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

UPASNA SHARMA1, PRIYANKA BANERJEE2, JYOTI JOSHI3, PRERNA KAPOOR4 and RAMESH KUMAR VIJH5

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 h²=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 milk 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.

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 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.
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 regions 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.

| 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 R² 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  | R²     | P     |
|-------|--------------|------|--------|-------|
|       | Marker name  | (F value) |       |       |
| Sire1 | 1 | BMS4037      | 6.628 | 0.018 | 0.010** |
|       | 2 | BMS1788      | 4.876 | 0.011 | 0.028*  |
|       | 4 | BMS1260      | 10.609| 0.052 | 0.001** |
|       | 4 | BMS2571      | 4.613 | 0.017 | 0.032*  |
|       | 6 | MNB-66       | 5.104 | 0.011 | 0.024*  |
|       | 7 | DIK2819      | 6.773 | 0.018 | 0.010** |
|       | 7 | BMS6117      | 8.990 | 0.041 | 0.003** |
|       | 14 | DIK4105    | 6.413 | 0.020 | 0.012*  |
|       | 17 | BL1029      | 3.991 | 0.018 | 0.047*  |
| Sire2 | 1 | ILSTS026     | 5.608 | 0.028 | 0.019*  |
|       | 3 | DIK1057      | 5.969 | 0.041 | 0.015*  |
|       | 3 | BMS1260      | 6.102 | 0.028 | 0.014*  |
|       | 4 | BMS1074      | 6.148 | 0.038 | 0.014*  |
|       | 6 | DIK1513      | 4.758 | 0.041 | 0.030*  |
|       | 14 | DIK2585     | 7.455 | 0.052 | 0.007** |
|       | 6 | MNB-66       | 4.198 | 0.011 | 0.042*  |
|       | 7 | BB719        | 6.945 | 0.039 | 0.009** |
| Sire3 | 4 | DIK4373      | 4.693 | 0.019 | 0.033*  |
|       | 6 | DIK1515      | 4.692 | 0.028 | 0.033*  |
| Sire4 | 2 | BM1223       | 5.593 | 0.038 | 0.019*  |
|       | 4 | DIK4373      | 3.952 | 0.000 | 0.048*  |
|       | 7 | BM1260       | 5.203 | 0.022 | 0.023*  |
|       | 7 | BM053        | 4.709 | 0.023 | 0.031*  |
|       | 10 | BMS1979    | 4.876 | 0.010 | 0.028*  |
|       | 14 | BMS003      | 6.908 | 0.019 | 0.009** |
| Sire5 | 8 | BMS4011      | 5.494 | 0.023 | 0.020*  |
|       | 2 | BMS003       | 6.408 | 0.019 | 0.009** |
|       | 6 | BMS1987      | 4.287 | 0.019 | 0.039*  |
|       | 8 | BMS2519      | 5.457 | 0.025 | 0.020*  |
|       | 8 | DIK3024      | 9.605 | 0.011 | 0.002** |
|       | 10 | BMS4311     | 7.068 | 0.028 | 0.006** |
|       | 14 | BMS2055      | 5.656 | 0.023 | 0.018*  |
| Sire6 | 1 | BMS4011      | 4.043 | 0.066 | 0.048*  |
|       | 3 | ILSTS029     | 4.243 | 0.065 | 0.043*  |
|       | 7 | BMS3020      | 5.023 | 0.068 | 0.028*  |
|       | 6 | MNB-66       | 4.397 | 0.054 | 0.040*  |
|       | 9 | BMS322       | 7.361 | 0.020 | 0.008** |
|       | 6 | BMS1909      | 4.171 | 0.061 | 0.045*  |
|       | 9 | BMS2063      | 4.596 | 0.032 | 0.035*  |
|       | 14 | BM8215      | 5.595 | 0.075 | 0.020*  |
| Sire7 | 5 | BMS4048      | 3.909 | 0.012 | 0.049*  |
|       | 3 | CSSM054      | 4.279 | 0.011 | 0.039*  |
|       | 7 | BMS3020      | 5.569 | 0.025 | 0.019*  |
|       | 8 | DIK2511      | 5.112 | 0.009 | 0.024*  |
|       | 3 | BMS827       | 4.181 | 0.011 | 0.042*  |
|       | 6 | DIK1513      | 6.101 | 0.015 | 0.014*  |
|       | 6 | DIK2585      | 4.506 | 0.011 | 0.034*  |
|       | 5 | BM1329       | 6.609 | 0.018 | 0.010*  |
|       | 8 | DIK3024      | 4.526 | 0.007 | 0.034*  |
|       | 7 | DIK4204      | 6.087 | 0.022 | 0.014*  |
|       | 9 | BM5190       | 6.461 | 0.021 | 0.011*  |
|       | 7 | DIK2816      | 3.879 | 0.007 | 0.050*  |
|       | 14 | BM8215      | 7.587 | 0.027 | 0.017*  |
| Sire8 | 3 | ILSTS029     | 6.301 | 0.024 | 0.012*  |
|       | 4 | BMS1237      | 6.316 | 0.017 | 0.012*  |
|       | 5 | DIK4816      | 5.718 | 0.004 | 0.017*  |
|       | 7 | DIK1517      | 4.652 | 0.022 | 0.032*  |
|       | 11 | BL1411       | 6.127 | 0.021 | 0.014*  |
|       | 4 | MNB-66       | 5.967 | 0.029 | 0.015*  |
|       | 7 | BMS522       | 12.011| 0.032 | 0.001***|

(Contd...)
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 | Position on LOD | % variance | F value | P value |
|----------|-------------|-----------------|------------|---------|--------|
| Sire1    | 3           | 19.3            | 1.508      | 1.508   | 3.251  | 0.040* |
|          | 3           | 84.8            | 1.731      | 2.436   | 3.739  | 0.025* |
|          | 3           | 94              | 2.031      | 2.866   | 4.399  | 0.013* |
|          | 3           | 104             | 3.272      | 4.634   | 2.886  | 0.029* |
|          | 3           | 64              | 4.431      | 6.336   | 3.263  | 0.004** |
|          | 3           | 70              | 2.556      | 3.621   | 5.559  | 0.004** |
|          | 7           | 60              | 1.228      | 1.709   | 2.641  | 0.073P |
|          | 9           | 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 | 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  | 0.000*** |
|          | 4           | 74              | 3.123      | 6.884   | 2.175  | 0.049* |
|          | 6           | 10              | 2.267      | 4.933   | 6.675  | 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           | 2.621      | 11.982  | 2.126  | 0.062P |
|          | 2           | 30.8            | 3.685      | 13.495  | 2.491  | 0.035* |
|          | 3           | 69.3            | 3.109      | 11.265  | 2.491  | 0.078P |
|          | 6           | 20              | 3.309      | 11.591  | 2.16   | 0.058P |
|          | 7           | 70              | 1.365      | 4.685   | 2.532  | 0.087P |
|          | 2@30.8:6@20.0 | 2.491          | 8.822      | 2.364   | 0.060P |
| Sire4    | 1           | 115.11          | 2.515      | 5.026   | 1.839  | 0.094P |
|          | 2           | 10.77           | 2.578      | 5.155   | 1.866  | 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   | 2.688  | 0.072P |
|          | 7           | 14              | 1.57       | 2.519   | 3.395  | 0.035* |
|          | 110.5       | 74.2            | 2.891      | 4.642   | 2.126  | 0.051P |
| Sire6    | 1           | 110.5           | 43.562     | 73.719  | 158.077| 0.000*** |
|          | 110.5       | 5.274           | 4.583      | 10.05   | 0.000***|
|          | 16.8        | 2.031           | 0.582      | 3.746   | 0.031* |
|          | 29.3        | 2.217           | 0.64       | 4.115   | 0.022* |
|          | 29.3        | 2.144           | 1.666      | 6.652   | 0.034* |
|          | 20          | 2.232           | 1.739      | 3.814   | 0.029* |

**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 (81.29–90.40) | 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 (118.34–125.00) |
| 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 (61.62–68.12) | –                 | –                 |

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

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.
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 meta-QTL positions have been identified. These regions in cattle have been identified. These regions in cattle have been reported by several authors in cattle (Cole 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 syteny between the cattle and buffalo chromosomes and have been established using the radiation hybrid panels (Amaral et al. 2008). Similarly, the syteny between cattle chromosomes and human chromosomes is well established and available in the Ensembl database (https://www.ensembl.org). Thus using the syteny 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 glycin, 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 underlie the identified QTL regions in buffaloes | Mapped entities in Reactome database | HGNC names of mapped genes |
|--------------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------|----------------------------|
| BTA1 (BBU1q)                                           | ENSBTAG000000014843 P35060 COPB2                                               |                                    |                            |
|                                                       | ENSBTAG000000007273 Q12841 FSTL1                                              |                                    |                            |
|                                                       | ENSBTAG000000047608 Q9GZV4 EIF5A                                                |                                    |                            |
|                                                       | ENSBTAG00000001390 Q8WVM7 STAG1                                               |                                    |                            |
|                                                       | ENSBTAG00000002075 Q14789 GOLGB1                                              |                                    |                            |
|                                                       | ENSBTAG000000014508 QSNDX5 PHC3                                               |                                    |                            |
|                                                       | ENSBTAG000000013938 O60513 B4GALT4                                           |                                    |                            |
|                                                       | ENSBTAG00000008862 Q9UNA3 A4GNT                                                |                                    |                            |
|                                                       | ENSBTAG000000010451 Q9UBH90 FBXO40                                             |                                    |                            |
|                                                       | ENSBTAG000000022155 P58012 FOXL2                                                |                                    |                            |
|                                                       | ENSBTAG000000030942 Q99442 SEC62                                               |                                    |                            |
|                                                       | ENSBTAG000000035437 P08473 MME                                                   |                                    |                            |
|                                                       | ENSBTAG00000014508 Q8NDX5 PHC3                                               |                                    |                            |
|                                                       | ENSBTAG00000001277 P09001 MRPL3                                               |                                    |                            |
|                                                       | ENSBTAG000000015895 Q93099 HGD                                                 |                                    |                            |
| BTA2 (BBU2q)                                           | ENSBTAG00000009517 Q8WXB1 METTL21A                                             |                                    |                            |
|                                                       | ENSBTAG000000022169 Q9UPW6 SATB2                                               |                                    |                            |
|                                                       | ENSBTAG00000014407 Q7Z7M9 GALNT5                                              |                                    |                            |
|                                                       | ENSBTAG000000021979 P63165 SUMO1                                               |                                    |                            |
|                                                       | ENSBTAG00000009859 P24534 EEF1B                                               |                                    |                            |
|                                                       | ENSBTAG00000020416 Q8IUC8 GALNT13                                              |                                    |                            |
|                                                       | ENSBTAG00000005476 Q9Y2X3 NOP58                                                |                                    |                            |
|                                                       | ENSBTAG000000016334 Q53TQ3 INO80D                                              |                                    |                            |
|                                                       | ENSBTAG00000007712 Q13103 SPP2                                                 |                                    |                            |
|                                                       | ENSBTAG00000002030 Q6ZVZ8 ASB1                                                 |                                    |                            |
|                                                       | ENSBTAG00000005864 P15509 CSF2RA                                               |                                    |                            |
|                                                       | ENSBTAG00000005562 Q7ST13 PGAP1                                               |                                    |                            |
|                                                       | ENSBTAG000000021691 Q94925 GLS                                                 |                                    |                            |
|                                                       | ENSBTAG00000007863 Q969M7 UBE2F                                               |                                    |                            |
| BTA3 (BBU6)                                            | ENSBTAG000000013961 P56524 HDC4                                                |                                    |                            |
|                                                       | ENSBTAG000000045948 Q8WXJ9 ASB1                                                |                                    |                            |
|                                                       | ENSBTAG000000021634 Q8TEY7 USP33                                               |                                    |                            |
|                                                       | ENSBTAG0000000014407 Q7Z7M9 GALNT5                                              |                                    |                            |
|                                                       | ENSBTAG000000021979 P63165 SUMO1                                               |                                    |                            |
|                                                       | ENSBTAG00000009859 P24534 EEF1B                                               |                                    |                            |
|                                                       | ENSBTAG000000020416 Q8IUC8 GALNT13                                              |                                    |                            |
|                                                       | ENSBTAG00000005476 Q9Y2X3 NOP58                                                |                                    |                            |
|                                                       | ENSBTAG000000016334 Q53TQ3 INO80D                                              |                                    |                            |
|                                                       | ENSBTAG00000007712 Q13103 SPP2                                                 |                                    |                            |
|                                                       | ENSBTAG00000002030 Q6ZVZ8 ASB1                                                 |                                    |                            |
|                                                       | ENSBTAG00000005864 P15509 CSF2RA                                               |                                    |                            |
|                                                       | ENSBTAG00000005562 Q7ST13 PGAP1                                               |                                    |                            |
|                                                       | ENSBTAG000000021691 Q94925 GLS                                                 |                                    |                            |
|                                                       | ENSBTAG00000007863 Q969M7 UBE2F                                               |                                    |                            |
| BTA4 (BBU8)                                            | ENSBTAG0000001127 P63211 GNTG1                                               |                                    |                            |
|                                                       | ENSBTAG0000001012 Q8SWWR8 NEU4                                                |                                    |                            |
|                                                       | ENSBTAG00000017764 Q9627 COP8S                                                |                                    |                            |
|                                                       | ENSBTAG00000035719 P21549 AGX7                                                |                                    |                            |
|                                                       | ENSBTAG00000017826 Q43175 PHGDL                                                |                                    |                            |
|                                                       | ENSBTAG00000005797 Q96155 SCY                                                  |                                    |                            |
|                                                       | ENSBTAG00000003424 Q13324 CRHR2                                               |                                    |                            |
| BTA9 (BBU10)                                           | ENSBTAG00000005083 P52061 PCMT1                                               |                                    |                            |
|                                                       | ENSBTAG000000018522 P62841 RPS15                                               |                                    |                            |
|                                                       | ENSBTAG00000009646 Q14353 GAMT                                                |                                    |                            |
| BTA14 (BBU15)                                          | ENSBTAG00000009394 Q669P5 FBXO32                                              |                                    |                            |
|                                                       | ENSBTAG00000008409 P11940 PCMT1                                               |                                    |                            |
|                                                       | ENSBTAG00000016194 P62888 RPL30                                               |                                    |                            |
|                                                       | ENSBTAG00000020399 P01106 MYC                                                  |                                    |                            |
|                                                       | ENSBTAG000000046358 Q96020 CCNE2                                              |                                    |                            |
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.
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, Miziaara M N, Mariani P, Caetano A R, Galvão S R, Tantia M S, Vijh R K, Mishra B, Bharti Kumar S T, Peli V A, Santana A M, Forntano 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, Mallecca 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 McEntyre L M. 2003. Intersektion 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 I. 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, Milacic M, Rothfels K, Shamoysky 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, Kornfing M, McKay S, Matthews L, May B, Milacic M, Rothfels K, Shamoysky 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): D415–87.

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 L, 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 Genetics* 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.

Khatarek M, Randhawa I and Raadmsa 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.

Mogis M O, Lipkin E, Khutoreskaya G, Tchourzyza 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 König 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.

Rebai 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, Upasma 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.