



OPEN ACCESS

EDITED BY

Ravesh Kumar Pathak,
Chung-Ang University, Republic of Korea

REVIEWED BY

Ravendra Chauhan,
Rhodes University, South Africa
Divya Mishra,
University of Allahabad, India
Akshay Singh,
Indian Council of Agricultural Research
(ICAR), India

*CORRESPONDENCE

Dwijesh Chandra Mishra
✉ dwij.mishra@gmail.com

†These authors have contributed equally to this work

RECEIVED 07 February 2023

ACCEPTED 11 April 2023

PUBLISHED 12 May 2023

CITATION

Mishra DC, Bhati J, Yadav S, Avashthi H, Sikka P, Jerome A, Balhara AK, Singh I, Rai A and Chaturvedi KK (2023) Comparative expression analysis of water buffalo (*Bubalus bubalis*) to identify genes associated with economically important traits. *Front. Vet. Sci.* 10:1160486. doi: 10.3389/fvets.2023.1160486

COPYRIGHT

© 2023 Mishra, Bhati, Yadav, Avashthi, Sikka, Jerome, Balhara, Singh, Rai and Chaturvedi. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comparative expression analysis of water buffalo (*Bubalus bubalis*) to identify genes associated with economically important traits

Dwijesh Chandra Mishra^{1*†}, Jyotika Bhati^{1†}, Sunita Yadav¹, Himanshu Avashthi¹, Poonam Sikka², Andonissamy Jerome², Ashok Kumar Balhara², Inderjeet Singh², Anil Rai¹ and Krishna Kumar Chaturvedi¹

¹ICAR-Indian Agricultural Statistics Research Institute, Indian Council of Agricultural Research (ICAR), PUSA, New Delhi, India, ²ICAR-Central Institute for Research on Buffaloes, Indian Council of Agricultural Research (ICAR), Hisar, India

The milk, meat, skins, and draft power of domestic water buffalo (*Bubalus bubalis*) provide substantial contributions to the global agricultural economy. The world's water buffalo population is primarily found in Asia, and the buffalo supports more people per capita than any other livestock species. For evaluating the workflow, output rate, and completeness of transcriptome assemblies within and between reference-free (RF) *de novo* transcriptome and reference-based (RB) datasets, abundant bioinformatics studies have been carried out to date. However, comprehensive documentation of the degree of consistency and variability of the data produced by comparing gene expression levels using these two separate techniques is lacking. In the present study, we assessed the variations in the number of differentially expressed genes (DEGs) attained with RF and RB approaches. In light of this, we conducted a study to identify, annotate, and analyze the genes associated with four economically important traits of buffalo, viz., milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency. A total of 14,201 and 279 DEGs were identified in RF and RB assemblies. Gene ontology (GO) terms associated with the identified genes were allocated to traits under study. Identified genes improve the knowledge of the underlying mechanism of trait expression in water buffalo which may support improved breeding plans for higher productivity. The empirical findings of this study using RNA-seq data-based assembly may improve the understanding of genetic diversity in relation to buffalo productivity and provide important contributions to answer biological issues regarding the transcriptome of non-model organisms.

KEYWORDS

water buffalo, transcriptome, annotation, GO terms, SSRs

Introduction

The domestic water buffalo (*Bubalus bubalis*) marks a key impact on the global agricultural economy through milk, meat, and draft power. The world's water buffalo population is largely found in Asia, and most people consider it the most promising livestock species for their livelihood (1, 2). Asia accounts for 97% of the total buffalo production with the largest population in India (>100 million) (2). More than half of the milk produced in

India comes from buffaloes, which also produce milk with higher levels of fat, particularly saturated fatty acids, than cattle (2). Buffaloes are resilient to the harsher environment and resistant to several bovine tropical diseases (1), thus may have better feed convergence while surviving on poor-quality roughage than cattle. Recent studies cataloging differentially expressed genes (DEGs) and variants (3, 4) with respect to important performance traits in water buffalo corroborate with the functional genetic diversity in this species.

Milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency traits define the overall productivity of buffaloes. Buffalo milk has the significance of having higher concentrations of fat, lactose, protein, ash, calcium, and vitamins A and C while having lower concentrations of cholesterol and the blue-green pigment (biliverdin) (5). Additionally, buffalo milk has bioactive pentasaccharides and gangliosides, which are absent in cow milk (6). Therefore, the study of genes related to milk volume is very important. The age at first calving can be used to determine a buffalo's fertility and productivity. Buffalo productivity is affected by delayed puberty onset and inadequate consecutive estrus detection (7). Reproductive efficiency is the primary factor affecting the productivity of buffaloes, which comprise early age at first calving (AFC) and optimum service period between the calvings (post-partum cyclicity) throughout the reproductive span in life. Thus, the identification of genes and the variants associated with these traits may support selective breeding for genetic improvement. Improved immunity is equally significant to propagate uterine cleansing to facilitate an early resumption of ovarian cyclicity (8, 9). Feed conversion efficiency (FCE) is defined as a dry matter intake (DMI) per unit body weight (g/day) gain determined as residual feed intake (RFI). It represents the difference in actual and predicted DMI of each individual heifer (10). Feed conversion efficiency is a heritable trait governed by common biomolecules as growth hormones are associated with milk volume, age at first calving, and post-partum cyclicity (11, 12).

Several studies have confirmed the discovery of differentially expressed as well as novel genes in mammals such as humans, buffaloes, sheep, goats, and pigs (3, 13–17). The genetic link and diversity among various buffalo breeds have primarily been studied using restriction fragment length polymorphism (RFLP) (18), random amplified polymorphic DNA (RAPD) (19), single nucleotide polymorphism (SNP) (4, 20), and simple sequence repeat (SSR) (21) markers. SSR markers have proven to be an incredibly powerful tool for researching genetic divergence and/or genetic resource conservation (22). Identification and characterization of genome-wide DEGs related to reproduction and production traits can be widely used for selective breeding, which may enhance productivity in buffaloes (23, 24).

Considering this, an attempt was made to identify the variants related to important traits, i.e., milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency, using transcriptomic data to improve breeding plans in water buffaloes. DEGs were identified, characterized, and annotated, in order to accelerate performance in buffaloes through molecular breeding. This is a unique study identifying the functional classifications of genes, variants, and SSRs related to desired traits in *B. bubalis*.

Materials and methods

Milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency were the four different traits for which datasets were collected. Four samples were selected for each trait (two each of low and high expression). The complete workflow of the study is presented in Figure 1.

Ethics statement

Animals ($n = 16$) were selected as per the referred design for selective genotyping of buffaloes, based on performance phenotype recorded at ICAR-Central Institute for Research on Buffaloes (ICAR-CIRB), Hisar, Haryana, India. Genotype data were generated for selected genotypes through outsourced services hired by the institute. Animals were maintained under farm management at the institute, and the experiment design was approved by the Institute Animal Ethics Committee (IAEC) with ethics approval number—406/GO/RBI/L/01/CPCSEA.

Animals and tissue collection

Whole blood tissues of individual animals were selected from unrelated pedigrees having extreme performance levels for complex traits as follows: milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency. Each trait had four samples comprising two each of low and high expressions.

RNA extraction, library preparation, and sequencing

For the transcriptome analyses of expression patterns in low- and high-performing Murrah buffaloes, cDNA was generated using a routine RNA library preparation HiSeq protocol developed by Illumina Technologies (San Diego, CA), using 1 μ g of total RNA as input. Using the High-Capacity cDNA Reverse Transcription kit (Life Technologies, Frederick, Maryland, USA), mRNA was first isolated from total RNA by performing a polyA selection step, followed by the construction of paired-end sequencing libraries with an insert size of \sim 300 bp. In brief, polyA selected RNA was cleaved as per Illumina protocol, and the cleaved fragments were used to generate first-strand cDNA using Super Script II reverse transcriptase and random hexamers. Subsequently, the second strand cDNA was synthesized with RNaseH and DNA polymerase enzyme, followed by adapter ligation and end-repair steps. The resulting products were amplified via PCR, and cDNA libraries were then purified and validated using the Agilent 2200 Tape Station system (Agilent Technologies Brasil Ltda, São Paulo, SP, Brazil). Paired-end sequencing was performed using the Illumina HiScanSQ platform. Samples were multiplexed with unique hexamer barcodes and run on multiple lanes to obtain 2×100 bp reads. Paired-end FASTQ files were subjected to standard quality control based on Phred scores of >20 , using the NGSQC Tool kit v2.2 (25) to obtain high-quality (HQ) filtered reads.

Transcriptome assembly and optimization

Raw reads from the four sets generated from the animal samples using Illumina HiSeq were filtered to generate clean data to remove adaptor sequences, reads with ambiguous sequences “N,” low-quality reads, and reads that were mostly repeated bases, such as polyT tracts using Trimmomatic 0.39 (26). The trimmed reads are evaluated with FastQC (27), a Java-based, quality control tool for high-throughput sequence data.

After obtaining clean reads and quality checks, RF transcriptome assembly was conducted with Trinity software v2.8.6 with default parameters (28, 29). Only assembled transcripts with lengths of >300 bp were included in further analysis.

Simultaneously, the raw reads were mapped to the *B. bubalis* genome (UOA_WB_1 accessed from https://www.ncbi.nlm.nih.gov/assembly/GCF_003121395.1/) with TopHat2 v2.0.13 (30), using Bowtie2 v2.2.6 (31) as the underlying aligner. Reads aligning to the UOA_WB_1 build were quantified, which disregarded any read/read pair that aligned to more than one location or more than one gene at a single location.

Differential expression analysis

For the RF assembly results, differential transcript expression for different datasets was calculated using an exact test in the Bioconductor R package (32) edgeR [Empirical Analysis of Digital Gene Expression Data in R (33)]. We used RSEM [RNA-Seq by Expectation-Maximization (34)] to generate read counts for the optimized assembled transcriptome to input into edgeR. EdgeR normalizes raw input data using a trimmed mean of M-values (TMM), and transcripts with artificially low counts (<1 count across all samples) after normalization were excluded before differential expression analysis was completed. The transcript level was quantified in terms of Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Differential expression (DE) was detected using the edgeR Bioconductor package with a log₂ fold change threshold of 2.

Differential expression analysis for the *B. bubalis* RB assembly was conducted using the Cufflink analysis tool between different samples of the same traits in pairs (high and low yielding). DE genes with log₂ fold change of ≥2.0, an adjusted *p*-value (padj) of < 0.05, and an adjusted FDR of <0.05 were subjected to further analyses.

Functional annotation of genes

Functional analysis of the DEGs was performed using Blast2GO v 2.5 (35). Blast2GO is a gene ontology-based annotation tool and found to be effective in the functional characterization of sequence data. The DEGs homologous with annotated proteins in the nr database were selected for functional characterization based on the maximum *E*-value (1E-3) and the minimum alignment size (HSP length 33) using BLASTX. The DEG sequences were then categorized according to the GO vocabularies into three categories, i.e., molecular function, biological process, and cellular component. The distribution of GO terms was analyzed at level 2

of the Directed Acyclic Graphs. Annotated DEGs were analyzed for pathway identification using KEGG.

SSR mining

Simple sequence repeats were identified using the MISA tool from the DEGs. The chromosome-wise distribution of DEGs, SSRs, and SNPs [extracted from DDRAD sequence data (4)] was graphically mapped using the CIRCOS (version 0.69) visualization tool.

Results

Analysis of our data with both cattle and water buffalo reference assemblies gave varied results for differential expression and annotation among different traits, viz., milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency.

Assembly benchmarking

A total of 857 million raw reads were generated (428,671,371 paired-end reads) by Illumina sequencing of the 16 *B. bubalis* samples for four traits, viz., milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency, with an average of ~26.7 million reads per sample. From these, ~773 million reads (90.7%) were attained after removing the adapters and trimming for quality. These post-cleaning reads passed the minimum quality standards of FastQC.

After read filtering, clean reads were assembled into 488,811, 86,054, 451,596, and 451,596 contigs, reaching a total length of 482,785,524 bp, 81,971,386 bp, 431,765,546 bp, and 529,684,800 bp for “milk volume,” “age at first calving,” “post-partum cyclicity,” and “feed conversion efficiency” traits, respectively. The average length of assembled contigs was 406, 383, 390, and 405 bp and N50 of 1,606, 1,728, 1,588, and 1,588 bp for “milk volume,” “age at first calving,” “post-partum cyclicity,” and “feed conversion efficiency” traits, respectively.

During the course of the abovementioned analysis, the *B. bubalis* genome was sequenced in 2019, and we proceeded with the RB assembly to compare the two assembly results. To evaluate the assembly quality, we mapped the Illumina clean reads on the water buffalo reference genome (UOA_WB_1, Accession GCA_003121395.1). Approximately 91.22% of the paired-end reads were mapped properly.

Differential expression analysis

A total of 14,201 and 279 DEGs were identified corresponding to RF and RB assembly, respectively, in four traits. The DEGs identified through RF assembly were more as compared to RB assembly. The number of upregulated genes was 7,190 and 126 while the downregulated genes were 7,011 and 153, respectively, expressed in RF and RB assembly (Table 1).

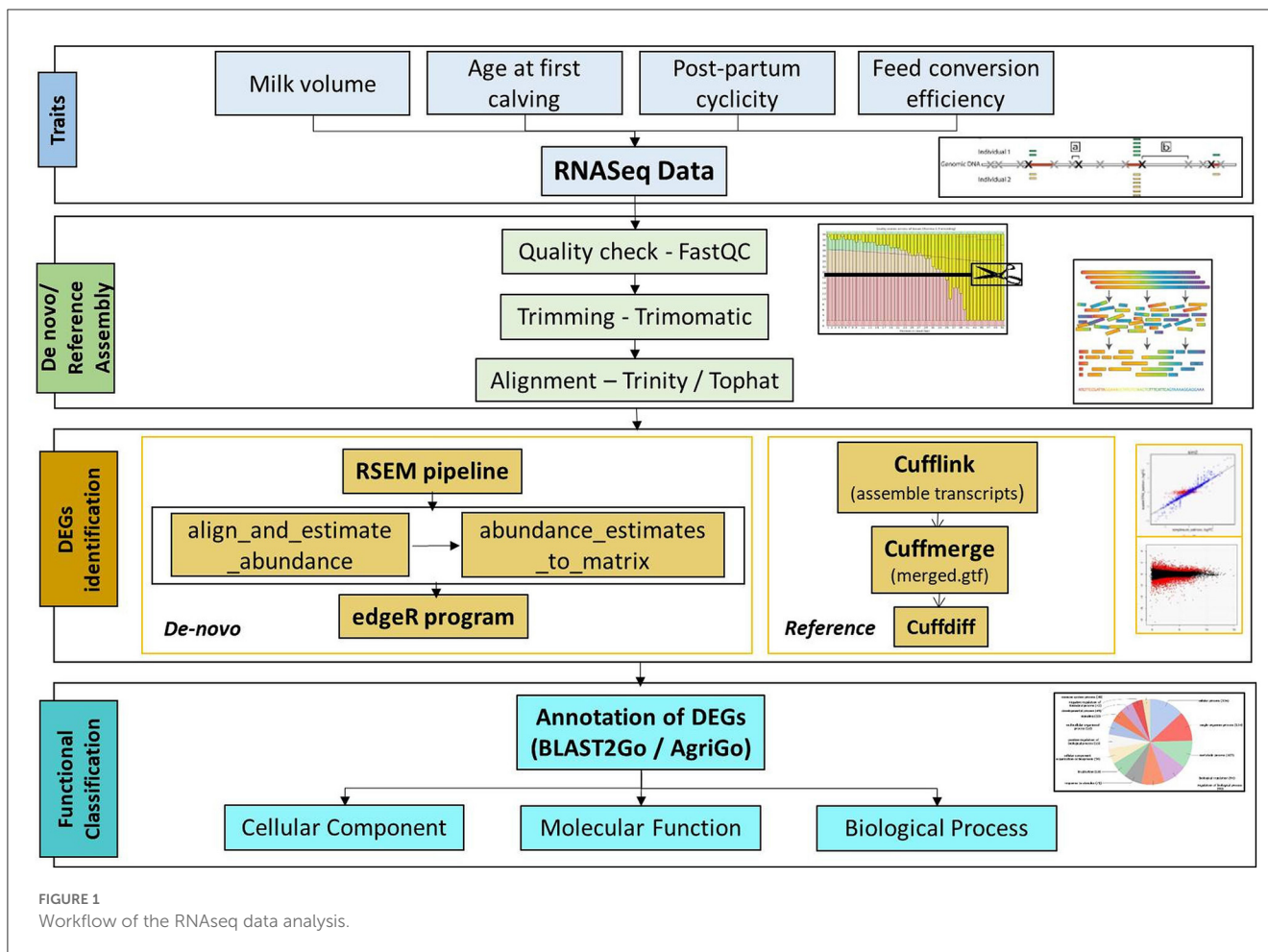


FIGURE 1 Workflow of the RNAseq data analysis.

TABLE 1 Number of differentially expressed genes in reference-free (RF) assembly and reference-based (RB) assembly.

Traits/genes	Reference-free (RF) assembly		Reference-based (RB) assembly	
	Up-regulated	Down-regulated	Up-regulated	Down-regulated
Milk volume (trait 1)	2,020	2,206	29	25
Age at first calving (trait 2)	726	525	37	62
Post-partum cyclicity (trait 3)	2,970	2,498	15	22
Feed conversion efficiency (trait 4)	1,474	1,782	45	44

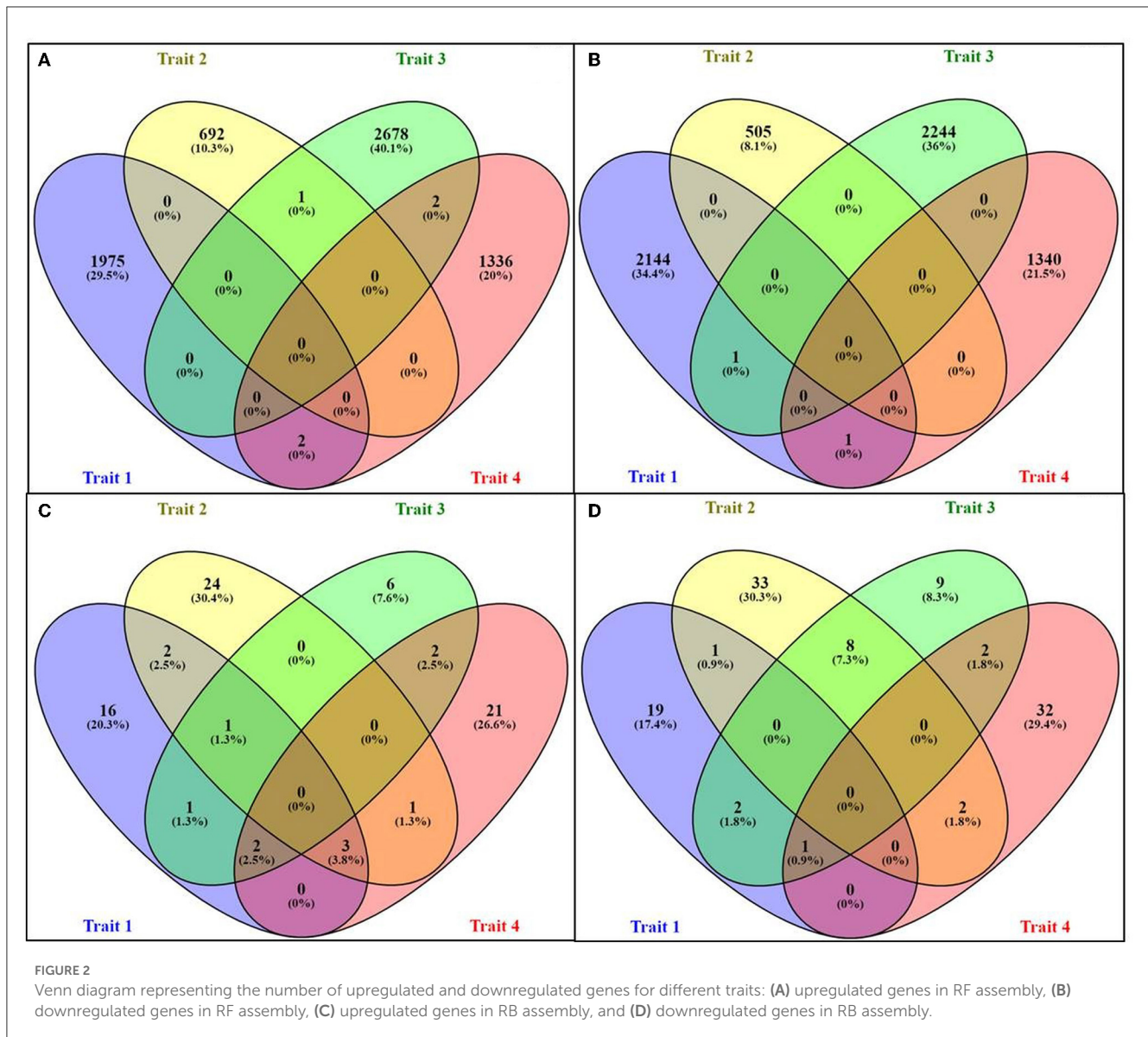
While making the comparison among the RF upregulated genes across all four traits, only two common DEGs were common between “milk volume” and “feed conversion efficiency,” and two DEGs were common between “post-partum cyclicity” and “feed conversion efficiency”. A single gene was common between “age at first calving” and “post-partum cyclicity” (Figure 2A). One DEG was found to be common among “milk volume,” “post-partum cyclicity,” and “feed conversion efficiency” traits (Figure 2B).

No common genes were identified among the four traits. A higher number of common DEGs were found in upregulated and downregulated categories through RB assembly. In total, 3.8% of all DEGs were upregulated in “milk volume,” “age at first calving,” and “feed conversion efficiency” traits, which were maximum. For traits, viz., “milk volume,” “age at first calving,” and “post-partum cyclicity,” only two common upregulated DEGs were identified

in the RB approach (Figure 2C). There were eight downregulated DEGs (7.3%) common in “age at first calving” and “post-partum cyclicity” (Figure 2D), fairly indicating the level of epigenetic regulation with respect to different traits either through DNA methylation and low expression of mRNA or demethylation.

Gene annotation

Gene ontology (GO) terms of identified genes (RF) were obtained using the BLAST2Go v 2.5 tool. The study revealed that one or more GO terms, viz., 30,290, 4,228, 43,142, and 17,097, were assigned to genes for milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency traits, respectively. GO



enrichment analysis classifies gene ontology terms into three broad categories, namely, cellular component, molecular function, and biological process (Figure 3). The binding function (GO: 0005488) was the most represented GO term in the molecular function category followed by cell and its part (GO: 0005623, GO: 0044464) and organelle and its part (GO: 0043226, GO:0044422) as cellular component terms for all the traits. Prominent GO terms that emerged from the RF assembly were similar to the classified terms that emerged from the RB assembly as 2,620 for milk volume, 440 for age at first calving, 3,644 for post-partum cyclicity, and 1,545 for feed conversion efficiency traits (Figure 4).

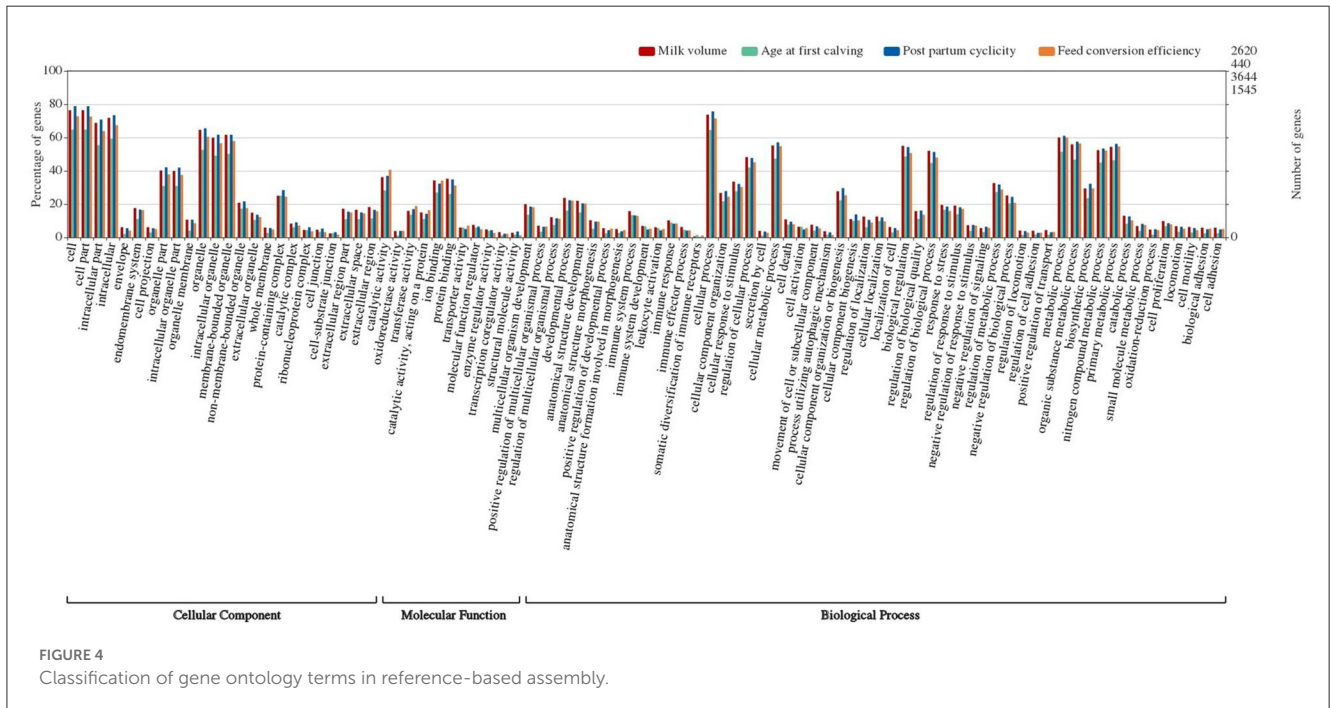
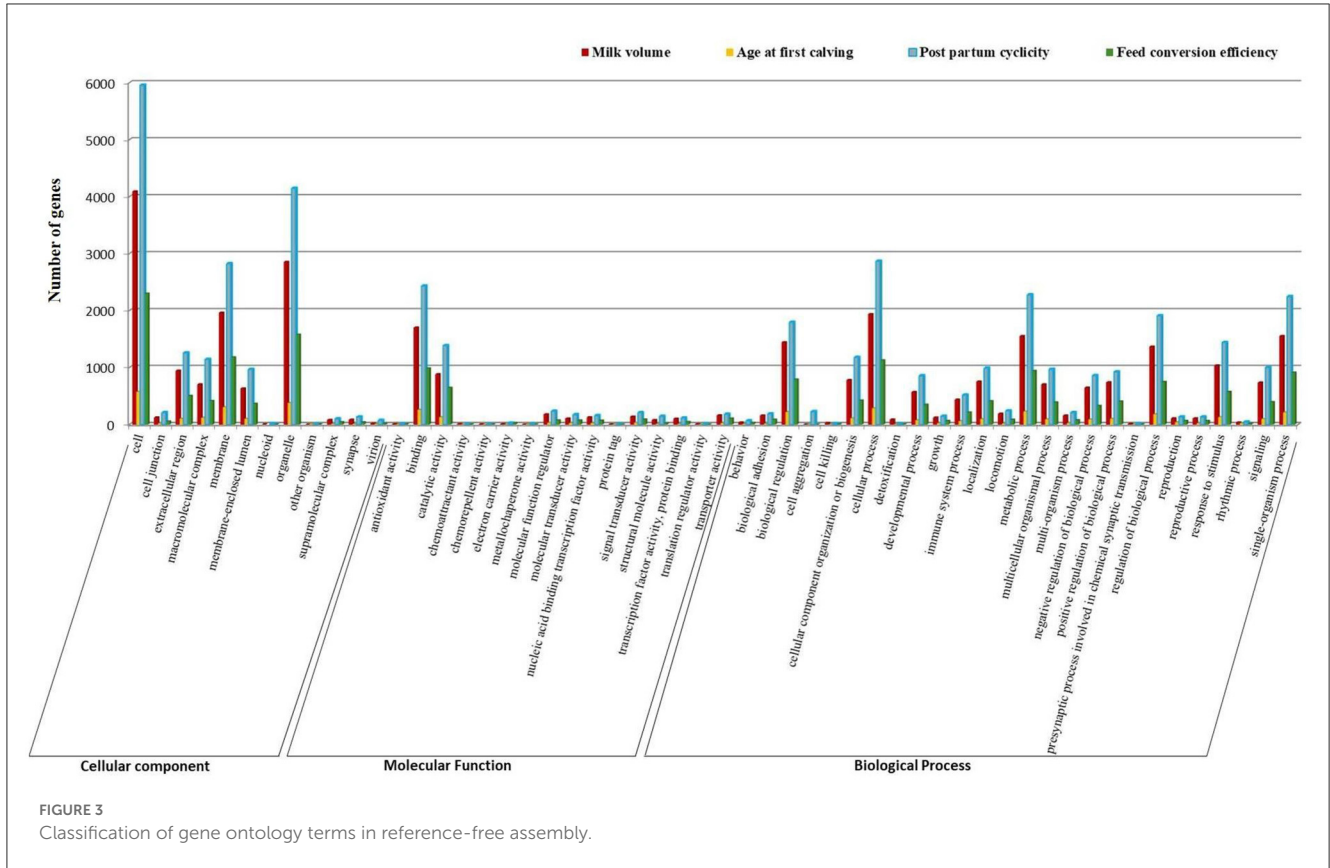
Variant distribution

Trait-wise distribution of three identified elements as DEGs, SSRs, and SNPs was mapped on the *Bos taurus* genome (Figure 5) and *B. bubalis* genome (Figure 6) using the CIRCOS tool. This

depicts the comparative view of the chromosomal-wise distribution of identified elements in various traits, viz., milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency. In the RB approach, there are 10,114 SSRs across all four traits, viz., milk volume (3,185), age at first calving (529), post-partum cyclicity (3,828), and feed conversion efficiency (2,572). Mononucleotide SSRs are 6,061 followed by 1,779 for dinucleotide SSRs, 1,417 for trinucleotide SSRs, 55 for tetranucleotide SSRs, 25 for pentanucleotide SSRs, and only two hexanucleotide SSRs. These identified SSRs were further filtered based on their chromosomal locations within the identified DEGs and SNPs. Figure 6 shows the mapping of these resulted in 161 SSRs.

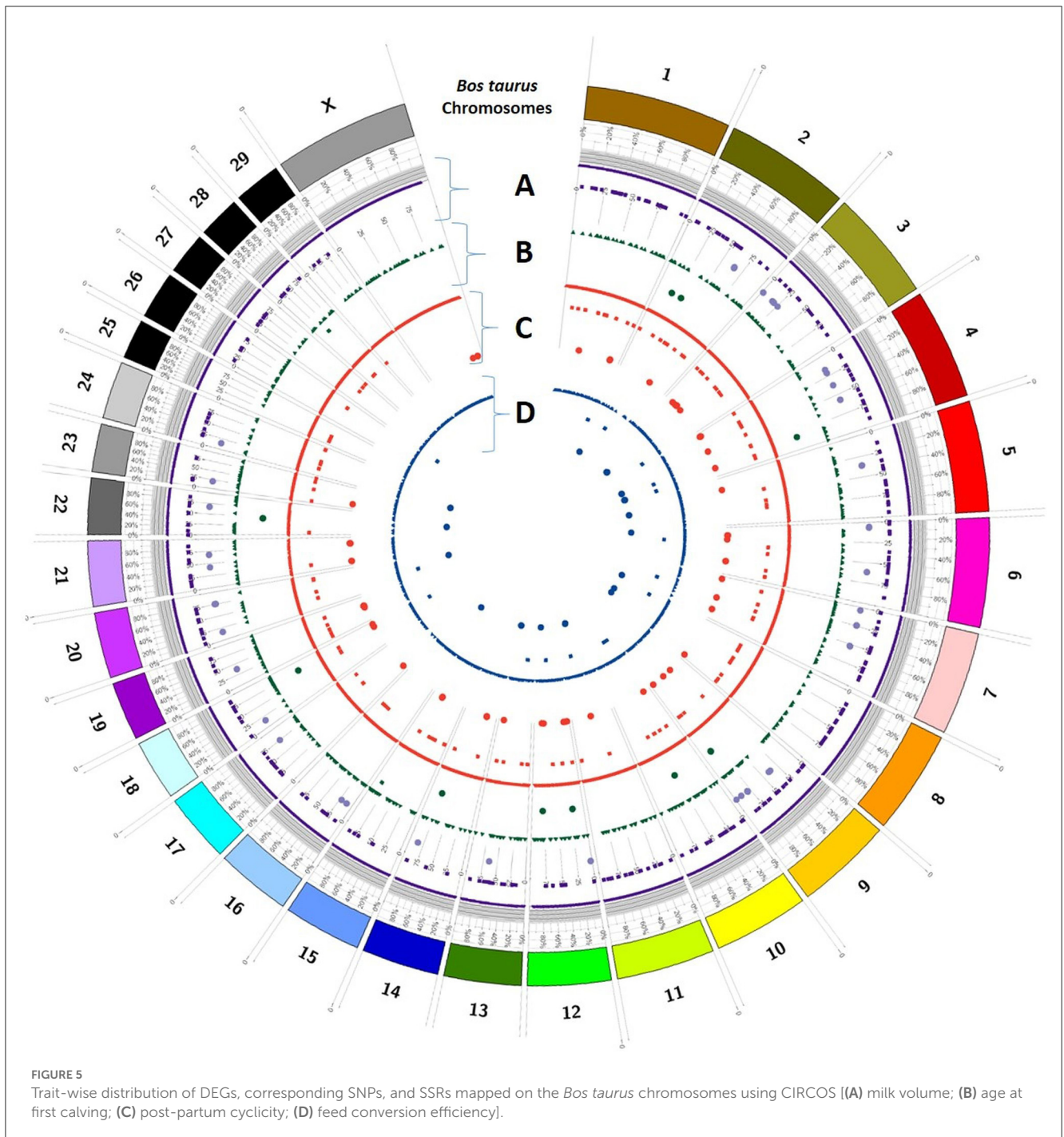
Discussion

The selection programs of domestic animals will be strengthened by a detailed grasp of the genetic variation underlying



complex phenotypes (36) as analyzed in this study. The current study compared the DEGs identified through RF and RB assembly approaches and shows their association with the buffalo traits considered in this study. An attempt was also made to show the chromosomal distribution of DEGs, SNPs, and SSRs in respect of

all four considered traits, viz., milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency of buffalo using the CIRCOS tool (Figures 5, 6). These maps depict that the maximum number of DEGs are found in post-partum cyclicity, followed by milk volume and feed conversion efficiency in the RF

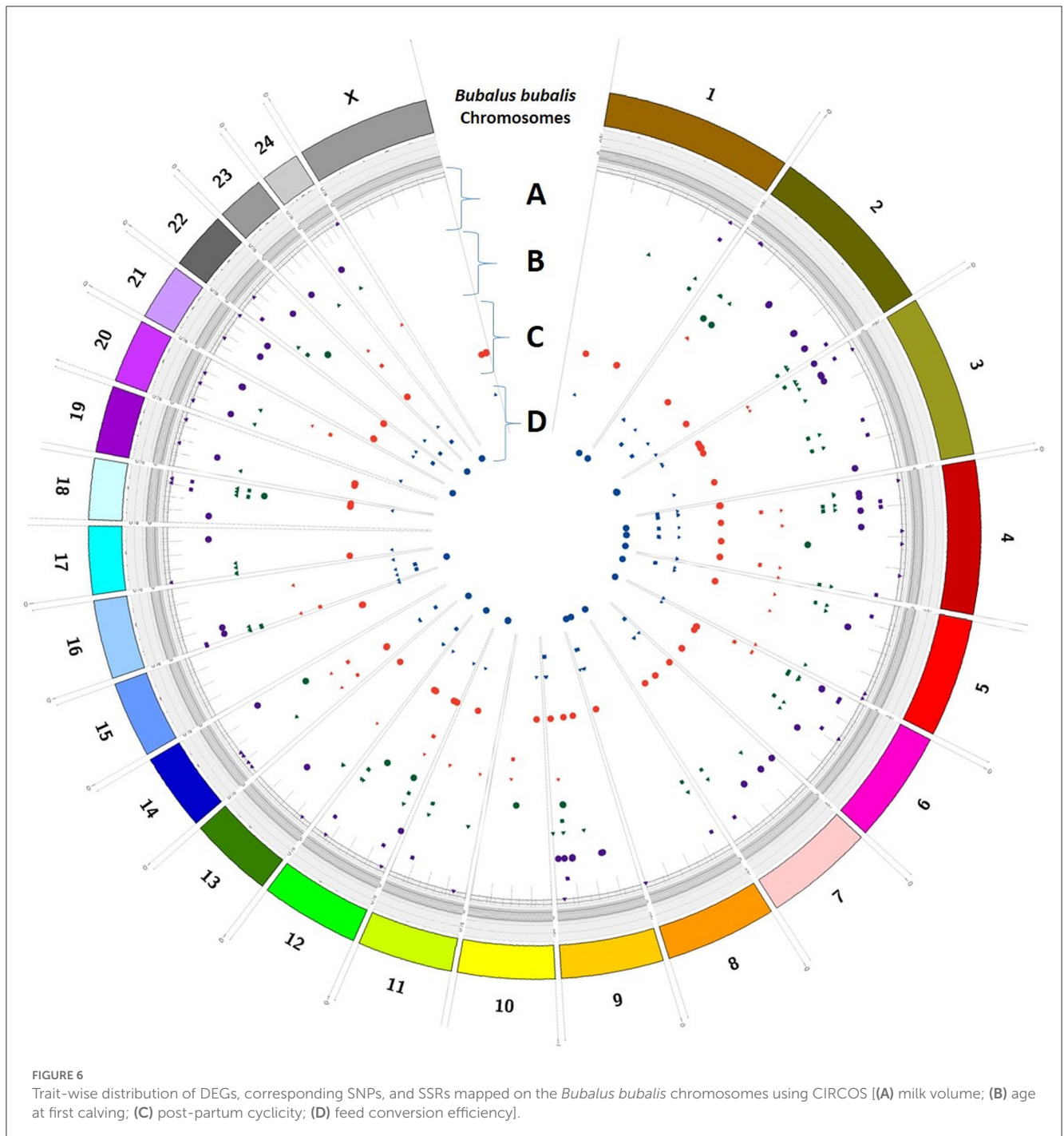


approach, and age at first calving has the maximum number of DEGs, followed by feed conversion efficiency and milk volume in the RB approach (Table 1).

Phosphatidylinositol 3,4,5 trisphosphate five phosphatase was identified as a common gene that was deregulated for functional diversity in two traits i.e., age at first calving and post-partum cyclicity, indicating the importance of the cell cycle progression in these traits and perhaps regulating the development of embryos (37, 38) (Supplementary Table S1). The FKBP4 gene is responsible for reproductive traits (39, 40). The presence of calcium channel, voltage-dependent, alpha-2/delta subunit 1 (CACNA2D1) gene

variant in respect of post-partum cyclicity and feed conversion efficiency traits in our study is crucial for its role in excitation-contraction coupling in neurons, glial cells, and muscle cells (41) (Supplementary Table S1).

The key genes, inositol 1,4,5-trisphosphate receptor (ITPR), branched-chain amino acid transaminase (BCAT), and other immunity-related genes, such as T-cell surface glycoprotein and AP-1 transcription factor, are identified as differential genes in our study that are mainly associated with all traits (42). Our study also shows that the identified candidate genes such as growth and hormone receptors, ribosomal proteins, sterol regulatory protein,



and GTPase are associated with milk volume as reported earlier by Surya et al. (43), Crisa et al. (44), Wickramasinghe et al. (45), Ma and Corl (46), and Lemay et al. (47), respectively, because of their involvement in histone modification, epidermal differentiation, cell adhesion, and cytoskeletal architecture (48). RNA binding FOX, transmembrane proteins, RNA binding proteins, cytosolic peptidase, and cell adhesion receptor genes were pertinent to age at first calving. These genes are primarily involved in cell proliferation, differentiation, adhesion, the mitotic cycle, DNA replication, RNA transcription, and apoptosis (49) (Supplementary Table S1).

As compared with RF, the RB assembly showed more common genes among the traits. There were a total of

28 genes that were common between different traits in RB compared to RF with only seven common genes (Figure 2). The genes, namely, sphingosine-1-phosphate (SIP), somatostatin (50, 51), BOLA class 1 histocompatibility antigen (52–54), interferon-stimulated/induced protein, and SRY-box transcription factor (55) have maximum homology when aligned with *B. bubalis* genome (UOA_WB_1, Accession GCA_003121395.1; Supplementary Table S2). The gene SIP is a bioactive lipid that acts through cell-surface receptors to promote cell signaling and causes a variety of cellular responses to help in developing immunity against diseases (50, 51). The major histocompatibility complex, such as BOLA class 1 histocompatibility antigen,

present in all mammalian species, is crucial for the immune system's development (54). This BOLA histocompatibility complex shows resistance to infectious diseases along with governing the milk volume (52, 53) (Supplementary Table S2). The genes governing disease resistance or susceptibility will positively affect milk productivity.

An immediate defense against viral infection identified in our study is provided by interferon-stimulated genes (ISGs) whose expression is induced by interferon signaling (56). These ISGs also act as potential biomarkers to avoid the occurrence of certain diseases in mammals and eliminate the incidence of adverse reactions to avoid the risk of further damage to the animals (57). This will help in increasing overall productivity that may be either due to an increase in milk volume or due to the application of feed conversion efficiency (Supplementary Table S2).

Post-partum cyclicity-related genes are myosin-related proteins, ribosylhydrolases, and cell adhesion receptors (58) that play a significant role in the growth and immunity of the bovine family. Furthermore, this study also identifies some vital genes (ligand-dependent nuclear receptors such as carboxylase and DLG2) related to feed conversion efficiency for regulating energy homeostasis, apoptosis, immune response, and cell growth in young heifers (59–61) (Supplementary Table S2).

In this study, the enriched GO terms revealed were related to milk production, reproduction, immunological response, and susceptibility/resistance to diseases. GO terms related to milk volume are associated with the biosynthesis of glycoproteins, fatty acids, glycerolipids, sterols, and other biological processes, such as oxidative stress, metabolic processes, transporter activity, divalent metal ion transport, calcium channel activity, acetyltransferase activity, and mRNA processing (Figures 3, 4). This confirms the importance of these processes in lactogenesis (62). The GO terms related to cellular component organization (GO:0071840), cell enzyme activity, or gene expression in response to stimulus (GO:0050896), regulation of biological functions (GO:0050789, GO:0048518, and GO:0048519), single-organism process (GO:0044699), and cell death (GO:0001906) govern physiological processes in animals. These terms are related to age at first calving and feed conversion efficiency (22). Important GO terms stimuli (GO:0050896), regulation of biological quality (GO:0065008), and molecular function (GO:0065009) are related to milk volume and feed conversion efficiency. Genes categorized under cell and cell-part localization (GO:0005623, GO:0030054, and GO:0044464) are prominent in all the considered four traits and are found to regulate different biological processes, viz., trans-membrane transport, regulation of signal transduction, milk production, and chemical transmission (49, 63) (Figures 3, 4). The gene condensing complex subunit 2 (Q3MHQ) encoded by GO term GO:0065007 is linked with the feed efficiency trait. The gene Q3MHQ regulates cell division and improves growth development in an animal by converting interphase chromatin into mitotic chromosomal condensation and is interestingly linked to metabolic pathways involved in feed conversion efficiency (64, 65).

Conclusion

In this study, an attempt has been made to compare reference-free and reference-based approaches to identify and

annotate differentially expressed genes in *B. bubalis* for four important traits, viz., milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency. Reference-free (RF) *de novo* transcriptome assembly approach is commonly used due to the non-availability of a complete reference genome with the high-quality genetic information of particular species. In this study, the RF approach identified 7,190 upregulated genes and 7,011 downregulated genes, whereas the RB approach identified 126 and 153 genes, respectively. The number of gene ontology terms associated with identified DEGs for the traits under consideration—milk volume, age at first calving, post-partum cyclicity, and feed conversion efficiency—were 30,290, 4,228, 43,142, and 17,097 for the RF approach, compared to 2,620, 440, 3,644, and 1,545 terms for the RB approach. The identified genes and GO terms will establish a sound base for biological postulates which will further improve future animal breeding programs to enhance animal productivity.

Data availability statement

The original transcriptome data presented in the study are publicly available. This data can be found here: NCBI Bioproject, accession number PRJNA934134.

Ethics statement

The animal study was reviewed and approved by Institute Animal Ethics Committee (IAEC) at ICAR-Central Institute for Research on Buffaloes, Hisar, India.

Author contributions

PS, AJ, AB, and IS conducted the experiments. JB and SY did data analysis. DM, JB, KC, and HA conceptualized the study and manuscript. All authors reviewed and approved the final manuscript.

Funding

This research was funded by the project Network Project for Agricultural Bioinformatics and Computational Biology under CABin Scheme, ICAR (F.no. Agril. Edn.4-1/2013-A&P).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Warriach H, McGill D, Bush R, Wynn P, Chohan K. A review of recent developments in buffalo reproduction—a review. *Asian-Australas J Anim Sci.* (2015) 28:451. doi: 10.5713/ajas.14.0259
- Niranjan S, Goyal S, Dubey P, Kumari N, Mishra S, Mukesh et al. Genetic diversity analysis of buffalo fatty acid synthase (FASN) gene and its differential expression among bovines. *Gene.* (2016) 575:506–12. doi: 10.1016/j.gene.2015.09.020
- Jerome A, Bhati J, Mishra DC, Chaturvedi KK, Rao AR, Rai A, et al. MicroRNA-related markers associated with corpus luteum tropism in buffalo (*Bubalus bubalis*). *Genomics.* (2020) 112:108–13. doi: 10.1016/j.ygeno.2019.01.018
- Mishra DC, Sikka P, Yadav S, Bhati J, Paul SS, Jerome A, et al. Identification and characterization of trait-specific SNPs using ddRAD sequencing in water buffalo. *Genomics.* (2020) 112:3571–8. doi: 10.1016/j.ygeno.2020.04.012
- El-Salam A, Mohamed H, El-Shibiny S. A comprehensive review on the composition and properties of buffalo milk. *Dairy Sci Tech.* (2011) 91:663–99. doi: 10.1007/s13594-011-0029-2
- Guzman JLG, Lázaro SF, do Nascimento AV, de Abreu Santos DJ, Cardoso DF, Becker Scalez DC, et al. Genome-wide association study applied to type traits related to milk yield in water buffaloes (*Bubalus bubalis*). *J Dairy Sci.* (2020) 103:1642–50. doi: 10.3168/jds.2019-16499
- Eastham NT, Coates A, Cripps P, Richardson H, Smith R, Oikonomou G. Associations between age at first calving and subsequent lactation performance in UK Holstein and Holstein-Friesian dairy cows. *PLoS ONE.* (2018) 13:e0197764. doi: 10.1371/journal.pone.0197764
- Japheth KP, Kumaresan A, Patbandha TK, Baithalu RK, Selvan AS, Nag P, et al. Supplementation of a combination of herbs improves immunity, uterine cleansing and facilitate early resumption of ovarian cyclicity: a study on post-partum dairy buffaloes. *J Ethnopharm.* (2021) 272:113931. doi: 10.1016/j.jep.2021.113931
- Sethi M, Shah N, Mohanty TK, Bhakat M, Dewry RK, Yadav DK, et al. The induction of cyclicity in postpartum anestrous buffaloes: a review. *J Exp Zool.* (2021) 24:989–97.
- Sikka P, Nath A, Paul SS, Andonissamy J, Mishra DC, Rao AR, et al. Inferring relationship of blood metabolic changes and average daily gain with feed conversion efficiency in murrah heifers: machine learning approach. *Front Vet Sci.* (2020) 7:518. doi: 10.3389/fvets.2020.00518
- Haldar A, Prakash BS. Effects of growth hormone-releasing factor on growth hormone response, growth and feed conversion efficiency in buffalo heifers (*Bubalus bubalis*). *Vet J.* (2007) 174:384–9. doi: 10.1016/j.tvjl.2006.10.003
- Reddy RM. Feed-milk conversion efficiency of dairy animals and methane emissions across different landholding groups. *International J Sci Res.* (2020) 10:1242–8. doi: 10.21275/SR21527174803
- Porcu E, Sadler MC, Lepik K, Auwerx C, Wood AR, Weihs A, et al. Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat Commun.* (2021) 12:5647. doi: 10.1038/s41467-021-25805-y
- Mishra DC, Smita S, Singh I, Devi MN, Kumar S, Farooqi MS, et al. Prediction of novel putative miRNAs and their targets in buffalo. *Indian J Anim Sci.* (2017) 87:59–63. doi: 10.56093/ijans.v87i1.66861
- Pokharel K, Peippo J, Honkatukia M, Seppälä A, Rautiainen J, Ghanem N, et al. Integrated ovarian mRNA and miRNA transcriptome profiling characterizes the genetic basis of prolificacy traits in sheep (*Ovis aries*). *BMC Genomics.* (2018) 19:104. doi: 10.1186/s12864-017-4400-4
- Liu Y, Wu X, Xie J, Wang W, Xin J, Kong F, et al. Identification of transcriptome differences in goat ovaries at the follicular phase and the luteal phase using an RNA-Seq method. *Theriogenology.* (2020) 158:239–49. doi: 10.1016/j.theriogenology.2020.06.045
- Gong X, Zheng M, Zhang J, Ye Y, Duan M, Chamba Y, et al. Transcriptomics-based study of differentially expressed genes related to fat deposition in Tibetan and yorkshire pigs. *Front Vet Sci.* (2022) 9:919904. doi: 10.3389/fvets.2022.919904
- El Nahas SM, Mossallam AAA. Acui identifies water buffalo CSN3 genotypes by RFLP analysis. *J Genet.* (2014) 93:e94–6. doi: 10.1007/s12041-014-0427-3
- Paraguison RC, Faylon MP, Flores EB, Cruz LC. Improved RAPD-PCR for discriminating breeds of water buffalo. *Biochem Genet.* (2012) 50:579–84. doi: 10.1007/s10528-012-9502-8
- Mishra DC, Yadav S, Sikka P, Jerome A, Paul SS, Rao AR, et al. SNPRBb: economically important trait specific SNP resources of buffalo (*Bubalus bubalis*). *Conserv Genet Resour.* (2021) 13:283–9. doi: 10.1007/s12686-021-01210-x
- Gargani M, Pariset L, Soysal MI, Ozkan E, Valentini A. Genetic variation and relationships among Turkish water buffalo populations. *Animal Genet.* (2010) 41:93–6. doi: 10.1111/j.1365-2052.2009.01954.x
- Pathak RK, Kim JM. Vetinformatics from functional genomics to drug discovery: insights into decoding complex molecular mechanisms of livestock systems in veterinary science. *Front Vet Sci.* (2022) 9:1008728. doi: 10.3389/fvets.2022.1008728
- Deng T, Liang A, Liang S, Ma X, Lu X, Duan A, et al. Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. *Front Genet.* (2019) 10:36. doi: 10.3389/fgene.2019.00036
- Liu S, Ye T, Li Z, Li J, Jamil AM, Zhou Y, et al. Identifying hub genes for heat tolerance in water buffalo (*Bubalus bubalis*) using transcriptome data. *Front Genet.* (2019) 10:209. doi: 10.3389/fgene.2019.00209
- Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE.* (2012) 7:e30619. doi: 10.1371/journal.pone.0030619
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* (2014) 30:2114–20. doi: 10.1093/bioinformatics/btu170
- Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data.* (2010). Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed March 30, 2023).
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech.* (2011) 29:644–52. doi: 10.1038/nbt.1883
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* (2013) 8:1494–512. doi: 10.1038/nprot.2013.084
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* (2013) 14:R36. doi: 10.1186/gb-2013-14-4-r36
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* (2012) 9:357–9. doi: 10.1038/nmeth.1923
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. Available online at: <http://www.R-project.org/> (accessed March 30, 2023).
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* (2011) 12:323. doi: 10.1186/1471-2105-12-323
- Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* (2008) 36:3420–35. doi: 10.1093/nar/gkn176
- Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. *Proc Royal Soc B.* (2016) 283:20160569. doi: 10.1098/rspb.2016.0569
- Qiao F, Law HC, Krieger KL, Clement EJ, Xiao Y, Buckley SM, et al. Ctdp1 deficiency leads to early embryonic lethality in mice and defects in cell cycle progression in MEFs. *Biol Open.* (2021) 10:bio057232. doi: 10.1242/bio.057232
- Dubon MAC, Pedrosa VB, Feitosa FLB, Costa RB, de Camargo GMF, Silva MR, et al. Identification of novel candidate genes for age at first calving in nellore cows using

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2023.1160486/full#supplementary-material>

- a SNP chip specifically developed for *Bos taurus indicus* cattle. *Theriogenology*. (2021) 173:156–62. doi: 10.1016/j.theriogenology.2021.08.011
39. Hulsege I, Woelders H, Smits M, Schokker D, Jiang L, Sorensen P. Prioritization of candidate genes for cattle reproductive traits, based on protein-protein interactions, gene expression, and text-mining. *Physiol Genomics*. (2013) 45:400–6. doi: 10.1152/physiolgenomics.00172.2012
40. Wathes DC, Cheng Z, Chowdhury W, Fenwick MA, Fitzpatrick R, Morris DG, et al. Negative energy balance alters global gene expression and immune responses in the uterus of postpartum dairy cows. *Physiol Genomics*. (2009) 39:1–13. doi: 10.1152/physiolgenomics.00064.2009
41. Magotra A, Gupta ID, Verma A, Alex R, Arya A, Vineeth MR, et al. Characterization and validation of point mutation in Exon 19 of Calcium channel, voltage-dependent, Alpha-2/Delta subunit 1(CACNA2D1) gene and its relationship with mastitis traits in Sahiwal. *Indian J Anim Res*. (2018) 52:61–4. doi: 10.18805/ijar.10990
42. Zheng X, Ju Z, Wang J, Li Q, Huang J, Zhang A, et al. Single nucleotide polymorphisms, haplotypes and combined genotypes of *LAP3* gene in bovine and their association with milk production traits. *Mol Biol Rep*. (2011) 38:4053–61. doi: 10.1007/s11033-010-0524-1
43. Surya T, Vineeth MR, Sivalingam J, Tantia MS, Dixit SP, Niranjana SK, et al. Genomewide identification and annotation of SNPs in *Bubalus bubalis*. *Genomics*. (2019) 111:1695–8. doi: 10.1016/j.ygeno.2018.11.021
44. Crisa A, Ferre F, Chillemi G, Muioli B. RNA-Sequencing for profiling goat milk transcriptome in colostrum and mature milk. *BMC Vet Res*. (2016) 12:264. doi: 10.1186/s12917-016-0881-7
45. Wickramasinghe S, Rincon G, Islas-Trejo A, Medrano JF. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics*. (2012) 13:45. doi: 10.1186/1471-2164-13-45
46. Ma L, Corl BA. Transcriptional regulation of lipid synthesis in bovine mammary epithelial cells by sterol regulatory element binding protein-1. *J Dairy Sci*. (2012) 95:3743–55. doi: 10.3168/jds.2011-5083
47. Lemay DG, Lynn DJ, Martin WF, Neville MC, Casey TM, Rincon G, et al. The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biol*. (2009) 10:R43. doi: 10.1186/gb-2009-10-4-r43
48. Du C, Deng T, Zhou Y, Ye T, Zhou Z, Zhang S, et al. Systematic analyses for candidate genes of milk production traits in water buffalo (*Bubalus bubalis*). *Anim Genet*. (2019) 50:207–16. doi: 10.1111/age.12739
49. Fan H, Wu Y, Qi X, Zhang J, Li J, Gao X, et al. Genome-wide detection of selective signatures in Simmental cattle. *J Appl Genet*. (2014) 55:343–51. doi: 10.1007/s13353-014-0200-6
50. Pyne NJ, Pyne S. Sphingosine 1-phosphate receptor 1 signaling in mammalian cells. *Molecules*. (2017) 22:344. doi: 10.3390/molecules22030344
51. Spiegel S, Milstien S. Sphingosine-1-phosphate: an enigmatic signalling lipid. *Nat Rev Mol Cell Biol*. (2003) 4:397–407. doi: 10.1038/nrm1103
52. Rupp R, Hernandez A, Mallard BA. Association of bovine leukocyte antigen (BoLA) DRB3.2 with immune response, mastitis, and production and type traits in Canadian holsteins. *J Dairy Sci*. (2007) 90:1029–38. doi: 10.3168/jds.S0022-0302(07)71589-8
53. Pashmi M, Qanbari S, Ghorashi SA, Sharifi AR, Simianer H. Analysis of relationship between bovine lymphocyte antigen DRB3.2 alleles, somatic cell count and milk traits in Iranian holstein p. *J Anim Breed Genet*. (2009) 126:296–303. doi: 10.1111/j.1439-0388.2008.00783.x
54. Behl JD, Verma NK, Tyagi N, Mishra P, Behl R, Joshi BK. The major histocompatibility complex in bovines: a review. *ISRN Vet Sci*. (2012) 28:872710. doi: 10.5402/2012/872710
55. Lefebvre V, Dumitriu B, Penzo-Méndez A, Han Y, Pallavi B. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int J Biochem Cell Biol*. (2007) 39:2195–214. doi: 10.1016/j.biocel.2007.05.019
56. Shaw AE, Hughes J, Gu Q, Behdenna A, Singer JB, Dennis T, et al. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol*. (2017) 15:e2004086. doi: 10.1371/journal.pbio.2004086
57. Zheng Z, Wang L, Pan J. Interferon-stimulated gene 20-kDa protein (ISG20) in infection and disease: review and outlook. *Intractable Rare Dis Res*. (2017) 6:35–40. doi: 10.5582/irdr.2017.01004
58. Abo-Ismael MK, Brito LF, Miller SP, Sargolzaei M, Grossi DA, Moore SS, et al. Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. *Genet Sel Evol*. (2017) 49:82. doi: 10.1186/s12711-017-0356-8
59. Sasago N, Abe T, Sakuma H, Kojima T, Uemoto Y. Genome-wide association study for carcass traits, fatty acid composition, chemical composition, sugar, and the effects of related candidate genes in Japanese black cattle. *Anim Sci J*. (2017) 88:33–44. doi: 10.1111/asj.12595
60. Sun C, Southard C, Witonsky DB, Kittler R, Di Rienzo A. Allele-specific down-regulation of *RPTOR* expression induced by retinoids contributes to climate adaptations. *PLoS Genet*. (2010) 6:e1001178. doi: 10.1371/journal.pgen.1001178
61. Setoguchi K, Watanabe T, Weikard R, Albrecht E, Kuhn C, Kinoshita A, et al. The SNPc.1326T>G in the non-SMC condensin I complex, subunit G(NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle. *Animal Genet*. (2011) 42:650–5. doi: 10.1111/j.1365-2052.2011.02196.x
62. Lemay DG, Neville MC, Rudolph MC, Pollard KS, German JB. Gene regulatory networks in lactation: identification of global principles using bioinformatics. *BMC Syst Biol*. (2007) 1:56. doi: 10.1186/1752-0509-1-56
63. Nayeri S, Stothard P. Tissues, metabolic pathways and genes of key importance in lactating dairy cattle. *Springer Sci Rev*. (2016) 4:49–77. doi: 10.1007/s40362-016-0040-3
64. Daetwyler H, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. (2014) 46:858–65. doi: 10.1038/ng.3034
65. Widmann P, Reverter A, Weikard R, Suhre K, Hammon HM, Albrecht E, et al. Systems biology analysis merging phenotype, metabolomic and genomic data identifies non-SMC condensin I complex, subunit G (NCAPG) and cellular maintenance processes as major contributors to genetic variability in bovine feed efficiency. *PLoS ONE*. (2015) 10:e0124574. doi: 10.1371/journal.pone.0124574