Supplementary Figures Available Online

Indian Journal of Animal Sciences **89**(5): 528–538, May 2019/Article
https://doi.org/10.56093/ijans.v89i5.90021

Identification of quantitative trait loci for milk protein percentage in Murrah buffaloes

UPASNA SHARMA[1], PRIYANKA BANERJEE[2], JYOTI JOSHI[3], PRERNA KAPOOR[4] and RAMESH KUMAR VIJH[5]

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana 132 001 India

Received: 26 June 2018 ; Accepted : 31 October 2018

ABSTRACT

Milk protein is an important constituent of milk in buffaloes and is moderately heritable. The milk protein percentage varies significantly between breeds/herds/species. Buffaloes can be selected for higher milk protein percentage and this paper provides QTLs for marker assisted selection in buffaloes. The milk protein percentage records on 2,028 daughters belonging to 12 half sib families were analyzed for the identification of QTLs on 8 chromosomes in buffaloes using chromosome scans. The single marker analysis revealed 74 markers to be associated with milk protein percentage in 12 sire families. When common markers were removed from the analysis, 51 markers remained. The Interval mapping using R/qtl identified 69 QTLs in 12 half sib families on 8 chromosomes of buffalo. The meta QTL analysis defined 25 consensus QTL regions in buffaloes for milk protein percentage. Most of the QTLs identified have been reported for cattle however few new chromosomal locations were also identified to be associated with milk protein percentage in buffaloes. Comparative genomics revealed 1117 genes underlying the QTL regions associated with milk protein percentage. Among these, 109 genes were directly associated with protein metabolism. The protein-protein interaction among the genes and gene ontology analysis and pathways have been identified. These 109 genes have potential to be candidate genes for milk protein percentage in buffaloes.

Key words: Buffaloes, Candidate genes, Milk protein percentage, QTLs

Most of the economic traits in buffaloes are quantitative in nature which means that they are affected not only by the environmental factors but also by a large number of genes. These large number of genes with small effects cumulatively and in a coordinated manner affect the same phenotype trait which exhibits a continuous distribution of phenotypic expression. The recent developments in molecular biology and analytical procedures have made it possible to decipher the genetic architecture of these quantitative traits by identification of chromosomal loci affecting these traits. Milk protein is an important constituent of milk. Milk protein percentage is a quantitative trait and has a positive correlation with milk fat percentage. Milk protein percentage is high in buffaloes compared to cattle. Large variation of milk protein percentage has been reported in cattle and similar situation might exist in buffaloes. The stage of lactation and seasonal variations along with the total milk yield of the animals/breeds/herd shall also impact the protein percentage in buffaloes. The heritable nature of the milk protein percentage (moderate heritability of h^2=0.5) in buffaloes makes it possible to increase it through selection of buffaloes using QTL markers. Till date there are no reports of markers linked to protein percentage in buffaloes nor any of the QTL markers have been identified. The present study was undertaken for the identification of quantitative trait loci or the chromosomal regions associated with protein percentage in buffaloes and the identification of candidate genes associated with this trait.

Present address: [1]Research Associate (upasna30@gmail.com), [4]Senior Research Fellow (prernakapoor31@gmail.com), [5]Principal Scientist (rameshkvijh@gmail.com). [2]Post Doctoral Fellow (priyankabnrj@gmail.com), Technical University of Denmark, Lyngby, Kobenhavn, Denmark. [3]Post Doctoral Fellow (jyotijoshi111@gmail.com), Dalhousie University, Nova Scotia, Canada.

MATERIALS AND METHODS

The reference family of buffaloes (animals of known pedigree and having both genotype and phenotype information) was created for the work (Vijh *et al.* 2013, 2018). The accuracy of the paternity records was authenticated using a set of DNA markers and only the daughters with confirmed paternity were recorded for phenotypes (Vijh *et al.* 2014). The daughters belonging to 12 half sib families were recorded for milk protein percentage during three stages of lactation. The test day protein record was generated by mixing the milk of morning

and evening and then estimating the protein percentage of milk through an automated milk analysis system. The mean of milk protein percentage of the three records was utilised for the analysis. The genotype data on 8 chromosomes of these animals was generated and has been reported (Vijh *et al.* 2013).

The single marker analysis of data was carried out using QTL cartographer software. For the identification of QTLs, we utilised the software R/qtl. The data generated fits into half sib design and the method used for the analysis of data was the most simple type, i.e. backcross. A total of 12 data set were created, one for each sire family. For a single QTL model, we utilised the standard interval mapping and Haley-Knott regression algorithm. The values of LOD score were utilised as an evidence of the existence of QTL. The statistical significance of LOD was tested using permutation test (Churchill and Doerge 1994) with 1000 replicates. The multiple QTL models as implemented in R/qtl were used as they have increased power to detect QTL, better separation of linked QTLs and defining epistatic interactions. We fitted the Full QTL model using the *fitqtl* function of R/qtl package and this also included the interaction among the QTLs. The ANOVA table indicates the overall fit of the model; the LOD score obtained is relative to the null model (with no QTL). The drop one QTL model was utilised to see the effect of each identified QTL and its related interactions.

QTL meta-analysis was carried out to synthesize QTL information from 12 independent half sib family analysis results and also to refine the chromosomal region involved using Biomercator software v4.2. The QTL meta-analysis algorithm developed by Goffinet and Gerber (2000) was used. We fitted five models using Gaussian distribution and the best fit was determined by means of the maximum likelihood method and Akaike information content. Using the select model, consensus QTL positions were determined as the mean of QTL distribution maximizing the likelihood and confidence interval.

Once the metaQTL regions with their confidence interval were known we utilised webserver AnnotQTL (http://annotqtl.genouest.org) for the identification of genes underlying the QTL region. Each metaQTL region was taken as an input. We utilised the Buffalo-Cattle synteny reported on the basis of radiation hybrid panel (Amaral *et al.* 2008). This provided a list of genes underlying the identified QTLs in buffaloes as Ensembl IDs and also provided a list of human genes assuming the synteny between cattle and human. The genes were then mapped using Reactome database (available at https://reactome.org) (Croft *et al.* 2014, Fabregat *et al.* 2016) to identify the genes related to protein metabolism, amino acids and their derivatives in cattle. We downloaded the Uniprot IDs associated with the identified genes. The Uniprot ID of these genes were used as input for the webserver STRING (https://string-db.org) (Szklarczyk *et al.* 2015) to obtain network of predicted association for a particular group of proteins/genes.

RESULTS AND DISCUSSION

The pricing of the buffalo milk as of today is based on the quality of milk and its fat percentage, however other milk components like milk protein percentage will also soon become important owing to processing and quality of the products obtained from them. Buffalo is known to have higher protein and fat percentage in its milk and these traits are positively correlated and have negative correlation with milk yield. These two important components can vary tremendously from one breed/ herd to another and have got moderately high heritability values (much higher than milk yield) making them suitable candidates for selection in buffaloes. To detect the association between microsatellites and protein percentage, we utilised single marker analysis as well as Interval mapping. As the protein percentage was recorded on daughters of 12 sires, the analysis was also carried out taking sires' family as a unit. There were 2028 protein percentage records with a mean of approximately 169 records per sire family. The number of daughters recorded for each sire, the mean, minimum and maximum milk protein percentage during the first lactation has been given in Table 1.

Table 1. Sire wise distribution of daughters with milk protein percentage records, mean, minimum and maximum protein percentage in each sire family

Sire	No. of Daughters	Mean±SE	Min. protein %	Max. protein %
Sire1	156	3.43±0.04	1.10	5.86
Sire2	162	3.64±0.06	1.87	5.86
Sire3	87	3.47±0.07	1.60	5.50
Sire4	205	3.45±0.04	1.91	5.27
Sire5	263	3.51±0.04	0.64	6.50
Sire6	69	3.59±0.10	2.27	8.94
Sire7	325	3.63±0.04	1.78	7.50
Sire8	264	3.49±0.03	2.07	6.07
Sire9	215	3.58±0.04	2.01	6.00
Sire10	158	3.53±0.04	1.34	6.09
Sire11	77	3.54±0.07	1.60	5.56
Sire12	47	3.38±0.09	1.59	4.69

To detect associations between molecular markers and milk protein percentage, we initially utilised the single marker analysis. This analysis is based on the fact that molecular marker genotypes can be classified into groups means that marker genotypes can be used as classifying variables for ANOVA, or as variables for regression analysis. The null hypothesis is genotypic classes do not differ in phenotype for a given molecular marker. Single marker analysis calculates whether phenotype values differ among genotypes for a given molecular marker. It does not require the linkage map to be available for the species (Collard *et al.* 2005). In the present half sib family analysis, we tested for linkage of a marker to a QTL and the evidence of linkage was measured by a LOD score representing likelihood ratio. The analysis of 12 half sib families revealed markers which were significantly associated with the milk

protein percentage. The LOD score and R^2 values are given in Table 2.

Table 2. Significant QTL of protein percentage for 12 half sib families using single marker analysis

Sire	Significant			LOD	R^2	P (F value)
	Chromosome	Marker	Marker name			
Sire1	1	4	BMS4037	6.628	0.018	0.010*
	4	2	BMS1788	4.876	0.011	0.028*
	4	7	BM1260	10.609	0.052	0.001**
	4	9	BMS2571	4.613	0.017	0.032*
	6	4	MNB-66	5.104	0.011	0.024*
	7	5	DIK2819	6.773	0.018	0.010**
	7	6	BMS6117	8.990	0.041	0.003**
	14	2	DIK4015	6.413	0.020	0.012*
	14	7	BL1029	3.991	0.018	0.047*
Sire2	2	1	ILSTS026	5.608	0.028	0.019*
	3	2	DIK1057	5.969	0.041	0.015*
	3	8	DIK2511	5.523	0.020	0.020*
	4	9	BMS2571	6.102	0.028	0.014*
	4	10	BMS1074	6.148	0.038	0.014*
	6	2	DIK5153	4.758	0.041	0.030*
	6	3	DIK5285	7.455	0.052	0.007**
	6	4	MNB-66	4.198	0.011	0.042*
	7	8	BB719	6.945	0.039	0.009**
Sire3	4	6	DIK4373	4.693	0.019	0.033*
	6	2	DIK5153	4.692	0.028	0.033*
Sire4	2	5	BM1223	5.593	0.038	0.019*
	4	6	DIK4373	3.952	0.000	0.048*
	4	7	BM1260	5.203	0.022	0.023*
	7	2	MB053	4.709	0.023	0.031*
	7	10	BMS1979	4.876	0.010	0.028*
Sire5	1	8	BMS4011	5.494	0.023	0.020*
	2	3	BMS803	6.908	0.019	0.009**
	2	6	BMS1987	4.287	0.019	0.039*
	2	8	BMS2519	5.457	0.025	0.020*
	6	8	DIK3024	9.605	0.011	0.002**
	6	10	BM4311	7.068	0.028	0.008**
	9	5	TGLA261	7.430	0.051	0.007**
	14	9	BMS2055	5.656	0.023	0.018*
Sire6	1	8	BMS4011	4.043	0.066	0.048*
	3	5	ILSTS029	4.243	0.065	0.043*
	3	7	BM3020	5.023	0.068	0.028*
	6	4	MNB-66	4.397	0.054	0.040*
	7	9	BMS522	7.361	0.020	0.008**
	9	6	BMS1909	4.171	0.061	0.045*
	9	9	BMS2063	4.596	0.032	0.035*
	14	6	BM8215	5.595	0.075	0.020*
Sire7	1	5	BMS4048	3.909	0.012	0.049*
	3	3	CSSM054	4.279	0.011	0.039*
	3	7	BM3020	5.569	0.025	0.019*
	3	8	DIK2511	5.112	0.009	0.024*
	4	3	BMS827	4.181	0.011	0.042*
	6	2	DIK5153	6.101	0.015	0.014*
	6	3	DIK5285	4.506	0.011	0.034*
	6	5	BM1329	6.609	0.018	0.010*
	6	8	DIK3024	4.526	0.007	0.034*
	7	3	DIK4204	6.087	0.022	0.014*
	9	6	BMS1909	6.461	0.021	0.011*
	9	7	DIK2816	3.879	0.007	0.050*
	14	6	BM8215	5.787	0.027	0.017*
Sire8	3	5	ILSTS029	6.301	0.024	0.012*
	4	4	BMS1237	6.316	0.017	0.012*
	4	5	DIK4816	5.718	0.004	0.017*
Sire9	1	7	DIK5127	4.652	0.022	0.032*
	4	11	BL1121	6.127	0.021	0.014*
	6	4	MNB-66	5.967	0.029	0.015*
	7	9	BMS522	12.011	0.032	0.001***
Sire10	3	1	URB006	4.648	0.033	0.032*
	4	1	DIK4364	3.994	0.024	0.047*
	4	9	BMS2571	4.246	0.031	0.041*
	6	7	BMS690	4.465	0.020	0.036*
	7	5	DIK2819	6.806	0.045	0.010**
	14	8	BM1577	4.122	0.014	0.044*
	14	9	BMS2055	6.884	0.036	0.009**
Sire11	2	8	BMS2519	5.609	0.060	0.020*
	4	3	BMS827	7.205	0.103	0.008**
	4	6	DIK4373	5.945	0.045	0.016*
	14	4	ILSTS011	6.115	0.070	0.015*
Sire12	1	6	BMS119	5.759	0.112	0.020*
	1	8	BMS4011	5.927	0.128	0.018*

***P<0.001; **P<0.01; *P<0.05.

All the sire families exhibited markers having association with the protein percentage in milk. However different sire families did not reveal associated markers on all the chromosomes. A total of 74 associated markers were observed in the 12 sire families. Common markers if deleted from the count revealed only 51 microsatellite markers to be associated with the protein percentage milk. Markers (23) were common among the families for the same chromosome analyzed. Single marker analysis is considered inferior to two marker analysis by several research workers (Haley and Knott 1992, Knott and Haley 1992, Kao 2000, Rebai *et al.* 1995). However, Coffman *et al.* (2003) have reported intersection tests for single marker QTL analysis can be more powerful than two marker QTL analysis (Doerge *et al.* 1997).

In the single marker analysis, different markers have been identified in different sire families and this can be attributed to heterozygosity of the sire at a particular marker loci. All the sires may not be heterozygous for all the loci and QTLs can only be identified in sires in which the markers are segregating. However, the primary disadvantage of the single marker analysis is need to omit the individuals with missing marker information. Such a situation was however not in present analysis since all the markers in all the individuals of the 12 families had marker genotypes. The second notable shortcoming is that one cannot inspect the position between the markers and one receives only a poor information about the position of QTLs. We thus also applied interval mapping and Multiple QTL models for analysis.

The interval mapping as implemented in R/qtl was carried out using the Haley-Knott regression and extended Haley-Knott regression method to fit the multiple QTL model. The chromosome scans of all the 8 chromosomes studied in the experiments revealed large number of QTLs for milk protein percentage in buffaloes. The details of the sire families, detected QTLs, their LOD score, percentage of the phenotypic variance explained and level of significance are depicted in the Table 3. The interaction among the QTLs where ever significant are also mentioned.

Table 3. Significant QTL locations for milk protein percentage for 12 half sib families using Interval Mapping (Haley-Knott regression and extended Haley-Knott regression) as implemented in R/qtl.

Sire No.	Chromosome No.	Position on Chromosome	LOD	% variance	F value	P value (F)
Sire1	3	19.3	1.508	2.104	3.251	0.040*
	3	84.8	1.731	2.436	3.739	0.025*
	4	94	2.031	2.866	4.399	0.013*
	4	104	3.272	4.634	2.386	0.029*
	6	64	4.431	6.336	3.263	0.004**
	6	70	2.556	3.621	5.559	0.004**
	7	60	1.228	1.709	2.641	0.073P
	7	104	2.854	4.054	2.074	0.057P
	9	30.2	1.256	1.761	2.703	0.069P
	4@104.0:6@64.0		2.811	3.966	3.064	0.017*
	7@104.0:14@5.1		1.899	2.676	2.054	0.087P
Sire2	2	94.8	1.449	3.004	2.952	0.056P
	3	19.3	4.2	9.07	2.971	0.009**
	4	74	3.123	6.884	2.175	0.049*
	6	10	2.267	4.933	4.675	0.011*
	6	20	3.716	7.967	2.61	0.020*
	7	94	1.366	2.828	2.779	0.066P
	2@106.8:4@74.0		2.044	4.434	2.101	0.084P
	3@19.3:6@20.0		2.494	5.252	2.581	0.040*
Sire3	1	135.1	3.261	11.982	2.126	0.062P
	2	30.8	3.685	13.495	2.431	0.035*
	3	69.3	3.081	11.265	1.999	0.078P
	6	20	3.309	11.991	2.16	0.058P
	7	70	1.365	4.685	2.532	0.087P
	2@30.8:6@20.0		2.491	8.822	2.384	0.060P
Sire4	1	115.11	2.515	5.026	1.839	0.094P
	2	10.77	2.578	5.155	1.886	0.085P
	7	126	1.968	3.909	4.291	0.015*
Sire5	2	46.8	1.201	1.9	2.611	0.076P
	3	79.3	1.1	1.738	2.388	0.094P
	4	94	1.227	1.941	2.668	0.072P
	4	94	1.833	2.947	3.972	0.020*
	7	14	1.57	2.519	3.395	0.035*
	9	74.2	2.891	4.642	2.126	0.051P
Sire6	1	110.5	43.562	73.719	158.077	0.0000***
	1	110.5	5.274	4.583	10.05	0.000***
	2	16.8	2.031	0.582	3.746	0.031*
	3	29.3	2.217	0.64	4.115	0.022*
	3	29.3	2.144	1.666	3.652	0.034*
	4	20	2.232	1.739	3.814	0.029*
	6	10	22.261	37.958	27.743	0.000***
	7	80	43.098	71.28	152.849	0.0000***
	7	104	26.474	54.231	39.636	0.000***
	9	90.2	1.714	0.486	3.125	0.053P
	14	19.1	5.143	4.448	9.752	0.000***
	1@110.5:7@80.0		41.847	65.087	209.353	0.0000***
	6@10.0:7@104.0		20.869	33.538	36.768	0.000***
Sire7	1	65.1	1.303	1.356	2.732	0.067P
	1	69.1	1.439	1.798	3.147	0.044*
	2	96.8	4.753	6.085	3.55	0.002**
	3	74.9	1.147	1.192	2.403	0.092P
	4	0	11.182	12.515	8.407	0.000***
	4	60	1.702	2.131	3.73	0.025*
	6	10	4.666	5.97	3.483	0.002**
	7	24.4	1.194	1.241	2.5	0.084P
	14	25.1	10.654	11.878	7.978	0.000***
	2@96.8:6@10.0		3.521	4.466	3.909	0.004**
	4@0.0:14@25.1		10.491	11.682	11.77	0.000***
Sire8	1	75.1	2.476	4.198	1.86	0.088P
	7	94	2.439	4.08	1.803	0.099P
Sire9	1	65.1	3.645	6.45	2.707	0.015*
	2	76.8	1.348	2.326	2.929	0.056P
	4	104	3.508	6.255	2.523	0.023*
	7	120	3.708	6.627	2.673	0.016*
	7	120	6.433	11.733	4.924	0.000***
	14	65.1	1.31	2.281	2.76	0.066P
	1@65.1:7@120.0		2.267	3.951	2.487	0.045*
Sire10	1	119.1	2.619	6.763	1.89	0.086P
	7	44	2.556	6.596	1.844	0.095P
Sire11	1	95.1	1.843	7.896	3.846	0.026*
	4	30	1.497	5.917	2.904	0.062P
	6	0	1.317	5.551	2.704	0.074P
	6	4	1.705	6.78	3.328	0.042*
	7	40	3.733	15.798	2.585	0.027*
	7	40	2.572	11.265	5.487	0.006**
Sire12	2	76.8	1.487	6.775	2.51	0.097P
	3	103.3	3.551	17.97	2.219	0.067P
	3	119.3	2.115	9.942	3.683	0.036*
	7	44	4.357	22.996	2.84	0.025*
	3@103.3:7@44.0		2.601	12.537	2.322	0.078P

***P<0.001; **P<0.01; *P<0.05.

Table 4. Meta QTL analysis of the QTL positions obtained from 12 half sib sire families

BTA Chromosome	BBU Chromosome	AIC value	Mean position (CI)	Mean position (CI)	Mean position (CI)	Mean position (CI)
1	BBU1q	88.03	65.97 (62.93–69.0)	96.11 (91.11–100.11)	112.3 (110.69–113.92)	135.11 (131.11–141.11)
2	BBU2q	61.41	10.77 (8.49–13.04)	39.68 (35.74–43.62)	85.84 (81.29–90.4)	96.24 (93.67–98.81)
3	BBU6	98.89	24.26 (21.81–26.71)	70.4 (68.48–72.32)	86.54 (83.06–90.01)	123.36 (118.34–125.0)
4	BBU8	101.87	9.83 (6.23–13.43)	28.0 (23.0–31.0)	69.59 (65.52–73.66)	100.88 (98.7–103.07)
6	BBU7	57.85	5.61 (3.65–7.58)	9.99 (7.36–12.63)	20.77 (17.24–24.3)	68.9 (65.2–72.6)
7	BBU9	203.61	23.71 (21.12–26.3)	46.36 (44.05–48.66)	74.8 (72.12–77.48)	121.63 (119.83–123.43)
9	BBU10	17.6	25.2 (20.2–32.2)	71.2 (66.2–75.2)	97.2 (90.2–98.64)	–
14	BBU15	14.22	14.13 (11.29–16.98)	63.12 (61.12–68.12)	–	–

Values in paranthesis represent the confidence interval of the meta QTL. BTA, *Bos taurus*; BBU, *Bubalus bubalis*.

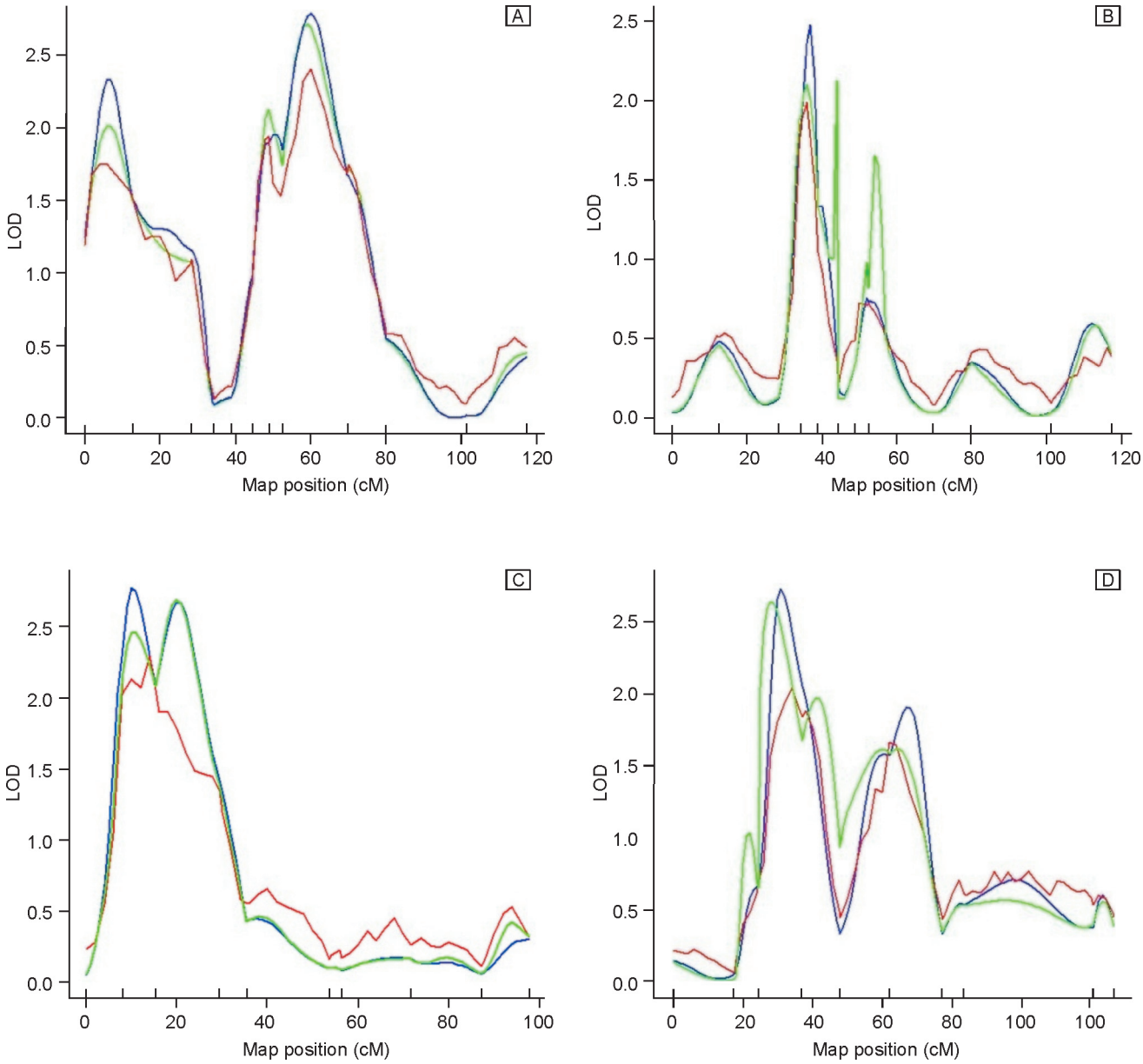

Fig. 1. Chromosome scans of BBU8 (A, B), BBU7 (C) and BBU9 (D) belonging to different sire families depicting significant QTL locations for milk protein percentage in buffaloes.

The representative chromosomal scans have been depicted as Fig. 1. Eleven interactions amongst the QTLs were found to be statistically significant. The most common chromosomes involved in the interactions (5 significant interactions) were chromosome 6 and chromosome 7. Both the chromosomes are known to harbour important genes related to the protein expression. Infact chromosome 6 harbours genes for not only proteins and fats but also for milk yield and somatic cell score in buffaloes. Total 69 QTLs were found to be associated with milk protein percentage in the 12 half sib families analysed. The QTLs identified in each of the half sibs families are independent of one another. QTLs' identification is dependent on the heterozygosity of the sire on particular marker loci and thus the different QTLs may be identified.

The genome scan carried out for detection of the QTLs indicate the chromosomal regions harbouring genes associated with the trait of interest. This however have a confidence interval of about 20 cM or more which is not sufficient for marker assisted selection. However, the QTLs can be localised more precisely using linkage disequilibrium analysis. However, the meta-analysis of the QTLs obtained in the different families can provide a means for improvement in the QTL localisations. This is however achieved at no additional cost and requires only the analysis of the already obtained data. Thus the QTLs identified using similar markers may yield different QTLs and also similar QTLs and this depends upon the heterozygosity of markers for a sire on different chromosomes. The positions of the QTL regions may differ from family to family and need to

be further analysed to find consensus regions. We detected significant differences in different families and this may be partly due to segregation of allele of the same QTL or due to segregation of various QTLs. The meta analysis of the identified QTL regions for protein percentage in buffaloes was carried out on QTLs on 8 chromosomes of buffaloes. The selection of the model from the 5 models tested for the analysis was based on Akaike Information Content. The meta analysis of the QTLs across 12 families of buffaloes are depicted in Table 4. The table depicts the chromosome number, metaQTL location and their confidence interval.

Most of the QTLs for protein percentage in cattle have been analysed in Hostein cattle and the effects of these QTLs have been found to be moderately leptokurtic by fitting Gamma distributions of QTL effects. The results summarize into the fact that there are few genes of large effect and few genes of small effect (Hayes and Goddard 2001). They also predicted that there may be 50 to 100 genes affecting a quantitative trait in dairy cattle, 17% of the genes could explain about 90% of the genetic variance. The analysis of the QTLs in the laboratory animals (Corva and Medrano 2001, Mackay 2001) have suggested to investigate the interaction among the loci (epistasis) which in the present study have been found to be statistically significant. The genotype environmental interactions, imprinting effects, and linked QTL in dairy cattle have been suggested for fully understanding the genetic architecture of quantitative traits. In the present study of buffaloes, the meta-analysis of the various QTLs as implemented in Biomercator version 4.2 revealed four consensus regions on the chromosome BBU1q. The mean position of the QTL regions were 65.97, 96.11, 112.3 and 135.11. The confidence interval of the meta-QTL regions have been given in the brackets. The BBU1q is equivalent to BTA 1 chromosome as per the radiation hybrid panel synteny (Amaral *et al.* 2008). There are no reports of QTLs in buffaloes and thus the comparison of the buffalo QTLs has been made with respect to the cattle chromosomes on which large literature is available for mostly Holstein Friesian cattle (Khatkar *et al.* 2014). The mean position of the meta QTL and its confidence interval has been reported in cattle by Viale *et al.* (2017) and Cole *et al.* (2016). Similarly for buffalo chromosome BBU2q (BTA 2), four meta-QTLs positions have been identified. The QTLs at these locations along with the confidence interval have been reported for cattle by Jiang *et al*. (2017), Chamberlain *et al.* (2012), Fang *et al.* (2014), Bagnato *et al.* (2008), Pimentel *et al.* (2011) and Heyen *et al*. (1999). For the buffalo chromosome BBU6 (BTA 3), the present analysis revealed four metaQTL regions. Similar locations have been reported by several workers in cattle (Cole *et al.* 2011, Pimentel *et al.* 2011, Alain *et al.* 2009, Mosig *et al.* 2001 and Buitenhuis *et al*. 2016). For the buffalo chromosome BBU7 (BTA 6), we identified 4 meta-QTL regions and similar locations have been reported to be associated with milk protein percentage in cattle by several authors (Cole *et al.* 2011, Pryce *et al.* 2010). The meta-QTL analysis of the results obtained from 12 sire families in buffaloes for the buffalo chromosome BBU9 (BTA 7), four consensus QTL regions for milk protein percentage were obtained and similar chromosomal location have been reported by several authors in cattle (Cole *et al.* 2011, Ron *et al.* 2004). Similarly, for buffalo chromosomes BBU10 (BTA 9) and BBU15 (BTA14), three and two meta QTL region were identified. These regions in cattle have been reported by Plante *et al.* (2001) and Cole *et al.* (2011) for BTA 9 and Russo *et al.* (2012) and Jiang *et al.* (2010) for QTL positions on chromosome BTA14. MetaQTL positions for protein percentage on 8 chromosomes for buffalo are shown in Fig. 2.

There is a great degree of synteny between the cattle and buffalo chromosomes and have been established using the radiation hybrid panels (Amaral *et al.* 2008). Similarly, the synteny between cattle chromosomes and human chromosomes is well established and available in the Ensembl database (https://www.ensembl.org/). Thus using the synteny of the chromosomal regions and using comparative genomics the genes underlying the QTL regions were established.

The results presented above are on the basis of partial genome scans (8 chromosome equivalent of cattle in buffaloes). It can be seen from the comparative analysis that the chromosome scans indicate the chromosome regions that harbour genes relate to protein percentage and are identified within a confidence of 2.5 to 3.0 cM. There are usually significant differences on the magnitude of effects of these QTLs. The differences may be there in the same experiment itself and may exist among the families. The total number of genes underlying the QTL regions of 8 BTA chromosomes were 1117 genes. The chromosome wise genes identified underlying QTL regions and related to protein metabolism are depicted in Fig. 3. These genes were mapped on Reactome database (https://reactome.org/) and 109 genes associated with protein metabolism were identified and shown in Table 5.

The enrichment analysis of the pathways and other functional enrichments revealed a total of 103 nodes and 188 edges in the analysis of genes in STRING webserver (https://string-db.org/). The protein-protein interaction enrichment value was highly significant and means that the proteins had more interactions among themselves than would be expected for a random set of protein of similar size. Such an enrichment points towards that the proteins are at least partially biologically connected as a group. The six KEGG pathways that were found to be highly enriched related to glycine, serine, threonine metabolism, arginine and proline metabolism, ubiquitin mediated proteolysis and GABAergic synapse, ribosomal and metabolic pathways. The functional enrichment of biological processes related to cellular protein metabolic processes, protein metabolic process, cellular protein modification processes, post translational protein modification and macromolecule modification. The molecular function related to catalytic activity and ubiquitin like protein transferase activity was highly enriched. The cellular component GO related to

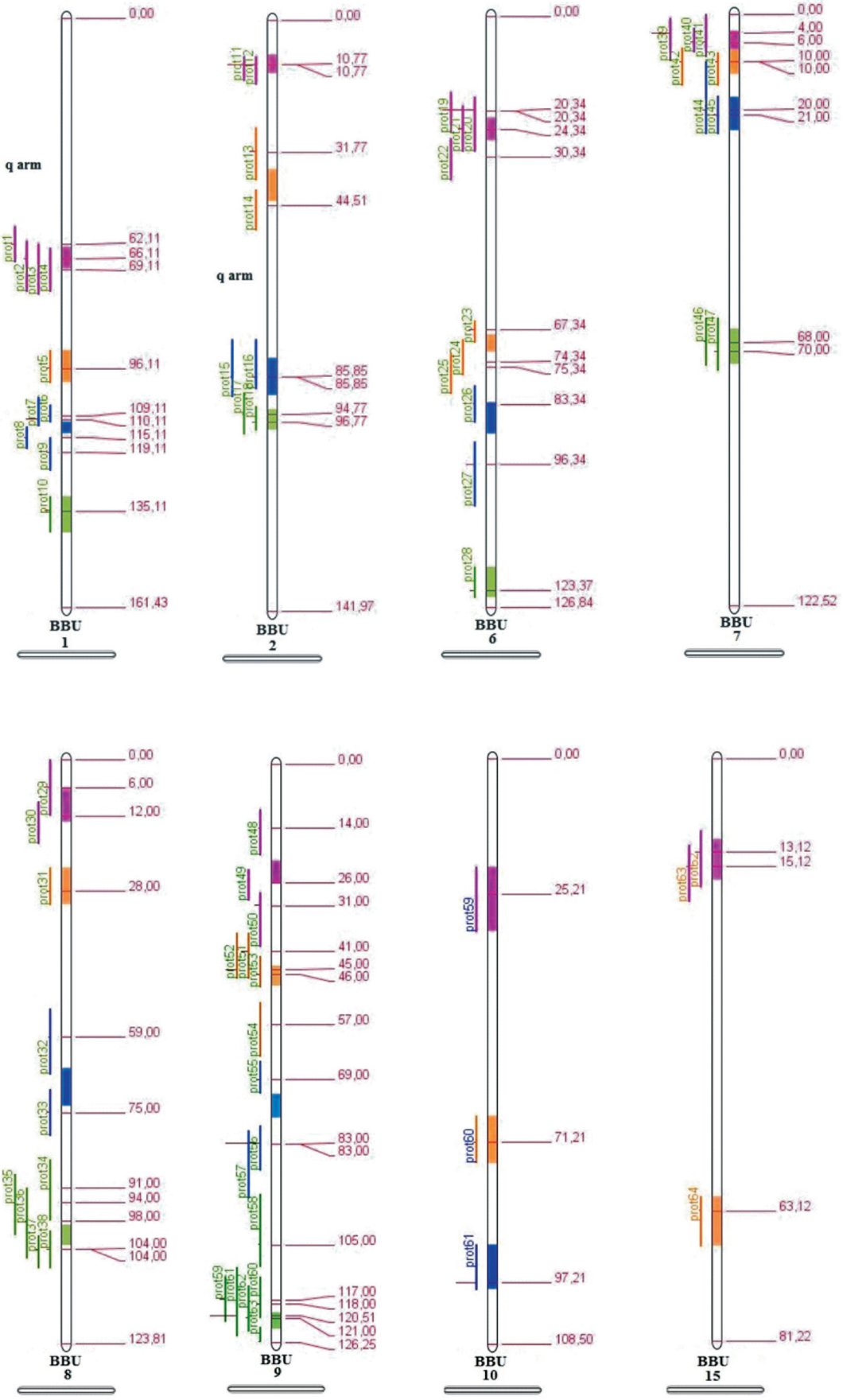

Fig. 2. The metaQTL positions for protein percentage on 8 chromosomes for buffalo.

Table 5. Genes associated with protein metabolism identified using Reactome database.

Chromosome number of cattle (Its equivalent in buffalo)	Cattle genes identified which underline the identified QTL regions in buffaloes	Mapped entities in Reactome database	HGNC names of mapped genes
BTA1 (BBU1q)	ENSBTAG00000014843	P35606	COPB2
	ENSBTAG00000007273	Q12841	FSTL1
	ENSBTAG00000047608	Q9GZV4	EIF5A2
	ENSBTAG00000001390	Q8WVM7	STAG1
	ENSBTAG00000002075	Q14789	GOLGB1
	ENSBTAG00000014508	Q8NDX5	PHC3
	ENSBTAG00000013938	O60513	B4GALT4
	ENSBTAG00000008862	Q9UNA3	A4GNT
	ENSBTAG00000001451	Q9UH90	FBXO40
	ENSBTAG00000022155	P58012	FOXL2
	ENSBTAG00000030942	Q99442	SEC62
	ENSBTAG00000039442	P82650	MRPS22
	ENSBTAG00000035437	P08473	MME
	ENSBTAG00000031277	P09001	MRPL3
	ENSBTAG00000010550	P02787	TF
	ENSBTAG00000015895	Q93099	HGD
BTA2 (BBU2q)	ENSBTAG00000005517	Q8WXB1	METTL21A
	ENSBTAG00000021691	Q9UPW6	SATB2
	ENSBTAG00000014407	Q7Z7M9	GALNT5
	ENSBTAG00000021979	P63165	SUMO1
	ENSBTAG00000009859	P24534	EEF1B2
	ENSBTAG00000020416	Q8IUC8	GALNT13
	ENSBTAG00000005476	Q9Y2X3	NOP58
	ENSBTAG00000017376	Q53TQ3	INO80D
	ENSBTAG00000016334	O00487	PSMD14
	ENSBTAG00000005562	Q75T13	PGAP1
	ENSBTAG00000021691	O94925	GLS
	ENSBTAG00000007863	O00487	PSMD14
BTA3 (BBU6)	ENSBTAG00000014833	Q969M7	UBE2F
	ENSBTAG00000013961	P56524	HDAC4
	ENSBTAG00000045948	Q8WXJ9	ASB17
	ENSBTAG00000021634	Q8TEY7	USP33
	ENSBTAG00000007712	Q13103	SPP2
	ENSBTAG00000002030	Q6ZVZ8	ASB18
	ENSBTAG00000005864	P15509	CSF2RA
	ENSBTAG00000020761	Q6GQQ9	OTUD7B
	ENSBTAG00000003586	Q9Y672	ALG6
	ENSBTAG00000002024	Q9Y576	ASB1
	ENSBTAG00000003376	Q5VVJ2	MYSM1
	ENSBTAG00000026114	O75396	SEC22B
	ENSBTAG00000001586	P53611	RABGGTB
	ENSBTAG00000030706	Q9BVH7	ST6GALNAC5
	ENSBTAG00000005854	O60476	MAN1A2
	ENSBTAG00000038480	Q92643	PIGK
	ENSBTAG00000007309	Q6FI13	HIST2H2AA4
	ENSBTAG00000032456	O14772	FPGT
	ENSBTAG00000018447	Q9Y6X2	PIAS3
	ENSBTAG00000031012	Q8WWR8	NEU4
	ENSBTAG00000017764	Q99627	COPS8
	ENSBTAG00000039719	P21549	AGXT
	ENSBTAG00000017826	O43175	PHGDH
	ENSBTAG00000005797	Q96I15	SCLY
BTA4 (BBU8)	ENSBTAG00000003424	Q13324	CRHR2
	ENSBTAG00000011127	P63211	GNGT1
	ENSBTAG00000014381	Q9Y239	NOD1
	ENSBTAG00000000820	O15155	BET1
	ENSBTAG00000018185	Q9Y574	ASB4
	ENSBTAG00000027134	Q96A65	EXOC4
	ENSBTAG00000002674	Q92621	NUP205
	ENSBTAG00000038235	P61952	GNG11
	ENSBTAG00000014112	O14576	DYNC1I1
	ENSBTAG00000001036	P31937	HIBADH
BTA6 (BBU7)	ENSBTAG00000001600	Q9H2U2	PPA2
	ENSBTAG00000014226	O94855	SEC24D
	ENSBTAG00000014913	Q9H8P0	SRD5A3
	ENSBTAG00000002428	Q92564	DCUN1D4
	ENSBTAG00000003619	P49207	RPL34
	ENSBTAG00000014226	P49207	RPL34
BTA7 (BBU9)	ENSBTAG00000019718	O95633	FSTL3
	ENSBTAG00000003018	P63146	UBE2B
	ENSBTAG00000000753	Q6P161	MRPL54
	ENSBTAG00000005272	Q9BU89	DOHH
	ENSBTAG00000010982	Q9UH77	KLHL3
	ENSBTAG00000008396	O60262	GNG7
	ENSBTAG00000002796	Q9Y6B6	SAR1B
	ENSBTAG00000002098	Q8N2W9	PIAS4
	ENSBTAG00000007644	P62841	RPS15
	ENSBTAG00000018322	O95486	SEC24A
	ENSBTAG00000021226	P21796	VDAC1
	ENSBTAG00000013113	P49427	CDC34
	ENSBTAG00000019718	P54368	OAZ1
	ENSBTAG00000004112	P49419	ALDH7A1
	ENSBTAG00000018522	P62841	RPS15
	ENSBTAG00000009646	Q14353	GAMT
BTA9 (BBU10)	ENSBTAG00000005083	P22061	PCMT1
	ENSBTAG00000009275	P25398	RPS12
	ENSBTAG00000010204	O95394	PGM3
	ENSBTAG00000001360	P35442	THBS2
	ENSBTAG00000047902	Q9H0U6	MRPL18
	ENSBTAG00000001271	Q9NTX7	RNF146
	ENSBTAG00000001289	Q9UGC7	MTRF1L
	ENSBTAG00000000679	P17987	TCP1
	ENSBTAG00000002829	Q5VY80	RAET1L
	ENSBTAG00000018996	P00747	PLG
	ENSBTAG00000018810	Q9NRJ4	TULP4
	ENSBTAG00000010196	Q8NFH3	NUP43
	ENSBTAG00000034531	O60260	PRKN
	ENSBTAG00000012403	P25398	RPS12
	ENSBTAG00000001360	P05089	ARG1
BTA14 (BBU15)	ENSBTAG00000014357	Q96MF7	NSMCE2
	ENSBTAG00000009394	Q969P5	FBXO32
	ENSBTAG00000008409	P11940	PABPC1
	ENSBTAG00000016194	P62888	RPL30
	ENSBTAG00000020399	P01106	MYC
	ENSBTAG00000004906	P34741	SDC2
	ENSBTAG00000046358	O96020	CCNE2
	ENSBTAG00000016278	Q8WU17	RNF139
	ENSBTAG00000007943	P62888	RPL30
	ENSBTAG00000016278	O14977	AZIN1

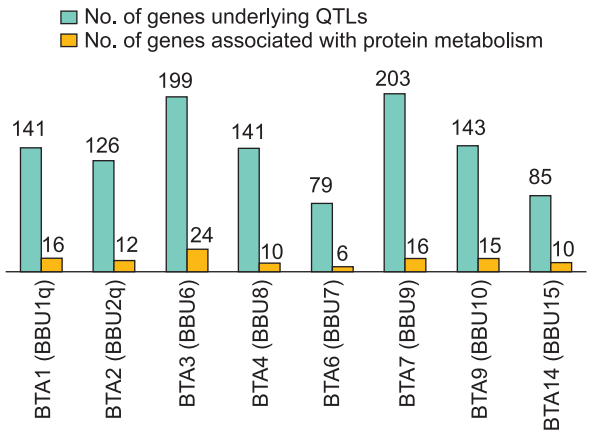

Fig. 3. Total number of genes underlying QTL regions and genes associated with protein metabolism.

cytoplasm, cytoplasmic part, intracellular cytosol and intracellular part were highly enriched. Gene Ontology terms which were highly enriched were 91 for biological function, 2 for molecular function and 26 for cellular component. The protein-protein interactions are depicted in Fig. 4.

Thus the multiple QTL models identified 69 QTLs in the 12 half sib families which were created for the purpose. The metaQTL analysis of the chromosomal regions identified through partial genome scan of 8 chromosomes. The analysis revealed 25 consensus regions on the chromosomes that were associated with protein percentage in buffaloes. Out of the 1117 genes which were identified using comparative genomics, 109 genes were associated with protein metabolism. These genes may be strong candidate genes for further analysis for their association

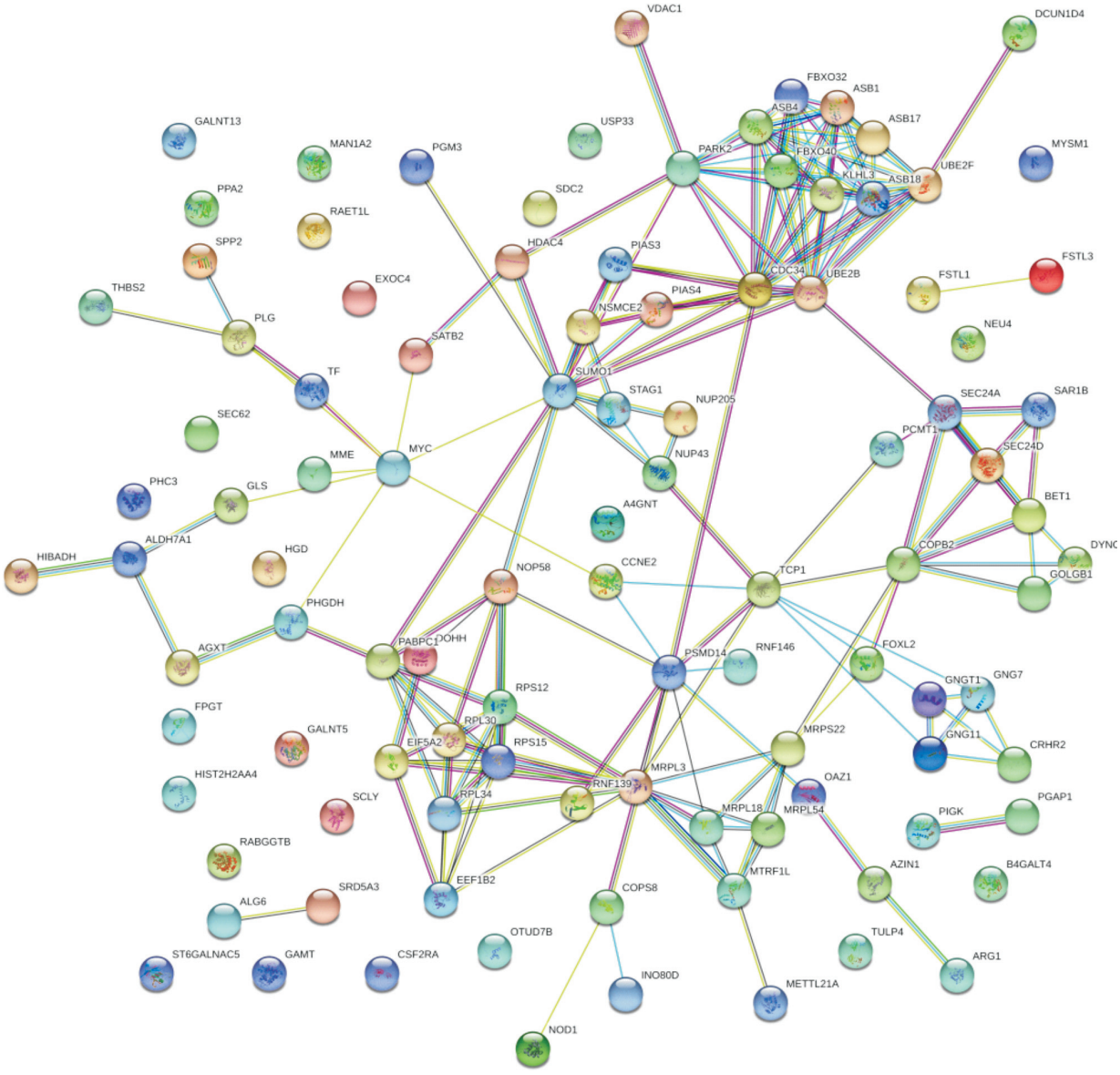

Fig. 4. Protein–protein interaction among 109 genes identified and underlying QTL regions in buffaloes and associated with protein metabolism (Supplementary figures are available Online).

with the milk protein percentage in buffaloes and utilised for marker assisted selection to increase milk protein percentage in buffaloes.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support of National Agriculture Innovation Project of Indian Council of Agricultural Research Grant no. 415401-02 under Component IV (Basic and Strategic Research in Agriculture and ICAR-Agricultural Consortium Platform on Genomics). The authors also acknowledge large number of field functionaries, block level and district level officers of BIRD, Allahabad who whole heartedly supported this scientific endeavour. The authors acknowledge the directions and monitoring of such a huge experiment by members of Consortium Advisory Committee of the project namely Drs A E Nivsarkar, S L Goswami, B K Joshi and Dinakar Raj.

REFERENCES

Alain K, Karrow N A, Thibault C, St-Pierre J, Lessard M and Bissonnette N. 2009. Osteopontin: an early innate immune marker of *Escherichia coli* mastitis harbors genetic polymorphisms with possible links with resistance to mastitis. *BMC Genomics* **10**: 444.

Amaral M E J, Grant J R, Riggs P K, Filho N B S E A R, Goldammer T, Weikard R, Brunner R M, Kochan K J, Greco A J, Jeong J, Cai Z, Lin G, Prasad A, Kumar S, Mathew G P S B, Kumar M A, Miziara M N, Mariani P, Caetano A R, Galvão S R, Tantia M S, Vijh R K, Mishra B, Bharani Kumar S T, Pelai V A, Santana A M, Fornitano L C, Jones B C, Tonhati H, Moore S, Stothard P and Womack J E. 2008. A first generation whole genome RH map of the river buffalo with comparison to domestic cattle. *BMC Genomics* **9**: 631–41.

Bagnato A, Schiavini F, Rossoni A, Maltecca C, Dolezal M, Medugorac I, Sölkner J, Russo V, Fontanesi L and Friedmann A. 2008. Quantitative trait loci affecting milk yield and protein percentage in a three-country Brown Swiss population. *Journal of Dairy Science* **91**(2): 767–83.

Buitenhuis B, Poulsen N A, Gebreyesus G and Larsen L B. 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. *BMC Genetics* **17**: 114.

Chamberlain A J, Hayes B J, Savin K, Bolormaa S, McPartlan H C, Bowman P J, Van der Jagt C, MacEachern S, Goddard M E. 2012. Validation of single nucleotide polymorphisms associated with milk production traits in dairy cattle. *Journal of Dairy Science* **95**(2): 864–75.

Churchill G A and Doerge R W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963–71.

Coffman C J, Doerge R W, Wayne M and McIntyre L M. 2003. Intersection tests for single marker QTL analysis can be more powerful than two marker QTL analysis. *BMC Genetics* **4**: 10.

Cole J B, Null D J and VanRaden P M. 2016. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. *Journal of Dairy Science* **99**(9): 7274–88.

Cole J B, Wiggans G R, Ma L, Sonstegard T S, Lawlor T J Jr, Crooker B A, Van Tassell 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**: 408.

Collard B C Y, Jahufer M Z Z, Brouwer J B and Pang E C K. 2005. An introduction to marker, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement. The basic concepts. *Euphytica* **142**: 169–96.

Corva P M and Medrano J F. 2001. Quantitative trait loci (QTLs) mapping for growth traits in the mouse. A review. *Genetic Selection Evolution* **33**: 105–32.

Croft D, Mundo A F, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar M R, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L and D'Eustachio P. 2014. The Reactome pathway knowledgebase. *Nucleic Acids Research* **42**: D472–7.

Doerge R W, Zeng Z B and Weir B S. 1997. Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Statistical Science* **13**: 195–219.

Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamovsky V, Webber M, Weiser J, Williams M, Wu G, Stein L, Hermjakob H, D'Eustachio P. 2016. The Reactome pathway Knowledgebase. *Nucleic Acids Research* **44**(D1): D481–87.

Fang M, Fu W, Jiang D, Zhang Q, Sun D, Ding X and Liu J. 2014. A multiple-SNP approach for genome-wide association study of milk production traits in Chinese Holstein cattle. *PLoS ONE* **9**(8): e99544.

Goffinet B and Gerber S. 2000. Quantitative trait loci: a meta-analysis. *Genetics* **155**(1): 463–73.

Haley C and Knott S. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–24.

Hayes B and Goddard M E. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genetic Selection Evolution* **33**: 209–29.

Heyen D W, Weller J I, Ron M, Band M, Beever J E, Feldmesser E, Da Y, Wiggans G R, VanRaden P M and Lewin H A. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiology Genomics* **1**(3): 165–75.

Jiang J, Shen B, O'Connell J R, VanRaden P M, Cole J B and Ma L. 2017. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. *BMC Genomics* **18**(1): 425.

Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y and Zhang Q. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS ONE* **5**(10): e13661.

Kao C H. 2000. On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics* **156**: 855–65.

Khatkar M, Randhawa I and Raadsma H. 2014. Meta-assembly of genomic regions and variants associated with female reproductive efficiency in cattle. *Livestock Science* **166**(1): 144–57.

Knott S and Haley C. 1992. Aspects of maximum likelihood methods for mapping of quantitative trait loci in line crosses. *Genetical Research* **60**: 139–52.

Mackay T F C. 2001. The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**: 303–39.

Mosig M O, Lipkin E, Khutoreskaya G, Tchourzyna E, Soller M

and Friedmann A. 2001. A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* **157**(4): 1683–98.

Pimentel E C, Bauersachs S, Tietze M, Simianer H, Tetens J, Thaller G, Reinhardt F, Wolf E and Konig S. 2011. Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Animal Genetic* **42**(3): 251–62.

Plante Y, Gibson J P, Nadesalingam J, Mehrabani-Yeganeh H, Lefebvre S,Vandervoort G and Jansen G B. 2001. Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *Journal of Dairy Science* **84**(6): 1516–24.

Pryce J E, Bolormaa S, Chamberlain A J, Bowman P J, Savin K, Goddard M E and Hayes B J. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *Journal of Dairy Science* **93**(7): 3331–45.

Rebaï A, Goffinet B and Mangin B. 1995. Comparing powers of different methods for QTL detection. *Biometrics* **51**: 87–99.

Ron M, Feldmesser E, Golik M, Tager-Cohen I, Kliger D, Reiss V, Domochovsky R, Alus O, Seroussi E, Ezra E and Weller J I. 2004. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. *Journal of Dairy Science* **2**(87): 476–90.

Russo V, Fontanesi L, Dolezal M, Lipkin E, Scotti E, Zambonelli P, Dall'Olio S, Bigi D, Davoli R, Canavesi F, Medugorac I, Foster M, Solkner J, Schiavini F, Bagnato A and Soller M. 2012. A whole genome scan for QTL affecting milk protein percentage in Italian Holstein cattle, applying selective milk DNA pooling and multiple marker mapping in a daughter design. *Animal Genetics* **43**(Suppl 1): 72–86.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou K P, Kuhn M, Bork P, Jensen L J and von Mering C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43**: D447–52.

Viale E, Tiezzi F, Maretto F, De Marchi M, Penasa M and Cassandro M. 2017. Association of candidate gene polymorphisms with milk technological traits, yield, composition, and somatic cell score in Italian Holstein-Friesian sires. *Journal of Dairy Science* **100**(9): 7271–81.

Vijh R K. 2013. Final Report of 'Quantitative trait loci for milk yield, fat and protein percentage in buffaloes' of World Bank Funded Project 'National Agriculture Innovation Project' of Indian Council of Agricultural Research, Grant No. 415401-02. ICAR-National Bureau of Animal Genetic Resources.

Vijh R K. 2014. Identification of quantitative trait loci for milk yield, fat and protein percentage in buffaloes. *Buffalo Reference Family Germplasm Catalogue*. Published by Indian Council of Agricultural Research. pp 671.

Vijh R K, Upasna S and Gokhle S B. 2018. Creation of a large reference family with phenotype recording and genotype data generation in buffaloes. *Indian Journal of Animal Sciences* **88**(2): 59–65.