RESEARCH ARTICLE
Open Access

Journal of Veterinary and Animal Sciences
ISSN (Print): 0971-0701, (Online): 2582-0605
https://doi.org/10.51966/jvas.2022.53.4.663-668

Mining of diverse short non-coding RNAs from transcriptome of milk somatic cells of Murrah buffalo

Pooja Chhabra[1], Brij Mohan Goel[2], Raja K. N.[3], Anil Kumar Mishra[4], Sonika Ahlawat[3], Mandeep Kaur[5], Ashish Kumar[6] and Reena Arora[4*]

ICAR-National Bureau of Animal Genetic Resources, Karnal-132 001, Haryana
Baba Mastnath University, Asthal Bohar, Rohtak-124021, Haryana, India

Citation: Pooja, C., Brij, M. G., Raja, K.N., Anil, K.M., Sonika, A., Mandeep, K., Ashish, K. and Reena, A. 2022. Mining of diverse short non-coding RNAs from transcriptome of milk somatic cells of Murrah buffalo. *J. Vet. Anim. Sci.* **53**(4): 663-668
DOI: https://doi.org/10.51966/jvas.2022.53.4.663-668

Received: 23.06.2022 *Accepted: 29.09.2022* *Published: 31.12.2022*

Abstract

The non-coding RNAs (ncRNA) are known to regulate expression of genes at the transcription, translation and processing levels. The present study was conducted to identify diverse short ncRNAs from milk somatic cells of lactating Murrah buffaloes. Elucidating the molecular drivers of lactation in dairy animals will help understand the process of lactation, eventually leading to improvement in milk production and quality. In order to discover the ncRNA, the transcriptome data of 12 samples of somatic cells from buffalo milk were analyzed. A web based pipeline, exceRpt was used to perform the analysis. The most abundant short ncRNA molecules discovered in buffalo milk were the miRNAs, followed by snRNAs. Least number of rRNAs was discovered in the investigated samples. The total number of rRNAs, tRNAs, snRNAs, snoRNAs and miRNAs were 12, 23, 72, 51 and 229 respectively, in the entire dataset. On matching with miRBase v22.1, a total of 1724, 897, 211 and 4 miRNAs were observed to be common to human, bovine, caprine and ovine genomes. The results provide information on the bioavailability of short ncRNAs in buffalo milk somatic cells, most of which are largely uncharacterized. The generated information is a step towards developing a database for ncRNAs in buffalo species.

Keywords: *Buffalo, data mining, lactation, RNA sequencing, ncRNA*

Buffalo species, with a population of 109 million, is the major milch species which contributes to 51% of the total milk production in India (BAHS, 2019). Since the demand for milk has been predicted to increase in the coming years, there is a lot of scope for genetic improvement of dairy animals to meet this demand. Milk secretion or lactation is an intricate process that is influenced

1. *PhD Scholar, Baba Mastnath University, Asthal Bohar, Rohtak*
2. *Assistant Professor, Baba Mastnath University, Asthal Bohar, Rohtak*
3. *Senior Scientist*
4. *Principal Scientist*
5. *Young Professional*
6. *PhD Scholar, Kurukshetra University, Kurukshetra*

**Corresponding author: rejagati@gmail.com; reena.arora@icar.gov.in*

by nutrition, climate, genes, hormones as well as many other factors (Kumar, 2017; Neelima *et al.,* 2021; Kalyan *et al*., 2022). Several genes and pathways work in synchrony for functioning of the mammary gland. The regulation of gene expression is a complex mechanism influenced by many biological factors and molecules. The advent of RNA sequencing technology has provided unprecedented opportunities in exploring the RNAome of any tissue to unravel the regulatory mechanisms of gene expression. The non-coding RNAs (ncRNA) are now known to govern expression of genes at the transcription, translation and processing levels (Zhang *et al*., 2019). The ncRNA are a family of RNA molecules that remain untranslated but function in regulating gene expression, biological pathways and processes. These untranslated RNA molecules include microRNAs (miRNA), small nucleolar RNAs (snoRNA), small nuclear RNAs (snRNA), Ribosomal RNAs (rRNA) and Transfer RNAs (tRNA). Recent studies have confirmed the role of miRNA in development of mammary gland in mice (Avril-Sassen*et al.,* 2009). The discovery of lncRNAs in mammary glands suggests their role in development and lactation (Yu *et al.,* 2017). The availability of whole genome and transcriptome sequences has facilitated the discovery and cataloging of this novel yet important class of biomolecules.

Elucidating the molecular drivers of lactation in dairy animals will help understand the process of lactation, eventually leading to improvement in milk production and quality (Yu *et al.,* 2017). Although some information is available on the lncRNAs in buffalo (Li *et al.,* 2020), there is still dearth of information on other short ncRNA in buffaloes. Therefore, this study was taken up with the objective to identify short ncRNAs from milk somatic cells of lactating Murrah buffaloes. Information on the regulatory molecules involved in lactation will provide a better insight into the complex biological processes involved in milk secretion.

Materials and methods

RNA sequencing data

The RNA sequencing data used in this study was generated in a previous study (Arora *et al*., 2019). The dataset was downloaded from GenBank with accession number GGRC00000000.1, under BioProject PRJNA453843 that contained 12 fastq files of buffalo milk somatic cells. The data used in this study was obtained from milk samples of 6.5 and 7 years old lactating buffaloes in the third parity (Arora *et al*., 2019).

Quality filtering of RNA sequencing data

The FastQC tool (Andrews, 2010) was used to check the quality of the raw reads. On the basis of the report generated by FastQC tool, the low quality reads were removed. A quality threshold of Q30 (average phred score ≥ 30) was applied and reads having quality score greater than threshold value were used for further analysis. The cutadapt tool (Martin, 2011) was used to clip the adapter sequences. The output of cutadapt was again validated by FastQC to ensure the high quality of the processed reads for further analysis.

Mining of short ncRNA

The exceRpt pipeline (Rozowsky*et al.,* 2019) was used to mine the diverse RNAs from the processed reads. STAR v2.4.2a (Dobin *et al*., 2013) and Bowtie v2.2.6 (Langmead and Salzberg, 2012) were used for alignment of the processed reads against the human reference genome. For annotation, different databases incorporated in exceRpt were used, rRNAs were extracted from 45S (Aguilera *et al.,* 2017), 5S (Szymanski *et al.,* 2002), miRNAs from miRBase v22.1 (Kozomara and Griffiths-Jones, 2011) and tRNAs from gtRNAdb (Chan and Lowe, 2016). The annotations were derived from GENCODE version 24 (Harrow *et al*., 2012). The alignment and annotation were done concurrently by the exceRpt pipeline (Rozowsky*et al.,* 2019) with default parameters.

Results and discussion

Emerging techniques in sequencing DNA/RNA have led to the discovery of hitherto unknown nucleic acid sequences whose functions are yet to be determined. Non-coding RNAs are a category of such sequences that have gained prominence due to their association in various biological activities. These molecules are still being identified and

characterized in different species. In order to understand the mechanism behind a biological process, information of all the participating molecules is necessary. Several studies have reported the role of ncRNAs in human and cattle milk secretion (Do and Ibeagha-Awemu, 2017; Tingo*et al*., 2021), however, negligible information is available for buffalo milk.

The RNA sequencing data for milk somatic cells from 12 lactating buffaloes was filtered for quality. The unprocessed reads/ transcripts for all the samples had a length varying from a single nucleotide to 101 nucleotides. Since ncRNAs have 15 or more nucleotides, all transcripts with less than 15 nucleotides were eliminated. Each sample yielded at least 47,000,000 raw reads after initial filtration. Further trimming and filtering of the raw reads provided not less than 45,000,000 reads for each sample. A total of 32,365,636 reads that were of poor quality, having adapter content or of short length were trimmed (Table 1). The processed reads were aligned to the human genome reference assembly (GRCh38.p13- GCA_000001405.28). Since the annotation level of the human genome is much better than the buffalo genome, it was selected over buffalo for alignment of the reads. The percentage of alignment (mapping percent) of the filtered and processed reads to the reference genome is displayed in Fig. 1. All the investigated samples aligned to the reference genome with an average mapping percentage of 97.49. The high quality, mapped reads were further annotated against different databases to identify the ncRNAs present in buffalo milk somatic cells. The diverse short ncRNAs (rRNAs, tRNAs, miRNAs, snRNAs and snoRNAs) identified in the dataset are given in Fig. 2.

The most abundant short ncRNA molecules discovered in buffalo milk were the miRNAs, followed by snRNAs. Least number of rRNAs were discovered in the investigated samples. These results may also depend on the priority order given in the pipeline. The total number of rRNAs, tRNAs, snRNAs, snoRNAs and miRNAs after removing the redundancy were 12, 23, 72, 51 and 229 respectively, in the entire dataset. Most of the rRNA is removed during sample preparation, as it is present in greater amount and may prevent the detection of other lesser expressed RNA molecules. As a consequence, lowest number of rRNAs could be detected in this study. There is still dearth of information on other ncRNA and their role in mammary gland.

Micro RNAs are the most widely studied class of short ncRNAs and have been detected in amniotic fluid, saliva, urine, blood as well as milk (Zhou *et al*., 2012). Several miRNAs have been identified in the mammary gland tissues of cattle, sheep and goat (Do and Ibeagha-Awemu, 2017). Previous studies have reported 230 known miRNAs in colostrum and

Table 1. Summary of raw, trimmed and processed reads across RNA sequencing data of milk somatic cells of buffalo.

Sample	Raw reads	Trimmed Reads	Processed reads
B1	98060582	2246640	95813942
B2	73845150	2503096	71342054
B3	63252478	1509736	61742742
B4	47773688	1838958	45934730
B5	88926132	3143020	85783112
B6	86709746	3775854	82933892
B7	80978704	3353628	77625076
B8	105728098	3700346	102027752
B9	108754608	4052926	104701682
B10	48109684	1223252	46886432
B11	114278262	3302408	110975854
B12	69643178	1715772	67927406

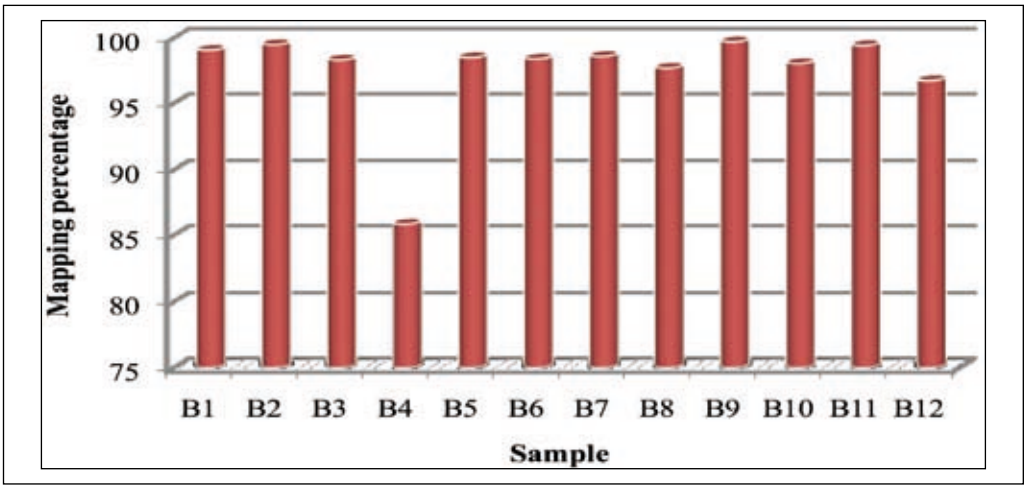

Fig. 1. Alignment of processed reads of buffalo milk somatic cells with human reference genome.

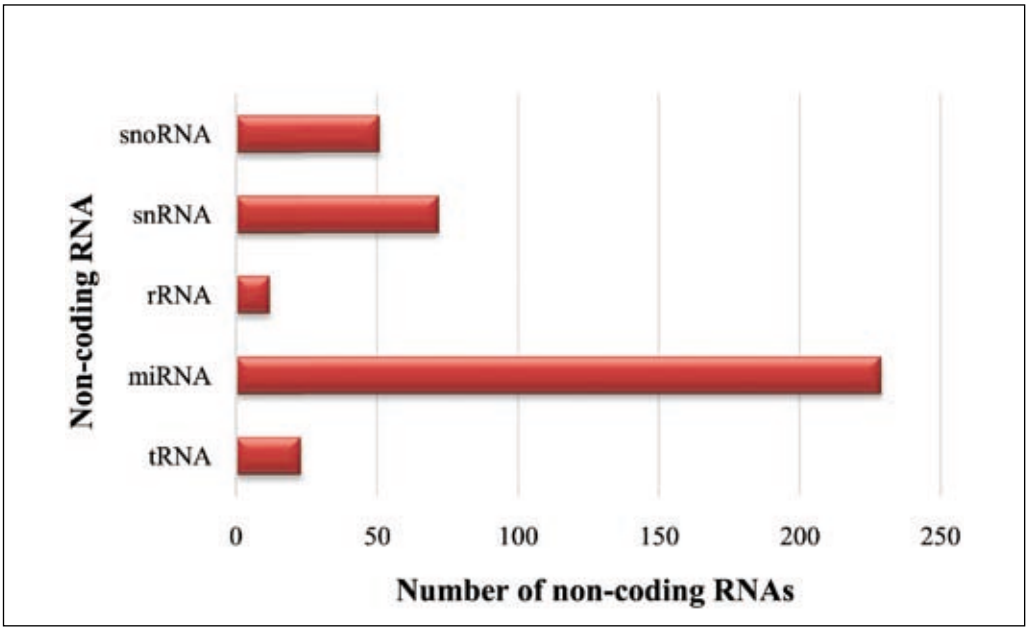

Fig. 2. Diverse short non-coding RNAs mined across all the samples of milk somatic cells of buffalo.

213 in mature milk (Gigli and Maizon, 2013). In order to search miRNA from our dataset, that were common to other species, the processed reads were matched with themiRbase v22.1, a database for all published miRNA (Kozomara and Griffiths-Jones, 2011). A total of 1724, 897, 211 and 4 miRNAs were observed to be common to human, bovine, caprine and ovine genomes. These results can be attributed to the fact that the human and bovine genomes are better annotated. Since no miRNA database is as yet available for the *Bubalus bubalis* genome, the present study adds to this information on this important milch species.

Conclusion

This present study is a step towards identification and cataloguing of short ncRNAs in buffalo milk. The results provide information on the bioavailability of short ncRNAs in buffalo milk somatic cells, most of which are largely uncharacterized. Further efforts are required to delineate their role in milk secretion. Many

molecular processes related to mammary gland development, health and milk secretion are affected by ncRNA in the genome (Shore *et al*., 2012) and therefore, they form promising candidates for improvement of milk production in buffaloes. These ncRNAs have the potential to be used as biomarkers once their role in regulating specific genes and pathways during lactation is known.

Acknowledgement

We acknowledge the Indian Council of Agricultural Research (ICAR), New Delhi for financial assistance under the Consortium Research Platform-Genomics (Animal Science). The help and cooperation of Director, ICAR-National Bureau of Animal Genetic Resources (NBAGR), Karnal for providing necessary facilities is thankfully acknowledged.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aguilera, P.M., Debat, H.J. and Grabiele, M. 2017. Dataset of the transcribed 45S ribosomal RNA sequence of the tree crop "yerba mate". *Data in brief*. **12**: 649-651.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arora, R., Sharma, A., Sharma, U., Girdhar, Y., Kaur, M., Kapoor, P., Ahlawat, S. and Vijh, R.K. 2019. Buffalo milk transcriptome: A comparative analysis of early, mid and late lactation. *Sci. Rep.* **9:** 5993.

Avril-Sassen, S., Goldstein, L.D., Stingl, J., Blenkiron, C., Le Quesne, J., Spiteri, I., Karagavriilidou, K., Watson, C.J., Tavare, S., Miska, E.A. and Caldas, C. 2009. Characterisation of microRNA expression in post-natal mouse mammary gland development. *BMC Genom*.**10**: 548.

BAHS-Basic Animal Husbandry & Fisheries Statistics (AH series -18*)*. 2019. Government of India, Ministry of Agriculture. Department of Animal Husbandry, dairying & Fisheries, KrishiBhavan, New Delhi. 1–166.

Chan, P.P. and Lowe, T.M. 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res*.**44**: 184–189.

Do, D.N. and Ibeagha-Awemu, E.M. 2017. Non-Coding RNA Roles in Ruminant Mammary Gland Development and Lactation.*Curr. Topics in Lactation. IntechOpen.* London. 10.5772/67194

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinfo.* (Oxford, England). **29**:15-21.

Gigli, I. and Maizon, D. 2013. microRNAs and the mammary gland: a new understanding of gene expression. *Genet. Mol.Biol.* **36**:465-474.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C. and Hubbard, T.J. 2012. GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.***22**:1760–1774.

Kalyan, U.P.V., Vasantha, S. K. I. and Tej, J.N.K. 2022. Effect of seasonal variation in Temperature Humidity Index on milk production and its composition in Murrah buffaloes. *J. Vet. Anim. Sci*. **53**(1): 89-93.

Kozomara, A. and Griffiths-Jones, S. 2011. miRBase: integrating microRNA

annotation and deep-sequencing data. *Nucleic Acids Res*. **39**: 152–157.

Kumar, V. 2017. Genetic and Breeding Aspects of Lactation. *Trends Adv. Vet. Genet*. IntechOpen, London. 10.5772/66179.

Langmead, B. and Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. ***9***:357–359.

Li, H., Huang, K., Wang, P., Feng, T., Shi, D., Cui, K., Luo, C., Shafique, L., Qian, Q., Ruan, J. and Liu, Q. 2020. Comparison of Long Non-Coding RNA Expression Profiles of Cattle and Buffalo Differing in Muscle Characteristics. *Front. Genet*. **11:**98. https://doi.org/10.3389/fgene.2020.00098

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* **17**:10-12.

Neelima, J., Sajith, P., Ally, K., Deepa, A. and Shibu, S. 2021. Effect of supplementation of rumen protected choline and methionine on milk yield and composition of early lactating dairy cows. *J. Vet. Anim. Sci*. **52**(2): 142-148.

Rozowsky, J., Kitchen, R.R., Park, J.J., Galeev, T.R., Diao, J., Warrell, J., Thistlethwaite, W., Subramanian, S.L., Milosavljevic, A. and Gerstein, M. 2019. exceRpt: A Comprehensive Analytic Platform for Extracellular RNA Profiling. *Cell syst.***8:**352–357.

Shore, A.N., Herschkowitz, J.I., and Rosen, J.M. 2012. Noncoding RNAs involved in mammary gland development and tumorigenesis: there is a long way to go. *J. Mammary Gland Biol. Neoplasia*. **17**:43–58.

Szymanski, M., Barciszewska, M.Z., Erdmann, V.A. and Barciszewski, J. 2002. 5S Ribosomal RNA Database. *Nucleic Acids Res.* **30**:176-80.

Tingo, L., Ahlberg, E., Johansson, L., Pedersen, S.A., Chawla, K., Saetrom, P., Cione, E. and Simpson, M.R. 2021. Non-Coding RNAs in Human Breast Milk: A Systematic Review. *Front. Immunol*. **12**.:725323.doi: 10.3389/fimmu.2021.725323

Yu, S., Zhao, Y., Lai, F., Chu, M., Hao, Y., Feng, Y., Zhang, H., Liu, J., Cheng, M., Li, L., Shen,W. and Min, L. 2017. LncRNA as ceRNAs may be involved in lactation process. *Oncotarget*. **8**:98014–98028.

Zhang, P., Wu, W., Chen, Q. and Chen M. 2019. Non-Coding RNAs and Their Integrated Networks. *J.Integr. Bioinform.* **16**.

Zhou, Q., Li, M., Wang, X., Li, Q., Wang, T., Zhu, Q., Zhou, X., Wang, X., Gao, X. and Li, X. 2012. Immune-related microRNAs are abundant in breast milk exosomes. *Int. J. Biol. Sci*. **8**:118–123. ■