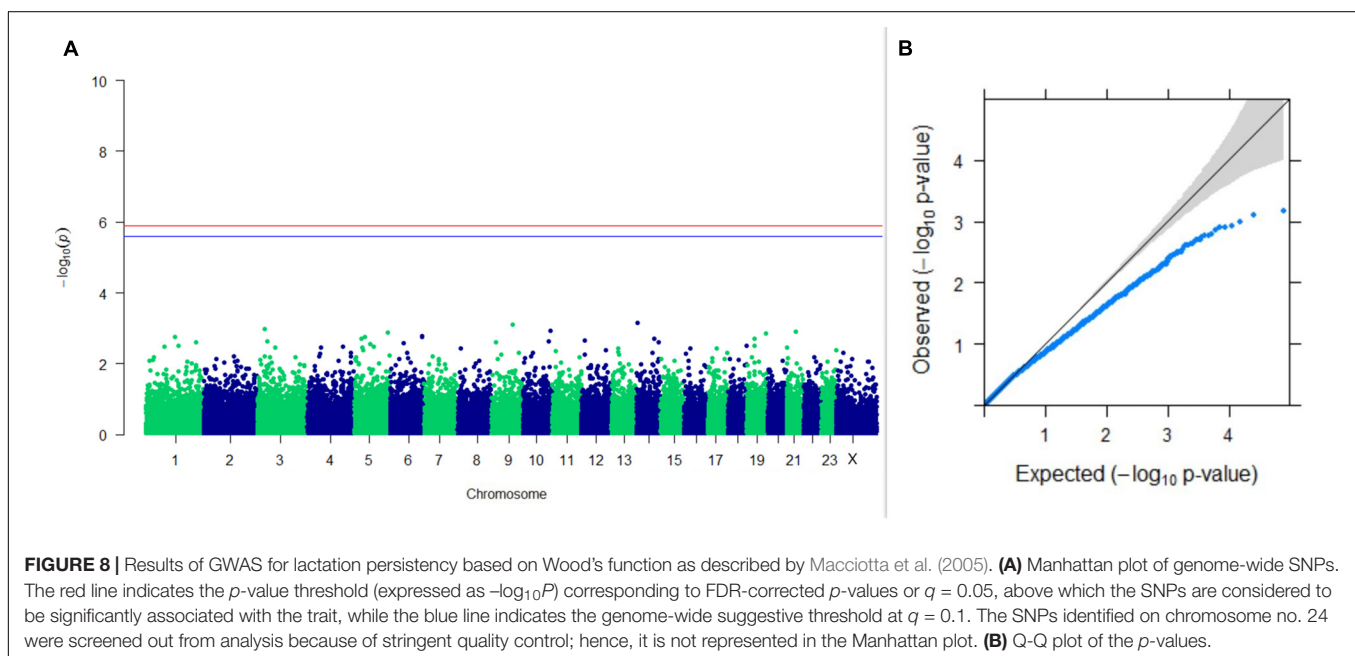
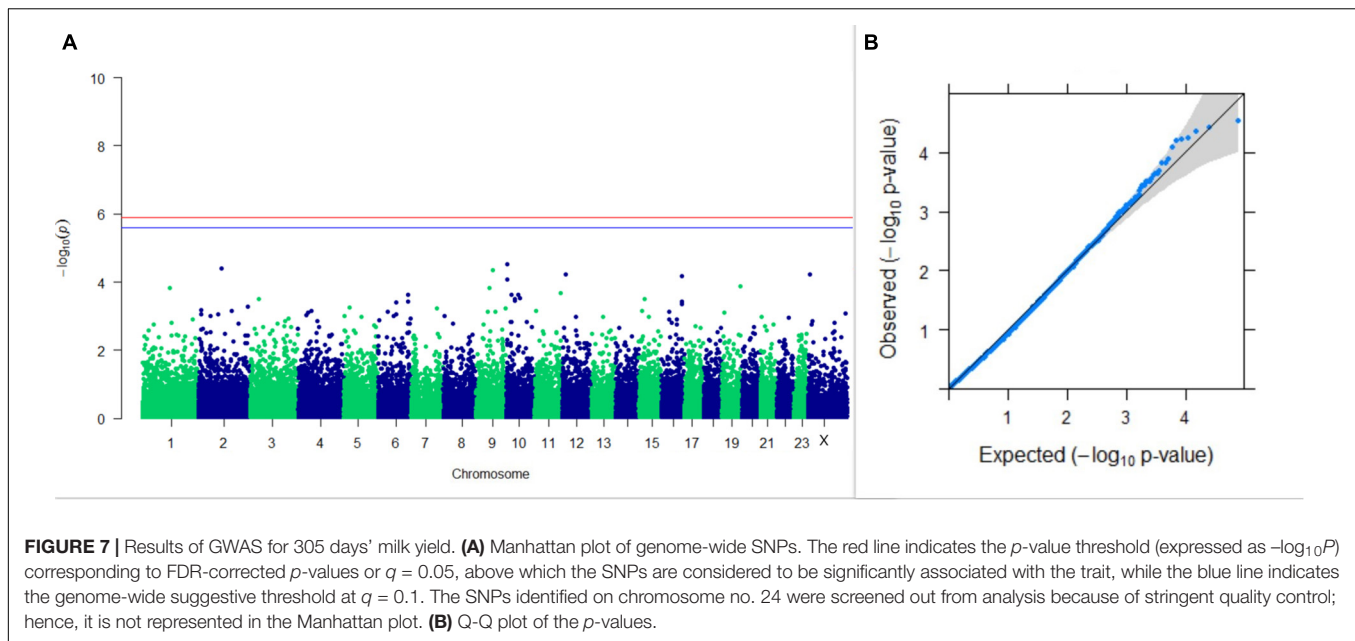


## DISCUSSION

The test-day model (TDM) as repeated measurements of milk yield traits is the method of choice for predicting 305 days' milk yield. TDMs account for environmental effects on each test day and is useful to model individual lactation curves (Schaeffer et al., 2000; Bignardi et al., 2012). PCA being a powerful multivariate technique is often used to predict 305 days' milk yield from the inter-related variables such as test-day records (Taggar et al., 2012). Wara et al. (2019) had performed GWAS on test day's milk yield in Vrindavani cattle; however, reports of GWAS on PCs of test days are very few. In the present study, PCA was applied

on seven test-day records of first lactation in order to perform GWAS on the PCs explaining maximum variation rather than the test-day records themselves. The first two PCs cumulatively explained 40–78% of total variation in test-day records of traits included, indicating their potential to be used as GWAS traits to identify novel SNPs for milk yield, fat percentages, and SNF percentages.

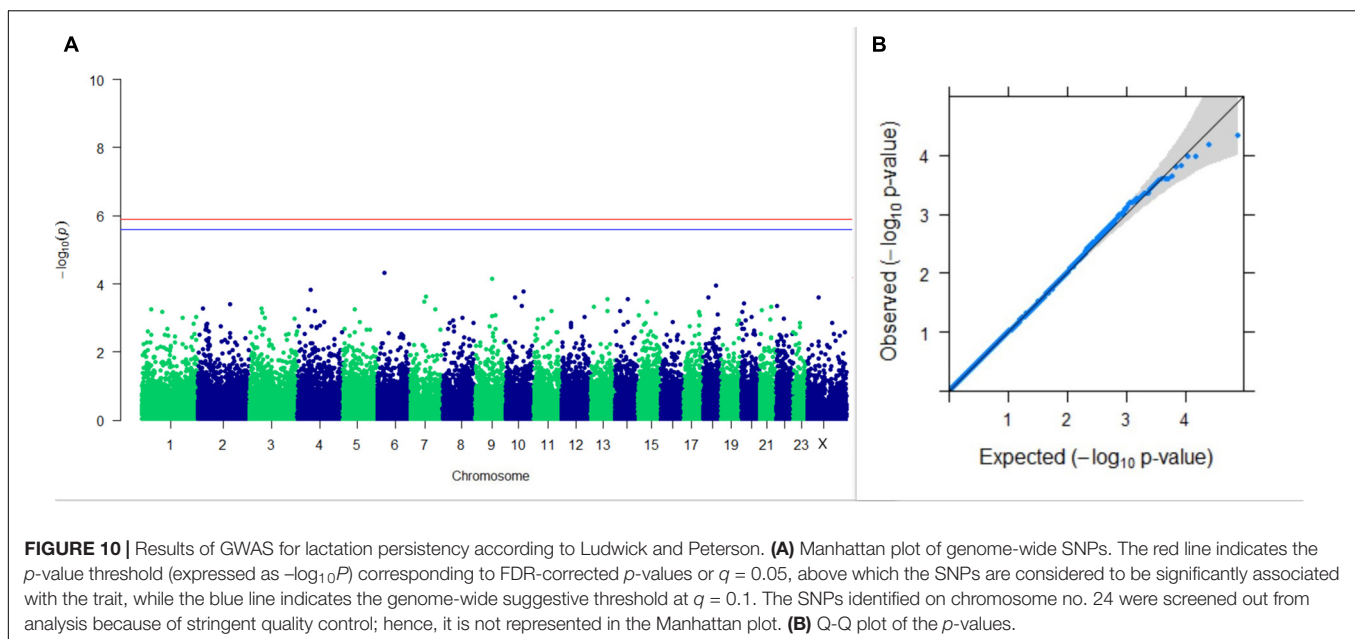
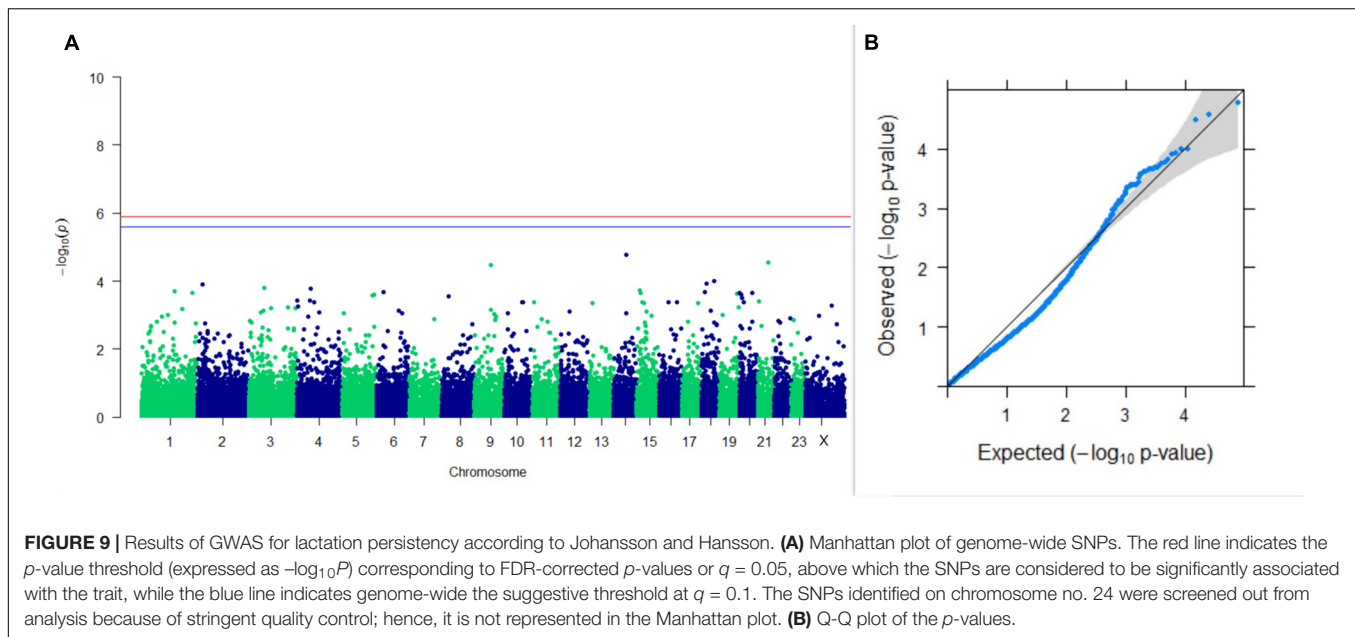
It is important to understand the reasons for considering 96 samples taken in the study as optimum, as in the buffalo farming scenario in Asia, particularly in India, the maximum number of buffaloes maintained at any large organized herd ranges from 250 to 500, with 100–200 breedable buffaloes having



complete phenotype information. The National Dairy Research Institute has the second-largest herd in the Indian Council of Agricultural Research (ICAR) with a well-managed herd of 250 breedable buffaloes. Furthermore, there is no buffalo sequencing consortium/project in operation (India). In such a case, one could afford this sample size to genotype with complete pedigree and with sufficient genetic diversity. However, the sample size is less for conducting GWAS and raises the question of whether the results are reliable enough. It was observed that the results obtained under the present study are encouraging and important, as one of the genes identified in the present study, i.e., *GRIA3*, was reported by Cole et al. (2011) in a sample of 1,654 US Holstein

cows. There are occasions where a similar sample size (96) has been used to understand body morphology traits through GWAS (Rahmatalla et al., 2018).

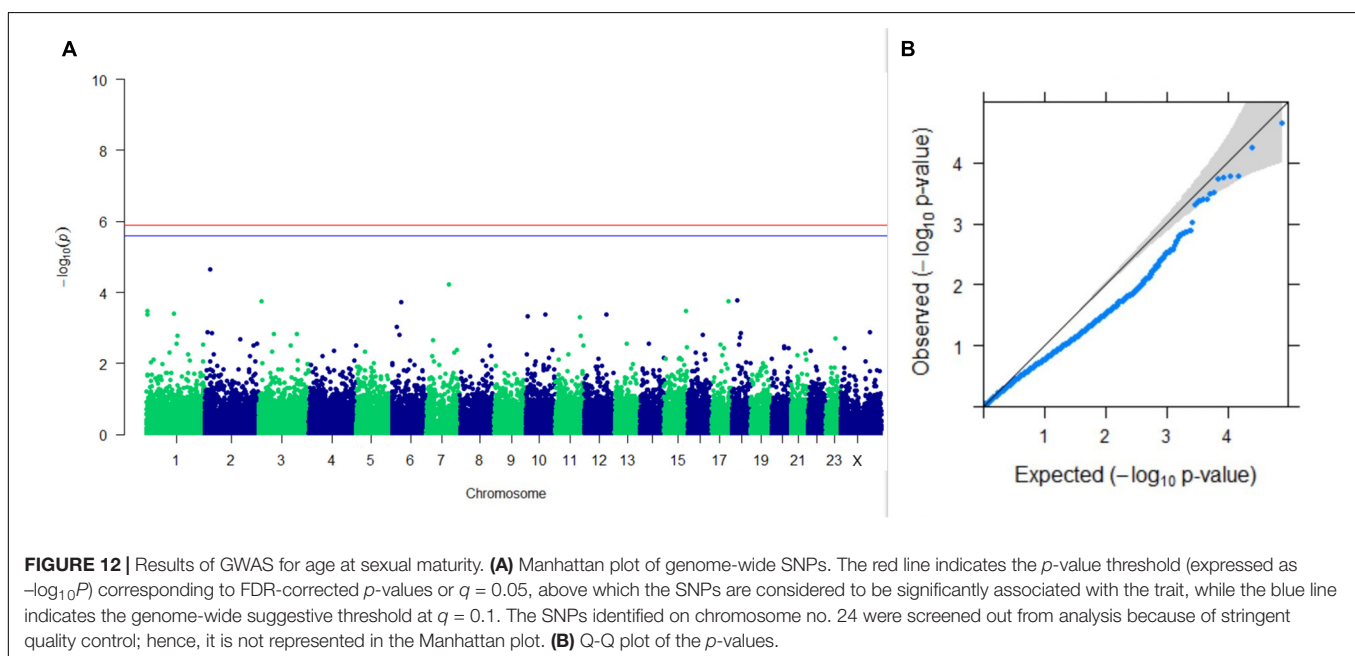
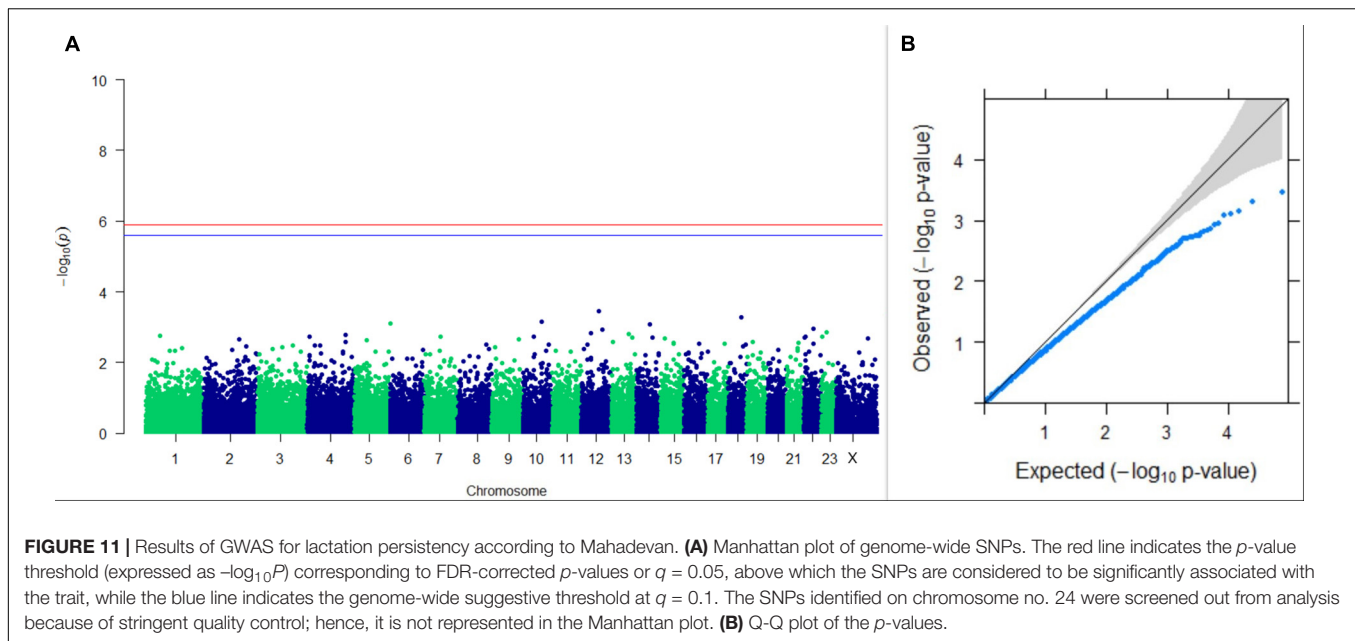
To identify novel SNPs associated with various economic traits, sequencing through the ddRAD approach was performed. Ruffalo et al. (2012) highlighted that a mapping quality of 15–40 is of the highest nature, and theoretical accuracy corresponds to  $\sim 100\%$ , while depending on the aligner, the actual accuracy of mapping or base call may vary from 40 to 60%. Similar mapping quality averaging 30 for all samples was obtained through BWA aligner, which usually implies good overall base quality of reads and few mismatches in the best possible alignment.



An earlier bovine SNP chip was used to study the traits of buffalo. Using a bovine SNP chip (Illumina BovineSNP50 BeadChip) for a GWAS in buffalo population, Wu et al. (2013) identified seven SNPs that were significantly associated with milk yield. Later on, Affymetrix's 90K SNP chip commercialization led to several GWAS on milk production traits in buffalo. de Camargo et al. (2015) in a GWAS on buffaloes reported that *LOC100847171*, *BCL6*, *RTP2*, *SST*, *PTGS2*, *LOC100295047*, and *LOC101908004* are associated with milk production; *KCTD8*, *LOC782855*, *LOC101904777*, *ESRRG*, *TRNAY-AUA*, and *GPATCH2* are associated with fat percentage; *LOC101903483*, *SART3*, *ISCU*, *CMKLRI*, *WSCD2*, *MFNG*, *CARD10*, and *USP18*

are associated with protein percentage. Liu et al. (2018) identified several candidate genes, namely, *MFSD14A*, *SLC35A3*, *PALMD*, *RGS22*, and *VPS13B*, for milk production in the Italian Mediterranean buffalo through GWAS. In another GWAS in buffalo, Iamartino et al. (2017) reported that genes regulating the D-glucose level in the blood affect milk production in buffalo.

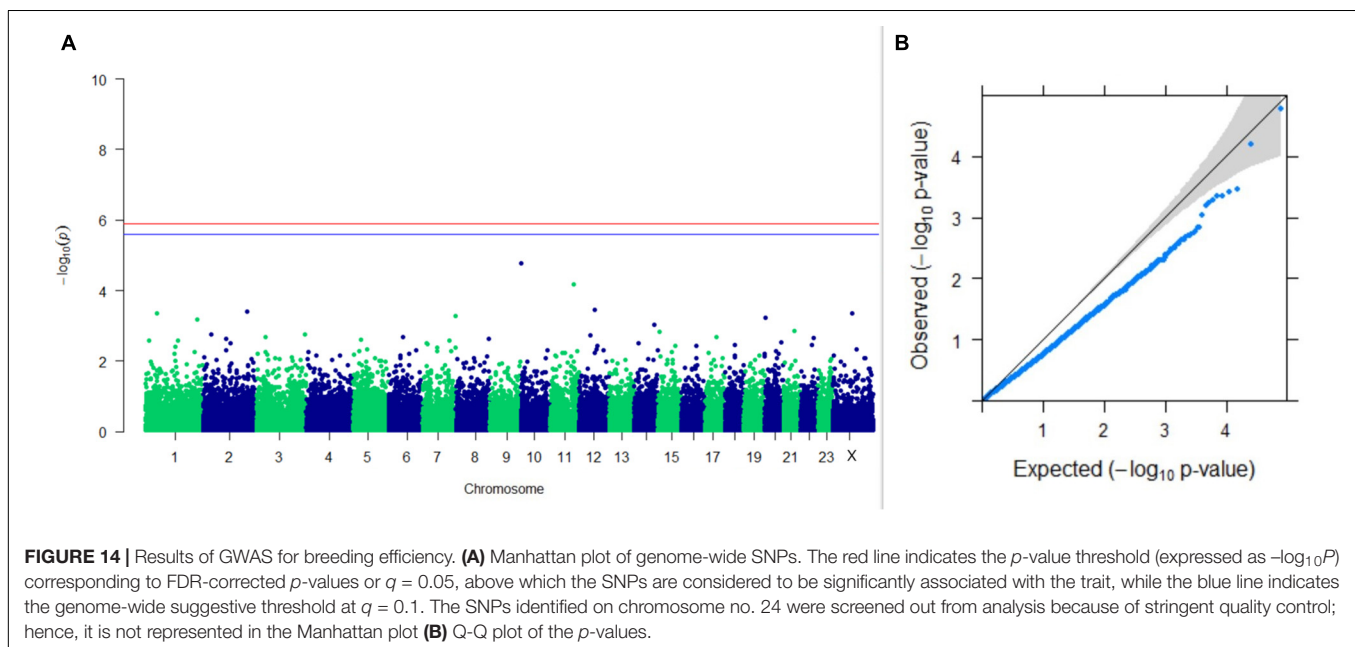
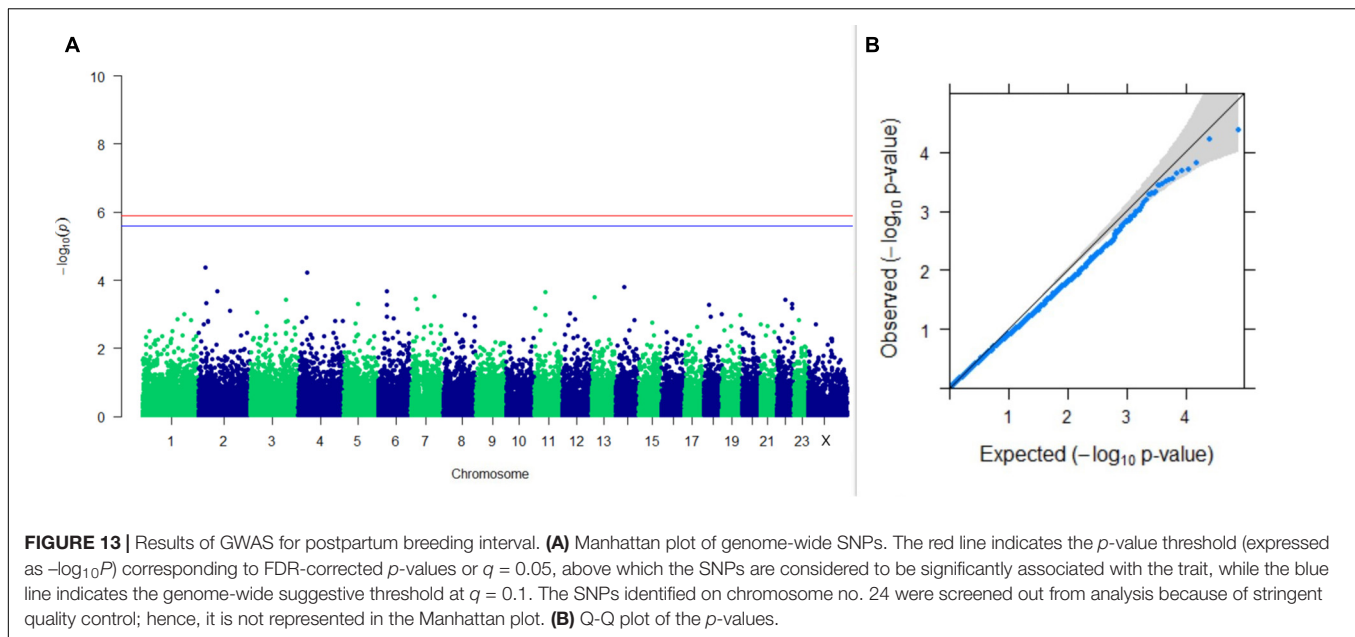
In the present study, GWAS performed on the milk yield PCs explaining maximum variation, which revealed that three SNPs were significantly associated with PC<sub>1</sub> of TDMY at the 5% FDR level. However, six SNPs present on chromosomes 1, 6, and 9 at 182,059,836, 35,648,660, and 48,559,942 bp, respectively, were above the suggestive



threshold of 10% FDR and are presented in **Table 1** as the top six SNPs associated with PC<sub>1</sub> of TDMY. However, for the rest of the traits studied, no SNPs were detected to be significantly associated (or as rejections) at 5% or even 10% FDR.

Upon scanning  $\pm 20$  kb around that SNP on the X chromosome, the *GRIA3* gene was found. Glutamate ionotropic receptor AMPA type subunit 3 (*GRIA3*) is reported to be very significantly associated with daughter pregnancy rate in US Holstein cows (Cole et al., 2011). In a study, a positive selection signal has been observed for the loci containing the *CNIH3* gene. *CNIH3* interacts with *GRIA1*, *GRIA2*, *GRIA3*, *GRIA4*, and

*GRIK1* belonging to a class of  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) glutamate receptors, thus regulating trafficking of AMPA receptors. AMPA receptors are known to participate in luteinizing hormone (LH) secretion (Utsunomiya et al., 2013). The nearest genes to the top 10 SNPs from PC<sub>1</sub> GWAS for milk yield were *ZNF292*, *LOC112444602*, *TIGD2*, *GRIK2*, and *LRRC34*. Similarly, for PC<sub>2</sub> of milk yield, GWAS revealed that *TCERG1*, *ANGPT4*, *ANKRD44*, *LOC112442949*, and *AKAP6* were the nearest genes from the top 10 SNPs. It could be observed that five SNPs present in the list of top SNPs from GWAS results of PC<sub>1</sub> and PC<sub>2</sub> of the test day's milk yield were also present in the top 10 SNP list



for 305 days' milk yield (**Table 1**), highlighting the efficacy of PCA in predicting 305 days' milk yield. Additionally, *MYNN* and *ACTRT3* were found nearest to the top SNPs of GWAS for 305 days' milk yield. Cai et al. (2020) have also reported the *ANKRD44* gene to be associated with milk yield in Nordic Holstein cattle.

A genome-wide scan for novel genes for fat percentage outlined *CCSER1*, *DAPK2*, *CAMTA1*, *ROR1*, *CCDC34*, *GSTA1*, *GSTA2*, *GRIK4*, *CACNG6*, *SH3BP5L*, and *ZNF672* as the nearest genes to the top SNPs obtained from GWAS. Kolbehdari et al. (2009) in a whole-genome study in Canadian Holstein cattle mapped the gene *DAPK2* to several milk production-related

QTLs. They also mapped *DAPK2* to a milk fat percentage QTL which concurs with the findings of our study. Genes other than *DAPK2* were not reported earlier to be associated with milk fat percentage in buffalo; however, *CAMTA1* may possibly have a role in fatty acid metabolism.

*TICAM2*, *TXN*, *HNF4G*, *SYBU*, *LOC104974614*, *TRIQK*, *CFAP44*, *CDH18*, *LOC782977*, and *SEC63* were found near the top SNPs associated with PCs of TDSNF. We could not find any previous reports stating the role of these genes in regulating milk SNF percentage; however, a functional enrichment analysis study by Zhou et al. (2019) revealed *HNF4G* and *SYBU* as candidate genes that regulate milk fat synthesis, transport, and

metabolism. All the genes involved with milk yield and its composition identified through the study were enriched for pathways in Cytoscape. A sub-network of the genes is depicted in **Supplementary Figure 2**. It can be observed that most of the genes identified belong to the glutamate receptor family, involved in the regulation of the neuronal system.

In Canadian Holstein cattle, Do et al. (2017) reported a significant genetic effect of *MAN1C1*, *MAP3K5*, *HCN1*, *TSPAN9*, *MRPS30*, *TEX14*, and *CCL28* genes on lactation persistency. Deng et al. (2019) reported that the *MAP3K5* gene regulates p38 MAPKs and Jun N-terminal kinases (JNKs) pathways involved in mammary gland development. In the present study, lactation persistency was calculated using four different methods, and it was observed that the top four SNPs identified by GWAS for persistency as estimated by the Johansson and Hansson (1940) method were in common with the GWAS results obtained for persistency as per the Ludwick and Petersen (1943) method. The genes present near the top SNPs were *DENND3*, *PSTPIP1*, *ADAMTS20*, *PRODH2*, *NPHS1*, *RREL2*, *FAM19A1*, *CLIC5*, *GPR12*, *DEFB125*, *STPG2*, *CFAP206*, *RINT1*, *EFCAB10*, *TCERG1*, and *C18H16orf78*. Upon network enrichment, it was observed that *TCERG1*, *RINT1*, *CLIC5*, and *PSTPIP1* are co-expressed with several other genes in the breast mammary tissue of human, indicating their possible role in persistency of lactation in dairy animals.

GWAS for fertility traits were performed on breeding efficiency, age at sexual maturity, and postpartum breeding interval. Genes present near the top SNPs of breeding efficiency were *APC*, *LOC112448352*, *NDFIP2*, *LOC107131404*, and *PDP1*. Mohamed et al. (2019) reported that *APC* gene in the canonical WNT signaling pathway plays a critical role in the regulation of ovarian development. Mis-regulation of this key pathway in the adult ovary is associated with subfertility in mice. The *APC* gene has also been mapped to the QTL region for conception rate in Holstein cattle (Kolbehdari et al., 2009). Genes identified near the top SNPs for age at sexual maturity are *PDE11A*, *SLAMF6*, *LOC104974658*, *GRID2*, *LOC104970778*, *EIF5A2*, and *RPL22L1*. Coyral-Castel et al. (2018) mapped the *SLAMF6* gene to the QTL region affecting female fertility located on the bovine chromosome three (QTL-F-Fert-BTA3). The *GRID2* gene is however reported to be associated with growth traits in Simmental cattle and may have role in regulating age at sexual maturity (Braz et al., 2020). Christensen et al. (2005) have reported *EIF5A2* as a candidate gene for infertility in human. Genes identified for postpartum breeding interval were *ERICH2*, *MALSU1*, *IGF2BP3*, *TCF24*, *PPP1R42*, *CCDC93*, *GRID2*, *EDIL3*, *LOC112449367*, *LOC101904981*, and *LOC112447353*. However, none of these genes was identified to be involved significantly in any pathway through the gProfiler.

Although we have already emphasized that this a preliminary GWAS on buffaloes covering a large set of economic traits (lactation, lactation persistency, and fertility), the results obtained in the present study are bias free, as indicated by the FDR. The findings of the present study in the Murrah population will

encourage researchers to come forward for GWAS in buffalo in a holistic manner.

## CONCLUSION

Optimum production and reproduction in buffaloes are long-standing questions in the Indian subcontinent. We present the analysis and identification of genomic regions that play role in shaping the selection and breeding decisions of Murrah buffalo for persistency of production and fertility. The dataset information presented in the paper is also submitted so that it can be used to compare and evaluate other breeds of buffalo based on the genomic information generated. The putative identified regions have a potential to improve the existing breeding decisions in this important dairy germplasm.

## DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the European Variation Archive repository, accession number PRJEB47270 (<https://www.ebi.ac.uk/eva/?eva-study=PRJEB47270>).

## ETHICS STATEMENT

The animal study was reviewed and approved by the ICAR-National Dairy Research Institute IAEC.

## AUTHOR CONTRIBUTIONS

VV conceptualized the work and interpreted the data. SC, GG, RA, and AM performed the collection of materials, and performed the data analysis. AV and SD provided the data and resources. VV and SC did writing of the manuscript and its critical evaluation. All authors contributed to the article and approved the submitted version.

## FUNDING

The authors thank the Director of ICAR-NDRI, Karnal for providing funds and support to carry out the study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.696109/full#supplementary-material>

**Supplementary Figure 1** | MDS plot based on the first and second MDS components showing population stratification.

**Supplementary Figure 2** | Sub-network of enriched pathways of genes identified responsible for milk yield and its composition.

## REFERENCES

- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed August 21, 2020).
- Barbosa, A. M. (2020). FuzzySim: applying fuzzy logic to binary similarity indices in ecology. *Methods Ecol. Evol.* 6, 853–858. doi: 10.1111/2041-210X.12372
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bignardi, A. B., El Faro, L., Rosa, G. J. M., Cardoso, V. L., Machado, P. F., and Albuquerque, L. G. D. (2012). Principal components and factor analytic models for test-day milk yield in Brazilian Holstein cattle. *J. Dairy Sci.* 95, 2157–2164. doi: 10.3168/jds.2011-4494
- Braz, C. U., Rowan, T. N., Schnabel, R. D., and Decker, J. E. (2020). Extensive genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. *Biorxiv*. [Preprint]. doi: 10.1101/2020.01.09.900902
- Brian, B. (2014). “BBMap: a fast, accurate, splice-aware aligner,” in *Proceedings of the 9th Annual Genomics of Energy & Environment Meeting*, (Berkeley, CA: Lawrence Berkeley National Lab).
- Bush, W. S., and Moore, J. H. (2012). Genome-wide association studies. *PLoS Comput. Biol.* 8:e1002822. doi: 10.1371/journal.pcbi.1002822
- Cai, Z., Dusza, M., Guldbandsen, B., Lund, M. S., and Sahana, G. (2020). Distinguishing pleiotropy from linked QTL between milk production traits and mastitis resistance in Nordic Holstein cattle. *Genet. Sel. Evol.* 52:19. doi: 10.1186/s12711-020-00538-6
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8
- Christensen, G. L., Ivanov, I. P., Atkins, J. F., Mielnik, A., Schlegel, P. N., and Carrell, D. T. (2005). Screening the SPO11 and EIF5A2 genes in a population of infertile men. *Fertil. Steril.* 84, 758–760. doi: 10.1016/j.fertnstert.2005.03.053
- Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Crooker, B. A., et al. (2011). Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics* 12:408. doi: 10.1186/1471-2164-12-408
- Coyral-Castel, S., Ramé, C., Cogne, J., Lecardonnell, J., Marthey, S., Esquerré, D., et al. (2018). KIRREL is differentially expressed in adipose tissue from ‘fertil+’ and ‘fertil-’ cows: in vitro role in ovary? *Reproduction* 155, 181–196. doi: 10.1530/REP-17-0649
- da Costa Barros, C., de Abreu Santos, D. J., Aspilcueta-Borquis, R. R., De Camargo, G. M. F., de Araújo Neto, F. R., and Tonhati, H. (2018). Use of single-step genome-wide association studies for prospecting genomic regions related to milk production and milk quality of buffalo. *J. Dairy Res.* 85, 402–406. doi: 10.1017/S0022029918000766
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012
- de Camargo, G. M. F., Aspilcueta-Borquis, R. R., Fortes, M. R. S., Porto-Neto, R., Cardoso, D. F., Santos, D. J. A., et al. (2015). Prospecting major genes in dairy buffaloes. *BMC Genomics* 16:872. doi: 10.1186/s12864-015-1986-2
- Deng, T., Liang, A., Liang, S., Ma, X., Lu, X., Duan, A., et al. (2019). Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. *Front. Genet.* 10:36. doi: 10.3389/fgene.2019.00036
- Do, D. N., Bissonnette, N., Lacasse, P., Miglior, F., Sargolzaei, M., Zhao, X., et al. (2017). Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. *J. Dairy Sci.* 100, 1955–1970. doi: 10.3168/jds.2016-11910
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354
- FAO (2014). *Food and Agriculture Organization the United Nations Statistics Division*. Available online at: <http://faostat3.fao.org/compare/E> (accessed November 12, 2020).
- Ganguli, N. C. (1981). Buffalo as a candidate for milk production. *Int. Dairy Fed. Bull.* 137, 48–56.
- García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., et al. (2012). Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28, 2678–2679. doi: 10.1093/bioinformatics/bts503
- Glickman, M. E., Rao, S. R., and Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J. Clin. Epidemiol.* 67, 850–857. doi: 10.1016/j.jclinepi.2014.03.012
- Gordon, I. (1996). *Controlled Reproduction in Cattle and Buffaloes*, Vol. 1. Wallingford: Centre for Agriculture and Bioscience International.
- Griffiths, M. W. (2010). *Improving the Safety and Quality of Milk: Improving Buffalo Milk*. Amsterdam: Elsevier. doi: 10.1533/9781845699437
- Iamartino, D., Nicolazzi, E. L., Van Tassell, C. P., Reecy, J. M., Fritz-Waters, E. R., Koltes, J. E., et al. (2017). Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS One* 12:e0185220. doi: 10.1371/journal.pone.0185220
- Johansson, I., and Hansson, A. (1940). Causes of variation in milk and butterfat yield of dairy cows. *Kungliga Lantbruksakademiens Handlingar* 79:127.
- Khedkar, C., Kalyankar, S., and Deosarkar, S. (2016). “Buffalo milk,” in *Encyclopedia of Food and Health*, eds B. Caballero, M. Finglas, and F. Toldrá (New York, NY: Academic Press), 522–528. doi: 10.1016/B978-0-12-384947-2.0.0093-3
- Kolbeldari, D., Wang, Z., Grant, J. R., Murdoch, B., Prasad, A., Xiu, Z., et al. (2009). A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *J. Animal Breed. Genet.* 126, 216–227. doi: 10.1111/j.1439-0388.2008.00793.x
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Liu, J. J., Liang, A. X., Campanile, G., Plastow, G., Zhang, C., Wang, Z., et al. (2018). Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. *J. Dairy Sci.* 101, 433–444. doi: 10.3168/jds.2017-13246
- Ludwick, T. M., and Petersen, W. E. (1943). A measure of persistency of lactation in dairy cattle. *J. Dairy Sci.* 26, 439–445. doi: 10.3168/jds.S0022-0302(43)92739-0
- Macciotta, N. P. P., Vicario, D., and Cappio-Borlino, A. (2005). Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. *J. Dairy Sci.* 88, 1178–1191. doi: 10.3168/jds.S0022-0302(05)72784-3
- Mahadevan, P. (1951). The effect of environment and heredity on lactation. II. persistency of lactation. *J. Agric. Sci.* 41, 89–93. doi: 10.1017/S0021859600058573
- Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., et al. (2018). A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27:e1608. doi: 10.1002/mpr.1608
- Mohamed, N. E., Hay, T., Reed, K. R., Smalley, M. J., and Clarke, A. R. (2019). APC2 is critical for ovarian WNT signalling control, fertility and tumour suppression. *BMC Cancer* 19:677. doi: 10.1186/s12885-019-5867-y
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7:e37135. doi: 10.1371/journal.pone.0037135
- Rahmatalla, S. A., Arends, D., Reissmann, M., Wimmers, K., Reyer, H., and Brockmann, G. A. (2018). Genome-wide association study of body morphological traits in Sudanese goats. *Anim. Genet.* 49, 478–482. doi: 10.1111/age.12686
- Raina, V., Narang, R., Malhotra, P., Kaur, S., Dubey, P. P., Tekam, S., et al. (2016). Breeding efficiency of crossbred cattle and Murrah buffaloes at organized dairy farm. *Indian J. Anim. Res.* 50, 867–871. doi: 10.18805/ijar.v0iOF.6663
- Ruffalo, M., Koyutürk, M., Ray, S., and LaFramboise, T. (2012). Accurate estimation of short read mapping quality for next-generation genome sequencing. *Bioinformatics* 28, i349–i355. doi: 10.1093/bioinformatics/bts408

- Sambrook, J., and Russell, D. W. (2006). Purification of nucleic acids by extraction with phenol: chloroform. *Cold Spring Harb. Protoc.* 2006:4455. doi: 10.1101/pdb.prot4455
- Schaeffer, L. R., Jamrozik, J., Van Dorp, R., Kelton, D. F., and Lazenby, D. W. (2000). Estimating daily yields of cows from different milking schemes. *Livest. Prod. Sci.* 65, 219–227. doi: 10.1016/S0301-6226(00)00153-6
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Taggar, R. K., Das, A. K., Kumar, D., and Mahajan, V. (2012). Prediction of milk yield in Jersey cows using principal component analysis. *Progr. Res.* 7, 272–274.
- Tomar, N. S. (1965). A note on the method of working out breeding efficiency in Zebu cows and buffaloes. *Indian Dairymen* 17, 389–390.
- Utsunomiya, Y. T., O'Brien, A. M. P., Sonstegard, T. S., Van Tassell, C. P., do Carmo, A. S., Mészáros, G., et al. (2013). Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS One* 8:e64280. doi: 10.1371/journal.pone.0064280
- Van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., et al. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* 5, 247–252. doi: 10.1038/nmeth.1185
- Venturini, G. C., Cardoso, D. F., Baldi, F., Freitas, A. C., Aspilcueta-Borquis, R. R., Santos, D. J. A., et al. (2014). Association between single-nucleotide polymorphisms and milk production traits in buffalo. *Genet. Mol. Res.* 13, 10256–10268. doi: 10.4238/2014.December.4.20
- Wara, A. B., Kumar, A., Singh, A., Arthikeyan, A. K., Dutt, T., and Mishra, B. P. (2019). Genome wide association study of test day's and 305 days milk yield in crossbred cattle. *Indian J. Anim. Sci.* 89, 861–865.
- Warriach, H. M., McGill, D. M., Bush, R. D., Wynn, P. C., and Chohan, K. R. (2015). A review of recent developments in buffalo reproduction—a review. *Asian-Australas. J. Anim. Sci.* 28:451. doi: 10.5713/ajas.14.0259
- Wood, P. D. P. (1967). Algebraic model of the lactation curve in cattle. *Nature* 216, 164–165. doi: 10.1038/216164a0
- Wu, J. J., Song, L. J., Wu, F. J., Liang, X. W., Yang, B. Z., Wathes, D. C., et al. (2013). Investigation of transferability of BovineSNP50 BeadChip from cattle to water buffalo for genome wide association study. *Mol. Biol. Rep.* 40, 743–750. doi: 10.1007/s11033-012-1932-1
- Zhou, C., Shen, D., Li, C., Cai, W., Liu, S., Yin, H., et al. (2019). Comparative transcriptomic and proteomic analyses identify key genes associated with milk fat traits in Chinese Holstein cows. *Front. Genet.* 10:672. doi: 10.3389/fgene.2019.00672
- Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., et al. (2009). A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42. doi: 10.1186/gb-2009-10-4-r42

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Vohra, Chhotaray, Gowane, Alex, Mukherjee, Verma and Deb. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.