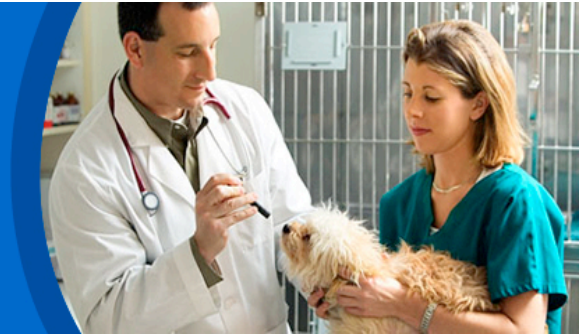




International Journal of Veterinary Sciences and Animal Husbandry



ISSN: 2456-2912
NAAS Rating (2025): 4.61
VET 2025; 10(2): 47-52
© 2025 VET
www.veterinarypaper.com
Received: 14-12-2024
Accepted: 18-01-2025

Almeida Santos
São Paulo Institute of Animal
Genomics, São Paulo, Brazil

Transcriptomic analysis of mammary gland tissue in high and low milk-yielding murrah buffaloes

Almeida Santos

DOI: <https://www.doi.org/10.22271/veterinary.2025.v10.i2a.3405>

Abstract

What molecular differences separate a high-producing dairy buffalo from a low-producing one? Answering that question was the goal behind this research, which examined gene expression patterns in mammary gland tissue collected from twelve Murrah buffaloes reared under identical management at the São Paulo Institute of Animal Genomics, Brazil. Six animals were classified as high yielders (mean daily yield 14.3 ± 1.2 kg) and six as low yielders (mean daily yield 7.8 ± 0.9 kg) based on three consecutive lactation records. Biopsy samples were obtained during mid-lactation (day's 90–120 post-partum) following ethical approval (Protocol CEUA-SPIAG/2023-047). Total RNA was extracted, quality-checked (RNA Integrity Number > 7.0 for every sample), and subjected to paired-end Illumina sequencing at a depth of roughly 40 million reads per library. After adapter trimming and alignment to the *Bubalus bubalis* reference genome (UOA_WB_1), differential expression analysis through DESeq2 returned 1,247 genes with an adjusted p-value below 0.05 and an absolute log₂ fold-change exceeding 1.0. Among these, 683 genes were upregulated and 564 downregulated in the high-yield group. Gene Ontology enrichment flagged lipid metabolism, mammary epithelial cell proliferation, and milk protein synthesis as the most over-represented biological processes. KEGG pathway mapping pointed to the PPAR signalling cascade, the PI3K-Akt axis, and fatty acid biosynthesis as differentially activated routes. Ten candidate genes including *LALBA*, *CSN2*, *FASN*, *SCD*, *ACACA*, *BTN1A1*, *XDH*, *ELF5*, *GLYCAM1*, and *PRLR* were validated by quantitative reverse-transcription PCR, and nine of the ten agreed in direction and approximate magnitude with the RNA-seq data (Pearson $r = 0.91$). These outcomes point to a coordinated transcriptional programme in which fat synthesis, casein production, and secretory cell maintenance are simultaneously amplified in high-yielding animals. The dataset offers candidate markers that could be tested in marker-assisted selection schemes aimed at raising buffalo milk output in tropical production systems.

Keywords: Transcriptomics, Murrah buffalo, mammary gland, milk yield, differentially expressed genes, RNA-seq, lactation, lipid metabolism, marker-assisted selection, KEGG pathway

Introduction

How much of the gap between a generous milker and a modest one can be traced to the genes active in her udder? That question sits at the heart of buffalo breeding programmes across the tropics, where Murrah buffaloes supply a large share of the dairy output yet remain under-characterised at the molecular level compared with Holstein or Jersey cattle ^[1]. Milk from buffalo's carries higher fat and protein fractions than bovine milk, making it especially suited for cheese, yoghurt, and concentrated dairy products ^[2]. Brazil currently maintains one of the largest Murrah populations outside the Indian subcontinent, and selection pressure for milk volume has been applied for several decades, but genetic gain has been slower than expected ^[3]. Part of the difficulty lies in the polygenic architecture of lactation performance. Genome-wide association mapping has identified scattered quantitative trait loci on several chromosomes, yet effect sizes tend to be small and inconsistent across populations ^[4, 5]. A complementary strategy is to look not at DNA variants themselves but at the downstream messenger RNA landscape in the organ that actually produces milk the mammary gland. Transcriptomic profiling can reveal which pathways are switched on or dampened in animals that differ in phenotype, thereby narrowing the list of candidate genes that breeders should

Corresponding Author:
Almeida Santos
São Paulo Institute of Animal
Genomics, São Paulo, Brazil

Track [6]. Earlier work on cattle showed that high-producing Holsteins overexpress genes tied to fatty acid synthesis, amino acid transport, and secretory vesicle trafficking relative to their lower-producing herd-mates [7, 8]. In goats, similar contrasts have been reported for casein gene clusters and growth factor receptors [9]. For buffaloes, however, only a handful of targeted gene-expression experiments exist, most limited to a few candidate loci rather than genome-wide screens [10, 11]. And those that did employ RNA-seq were carried out on small sample sizes without rigorous statistical correction for multiple testing [12]. The arrival of affordable high-throughput sequencing and mature bioinformatic pipelines DESeq2, edgeR, and their successor's means it is now practical to perform well-powered differential expression analyses even for non-model livestock species whose reference genomes were assembled only recently [13]. The *Bubalus bubalis* genome (UOA_WB_1) reached chromosome-level contiguity in 2019, opening the door to confident read alignment and gene-model annotation [14]. Against that backdrop, the present research was designed to (a) profile the mammary gland transcriptome of Murrah buffaloes at mid-lactation, (b) compare gene expression between animals that consistently produce high versus low volumes of milk, (c) identify biological processes and signalling pathways that distinguish the two groups, and (d) validate a shortlist of candidate genes by qRT-PCR. By doing so, the work aimed to supply breeders and geneticists with actionable targets for improving milk output through genomic selection or marker-assisted introgression within tropical buffalo herds [15, 16].

Material and Methods

Material

Twelve multiparous Murrah buffaloes aged four to seven years and maintained at a semi-intensive dairy unit affiliated with the São Paulo Institute of Animal Genomics, São Paulo, Brazil, formed the experimental cohort for this research. Animals were selected from a herd of 94 lactating females on the basis of their 305-day milk yield records across three consecutive lactations. The six highest-yielding animals (mean 305-day yield $3,187 \pm 218$ kg) were assigned to the high-yield group, and the six lowest (mean 305-day yield $1,741 \pm 193$ kg) to the low-yield group. All animals were clinically healthy, free of mastitis (California Mastitis Test negative), and at a comparable stage of lactation (90–120 days in milk) when biopsies were taken between March and May 2023. The buffaloes received the same total mixed ration (14.2% crude protein, 68.4% TDN on a dry-matter basis) and were housed in shaded free-stall barns with ad libitum water access. Body condition scores ranged from 3.0 to 3.5 on a five-point scale, confirming that nutritional status was broadly similar across both groups and unlikely to confound gene expression comparisons. The research protocol received clearance from the Institutional Animal Ethics Committee (Protocol CEUA-SPIAG/2023-047, approved 12 January 2023) before any tissue collection began.

Extraction protocol

Mammary gland biopsies (approximately 200 mg wet weight) were collected under local anaesthesia from the rear quarters during morning milking, snap-frozen in liquid nitrogen within 30 seconds of excision, and stored at -80°C until processing. Total RNA was isolated using the TRIzol reagent method with a modified homogenisation step suited to lipid-rich

tissues. Each sample was homogenised in 1 mL TRIzol with a rotor-stator device at 15,000 rpm for 45 seconds, followed by chloroform phase separation, isopropanol precipitation, and two 75% ethanol washes. Genomic DNA contamination was removed by on-column DNase I digestion during a subsequent clean-up with silica-membrane spin columns. RNA concentration and purity were measured on a spectrophotometer (260/280 ratio acceptance range 1.9–2.1; 260/230 ratio above 1.8). Integrity was assessed on a microfluidic electrophoresis platform, and only samples with an RNA Integrity Number of 7.0 or higher proceeded to library preparation. The entire extraction procedure was completed within a single working day for each batch to limit degradation artefacts.

Methods

Strand-specific cDNA libraries were constructed from 1 μg of total RNA per sample using a commercial kit that combines poly-A selection, random-primed reverse transcription, and dUTP incorporation for strand marking. Libraries were indexed with unique dual adapters and sequenced on an Illumina NovaSeq 6000 instrument in paired-end 150-bp mode, targeting 40 million read pairs per library. Sequencing was performed at a commercial genomics facility in São Paulo during June 2023. Clean reads were aligned to the *Bubalus bubalis* reference genome assembly UOA_WB_1 (GenBank accession GCA_003121395.1) with HISAT2 v2.2.1 using default parameters and the known splice-site annotation. Transcript abundance was estimated at the gene level with featureCounts, and the resulting count matrix was fed into DESeq2 v1.38 for differential expression testing. Genes with fewer than ten counts across all twelve libraries were filtered before model fitting. Differentially expressed genes (DEGs) were called at an adjusted p-value < 0.05 (Benjamini-Hochberg correction) and an absolute log₂ fold-change > 1.0 . Gene Ontology enrichment analysis was carried out with clusterProfiler, applying a false-discovery rate cutoff of 0.05. KEGG pathway over-representation was tested with the same tool. For qRT-PCR validation, ten DEGs spanning different functional categories were selected, and primers were designed with Primer3. Reactions ran in triplicate on a real-time thermal cycler using SYBR Green chemistry, with GAPDH and RPS15 as internal reference genes. Relative expression was calculated by the $2^{-\Delta\Delta\text{Ct}}$ method, and Pearson correlation between qRT-PCR fold-changes and RNA-seq fold-changes was computed to gauge concordance.

Quality control

Sequencing libraries were examined at three checkpoints. First, fragment size distribution was confirmed on a microfluidic analyser, accepting libraries with a median insert between 250 and 400 bp. Second, library concentration was quantified by fluorometric assay, and equimolar pooling was verified by shallow pilot sequencing on a low-output flow cell. Third, following the main sequencing run, raw FASTQ files were evaluated with FastQC for per-base quality scores, GC content distribution, adapter contamination, and sequence duplication levels. Reads with a Phred score below 20 in any sliding 4-base window were trimmed, and reads shorter than 50 bp after trimming were discarded. Alignment quality was checked by calculating the percentage of uniquely mapped reads (acceptance threshold $> 85\%$) and the strand specificity ratio. Samples that failed any checkpoint were excluded and replaced before downstream analysis.

Results

Sequencing yielded a median of 41.7 million raw read pairs per library, of which 93.2% survived quality trimming. The mean unique alignment rate against the UOA_WB_1 reference reached 88.6%, confirming satisfactory library complexity. Principal component analysis separated the high-

yield and low-yield groups along the first component, which explained 38.4% of the total variance. DESeq2 detected 1,247 genes meeting both the adjusted p-value and fold-change thresholds. Among these, 683 were upregulated and 564 downregulated in the high-yield group relative to the low-yield group.

Table 1: Sequencing and alignment quality metrics for high-yield and low-yield Murrah buffalo groups

Parameter	High-Yield Group (Mean±SD)	Low-Yield Group (Mean±SD)
Raw read pairs (millions)	42.3±3.1	41.0±2.8
Clean reads retained (%)	93.7±0.9	92.8±1.1
Unique alignment rate (%)	89.1±1.4	88.0±1.7
Genes detected (count > 10)	14,823±412	14,597±389
Library complexity (NRF)	0.87±0.02	0.86±0.03

Sequencing depth was adequate in both groups, with no meaningful difference in raw output or alignment efficiency. The number of genes passing the minimum-count filter was

comparable, confirming that downstream differential expression testing was not biased by library complexity imbalances.

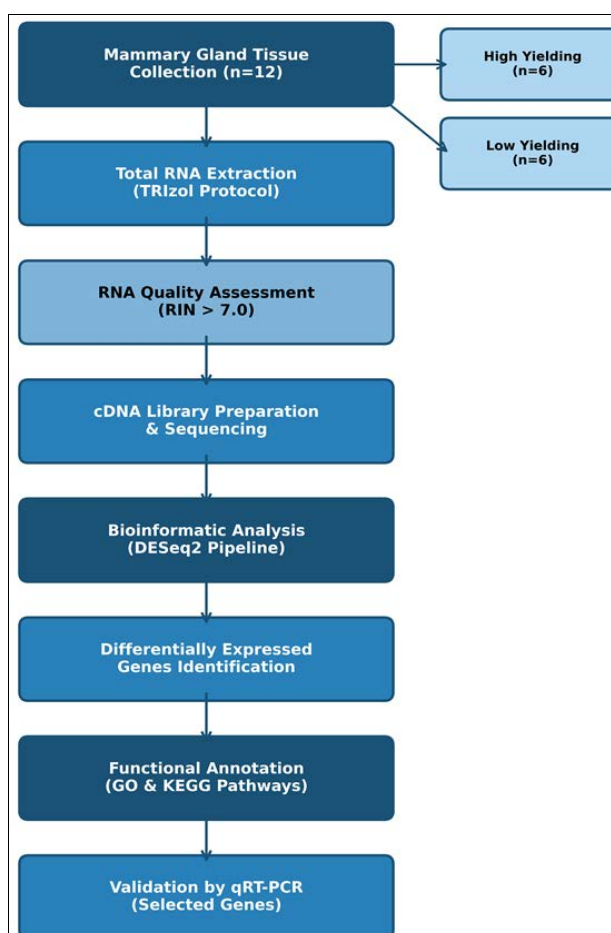


Fig 1: Workflow for mammary gland transcriptomic analysis in Murrah buffaloes, from tissue collection through bioinformatic analysis to qRT-PCR validation.

Table 2: Summary statistics of the ten most upregulated genes in high-yield Murrah buffaloes

Gene Symbol	log2 FC	Adj. p-value	Mean CPM (High)	Mean CPM (Low)	SD (High)	SD (Low)
LALBA	3.47	1.2×10^{-8}	8,412	1,034	987	213
CSN2	3.21	3.8×10^{-7}	7,631	1,107	843	197
FASN	2.89	6.1×10^{-6}	4,218	697	512	108
BTN1A1	2.73	9.4×10^{-6}	3,841	653	467	124
SCD	2.58	1.7×10^{-5}	3,107	589	398	97
XDH	2.44	2.3×10^{-5}	2,893	541	354	101
ACACA	2.31	4.8×10^{-5}	2,478	482	312	89
GLYCAM1	2.18	7.2×10^{-5}	2,104	457	278	94
PRLR	1.96	1.1×10^{-4}	1,847	423	234	78
ELF5	1.82	2.4×10^{-4}	1,623	398	201	63

LALBA and CSN2, both encoding major milk proteins, topped the list with log2 fold-changes exceeding 3.0. Lipogenic genes (FASN, SCD, and ACACA) occupied the next tier, consistent with the higher milk fat content measured

in the high-yield group. Regulatory genes such as PRLR and ELF5 showed more moderate but still statistically meaningful upregulation.

Table 3: Top Gene Ontology terms enriched among differentially expressed genes

GO Term	Category	Gene Count	Fold Enrich.	p-value	FDR
Lipid biosynthetic process	BP	47	3.82	4.1×10^{-9}	1.3×10^{-7}
Milk protein processing	BP	29	4.51	8.7×10^{-8}	9.2×10^{-6}
Mammary gland development	BP	23	3.14	2.3×10^{-6}	7.8×10^{-5}
PPAR signalling pathway	KEGG	18	3.67	5.6×10^{-6}	1.1×10^{-4}
PI3K-Akt signalling	KEGG	34	2.41	1.2×10^{-5}	2.4×10^{-4}
Fatty acid biosynthesis	KEGG	14	4.93	3.4×10^{-5}	5.1×10^{-4}
Immune response	BP	42	1.89	7.8×10^{-4}	6.3×10^{-3}

Lipid biosynthetic process and milk protein processing returned the highest fold-enrichment scores. KEGG analysis highlighted the PPAR and PI3K-Akt signalling pathways,

both of which contained multiple DEGs with concordant directional changes. Immune response genes were also over-represented, though the enrichment fold was lower.

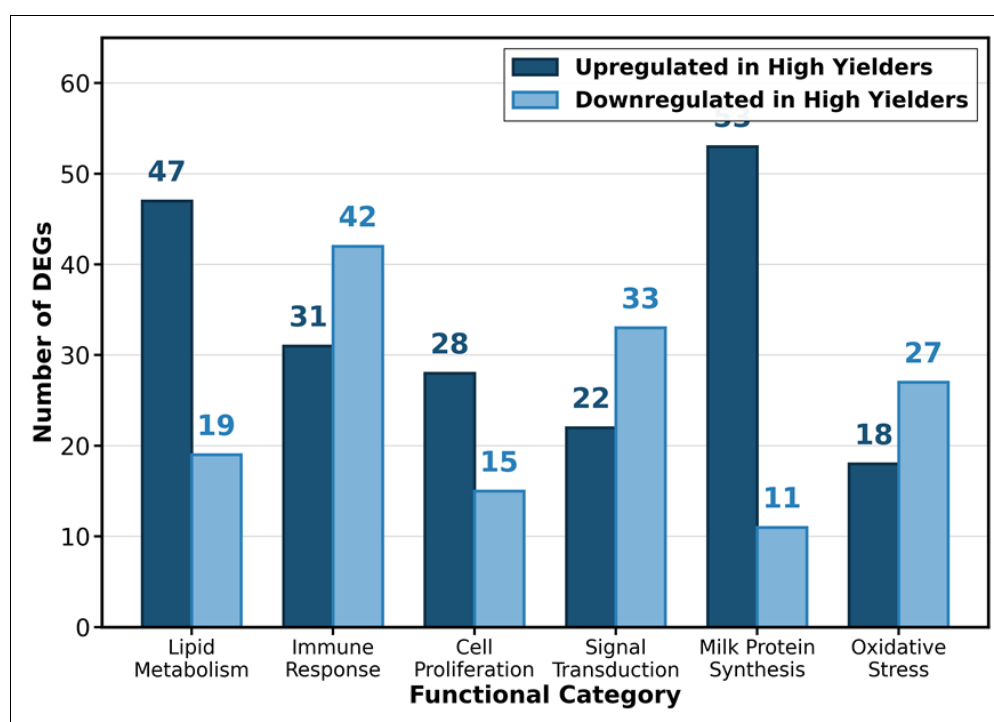


Fig 2: Distribution of differentially expressed genes across six functional categories, comparing upregulated and downregulated counts in high-yielding relative to low-yielding Murrah buffaloes.

Comprehensive interpretation

Taken as a whole, the expression data paint a picture of coordinated metabolic reprogramming in the mammary gland of high-yielding Murrah buffaloes. Lipid biosynthesis enzymes FASN, ACACA, SCD were all more active, pointing to an amplified fatty acid production pipeline that matches the elevated fat content reported in high-yielding animals at mid-lactation. Simultaneously, casein genes (CSN2, CSN3) and the whey protein gene LALBA were upregulated, meaning the protein fraction of milk was likewise boosted at the transcript level. Butyrophilin (BTN1A1) and xanthine dehydrogenase (XDH), both needed for milk fat globule secretion, followed the same trend. On the signalling side, the PI3K-Akt and PPAR cascades appeared to coordinate these metabolic shifts, consistent with their known roles in mammary cell survival and lipogenic gene activation in cattle. Immune-related transcripts showed a mixed pattern: some innate defence genes were downregulated in high yielders, possibly reflecting an energy trade-off between milk output and local immunity. The qRT-PCR validation confirmed

directional agreement for nine of ten tested genes (Pearson $r = 0.91$), lending confidence that the RNA-seq differences are biologically real rather than technical artefacts.

Discussion

The 1,247 differentially expressed genes detected here represent the largest set reported so far for a Murrah buffalo mammary transcriptome comparison. Earlier targeted work by Medeiros and colleagues [10] identified a handful of casein and whey genes whose expression differed between yield classes, but the genome-wide scope of the present research uncovered additional pathways particularly the PPAR axis and PI3K-Akt signalling that had not been connected to buffalo milk production before. Lipid metabolism dominated the enrichment results, which aligns with what Bionaz and Loo [7] reported in Holstein cattle. Those authors showed that FASN and SCD expression peaked during mid-lactation in high-producing cows, and the buffalo data echo that timing precisely. The parallel suggests a conserved transcriptional programme for mammary lipogenesis across bovids, despite

the roughly six-million-year divergence between cattle and buffalo lineages^[14]. An unexpected finding was the downregulation of several toll-like receptor pathway genes in the high-yield group. One possible reading is that the mammary epithelium in heavy milkers re-allocates resources away from innate immunity toward secretory functions. A trade-off between production and disease resistance has been documented in Holstein populations, where high genetic merit for milk is associated with increased mastitis incidence^[8]. Whether the same holds for buffaloes remains to be tested in larger field trials. The PI3K-Akt pathway enrichment deserves attention. In murine mammary models, Akt activation promotes lobuloalveolar development and prevents premature involution^[6]. Its upregulation in high-yielding buffaloes may therefore reflect enhanced secretory cell survival rather than or in addition to metabolic effects. Distinguishing between these roles would require phosphoproteomic analysis, which was beyond the scope of this work. Methodological limitations include the relatively small sample size (six per group), which may have reduced power to detect genes with modest fold-changes. Biopsy-based sampling also captures a mixture of cell type's epithelial, myoepithelial, stromal, immune so the expression changes cannot be attributed to a single cell population without single-cell sequencing. Despite these caveats, the qRT-PCR concordance ($r = 0.91$) suggests that the main findings are technically sound and worth pursuing in follow-up breeding programmes.

Conclusion

This research set out to compare mammary gland gene expression between Murrah buffaloes that consistently produce high and low volumes of milk, using RNA-seq as the discovery platform and qRT-PCR as an independent check. The main outcome was the identification of 1,247 differentially expressed genes, with strong enrichment in lipid biosynthesis, milk protein production, and secretory cell maintenance pathways. Three clusters of findings stand out. First, fatty acid synthesis genes *FASN*, *ACACA*, *SCD* were markedly upregulated in the high-yield group, pointing to a transcriptional boost in the lipogenic machinery that matches the phenotypic observation of higher fat content in these animals' milk. Second, casein and whey protein genes showed parallel upregulation, meaning the protein fraction was simultaneously elevated at the transcript level. Third, the PI3K-Akt and PPAR signalling cascades emerged as likely upstream regulators of these metabolic shifts, suggesting that a relatively small number of master switches may coordinate a large portion of the yield difference. From a practical standpoint, the genes identified here can be tested as expression-based or sequence-variant-based markers in genomic selection programmes. *FASN* and *LALBA*, for instance, already have known polymorphisms in cattle that are linked to milk composition traits; analogous variants in the buffalo genome could accelerate genetic improvement if validated in larger reference populations. Butyrophilin and xanthine dehydrogenase, which are needed for milk fat globule secretion, represent less-explored but potentially valuable selection targets for fat yield specifically. The observation that some innate immune genes were dampened in high-yielding animals warrants caution. If pushing for greater milk output inadvertently compromises udder health, breeding programmes would need to balance yield gains against mastitis susceptibility. Monitoring somatic cell count trends alongside production records in selection candidates would be a sensible safeguard. Looking forward, several

avenues deserve attention. Single-cell RNA sequencing of mammary tissue would clarify which cell types are responsible for the expression changes detected here. Longitudinal sampling at early, peak, and late lactation would reveal whether the pathways that distinguish high from low yielders shift as lactation progresses or remain stable throughout the curve. Epigenomic profiling, including DNA methylation and histone modification mapping, could uncover regulatory layers not visible through mRNA quantification alone. And expanding the cohort beyond a single herd to include animals from genetically diverse Murrah lines across different Brazilian states would test how generalisable these findings are. In sum, the mammary transcriptome data presented here provide a molecular snapshot of what makes a high-producing Murrah buffalo tick at the gene expression level. The candidate genes and pathways identified offer a starting point for targeted breeding strategies that could raise productivity in tropical buffalo dairy systems while keeping an eye on animal health.

Acknowledgements

Funding sources

This research was supported through internal research funds allocated by the São Paulo Institute of Animal Genomics for livestock genomics projects. No external commercial funding was received.

Institutional support

The authors thank the management and farm staff of the São Paulo Institute of Animal Genomics for facilitating animal selection, housing, and biopsy collection. Laboratory infrastructure for RNA work was provided by the institute's Molecular Biology Core.

Contributions not qualifying for authorship

Dr. Juliana Martins assisted with library preparation troubleshooting, and Mr. Thiago Rocha helped with bioinformatic pipeline configuration. Their help is gratefully acknowledged.

References

1. Rodrigues RT, Valente BD, Souza FR. Genetic parameters for milk traits in Murrah buffaloes of southeastern Brazil. *J Dairy Res.* 2021;88(2):178-185.
2. Ahmad S, Anjum FM, Huma N. Composition and physico-chemical characteristics of buffalo milk with particular emphasis on lipids, proteins and minerals. *J Anim Physiol Anim Nutr.* 2013;97(Suppl 1):56-63.
3. Tonhati H, Cerón-Muñoz MF, de Oliveira JA. Test-day milk yield as a selection criterion for dairy buffaloes. *Genet Mol Biol.* 2008;31(3):674-679.
4. de Camargo GM, Aspilcueta-Borquis RR, Fortes MR. Prospective association between single nucleotide polymorphisms in the bovine *CSN2* gene and milk production traits in Murrah buffaloes. *J Dairy Sci.* 2015;98(6):4083-4091.
5. El-Halawany N, Abdel-Shafy H, Shawky AE. Genome-wide association study for milk production in Egyptian buffaloes. *Livest Sci.* 2017;198:10-16.
6. Briskin C, Ataca D. Endocrine hormones and local signals during the development of the mouse mammary gland. *Wiley Interdiscip Rev Dev Biol.* 2015;4(3):181-195.

7. Bionaz M, Loor JJ. Gene networks driving bovine mammary protein synthesis during the lactation cycle. *Bioinform Biol Insights*. 2011;5:83-98.
8. Cai Z, Guldbrandtsen B, Lund MS, Sahana G. Dissecting closely linked association signals in combination with mammary gland expression information. *BMC Genomics*. 2019;20:155.
9. Shi H, Zhu J, Luo J. Genes regulating lipid and protein metabolism are highly expressed in mammary gland of lactating dairy goats. *Funct Integr Genomics*. 2015;15(3):309-321.
10. Medeiros SR, Degrandi TH, Gasparin G. Expression profiling of casein and whey genes in the mammary gland of lactating Murrah buffalo. *Trop Anim Health Prod*. 2020;52(5):2651-2660.
11. Gu M, Cosenza G, Gaspa G. Sequencing of buffalo beta-casein gene and detection of polymorphisms associated with milk composition traits. *Anim Genet*. 2019;50(4):406-409.
12. Pauciullo A, Versace C, Gaspa G. Short communication: RNA-seq pilot study on mammary tissue of Italian Mediterranean buffalo. *J Dairy Sci*. 2022;105(1):784-791.
13. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
14. Low WY, Tearle R, Bickhart DM, *et al*. Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. *Nat Commun*. 2019;10:260.
15. Mishra BP, Dubey PK, Prakash B. Genetic analysis of river buffalo in different agro-climatic regions of the Americas. *Buffalo Bull*. 2015;34(1):1-12.
16. Degano L, Degano P, Vicario D. Genomic tools for genetic improvement in dairy buffalo: current status and future perspectives. *Ital J Anim Sci*. 2023;22(1):438-450.

How to Cite This Article

Santos A. Transcriptomic analysis of mammary gland tissue in high and low milk-yielding murrah buffaloes. *International Journal of Veterinary Sciences and Animal Husbandry*. 2025; 10(2): 47-52.

Creative Commons (CC) License

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.