RESEARCH ARTICLE

Crossbred Bull Fertility Prediction Using a Transcriptome-based SNP Methodology

Manish Kumar Sinha[1], Arumugam Kumaresan[2], Shivanagouda Patil[2], Aranganathan V[1]*

Abstract

The utilization of high-throughput data sequencing has been instrumental in the generation of extensive datasets derived from diverse studies, with a primary focus on elements such as epigenetics, genomic variations, genome-wide associations, and differential gene expression in the context of bull fertility. However, the task of achieving effective prediction remains a formidable challenge. The current investigation aims to examine variations identified through RNA-Seq, with a particular emphasis on coding region alterations, and subsequently, map these variations to the region of genes associated with fertility functions and proximity. In this study, transcriptome analysis was carried over sperm samples obtained from 12 bulls, half of which exhibited high fertility (n=6) and the other half low fertility (n=6) based on their conception rates. Bioinformatics tools align the data; genetic variants were identified and mapped subsequently to known fertility quantitative trait loci (QTL). After filtering for fertility-specific single nucleotide polymorphisms (SNPs) with missense mutations, we identified a total of six significant biomarkers. This research demonstrates the efficacy of our innovative approach for screening genetic variants, ultimately leading to the identification of fertility markers within bovine spermatozoa.

Key words: Bovine spermatozoa, Fertility markers, Quantitative trait loci (QTL), RNA-seq, Single nucleotide polymorphisms (SNPs).

Ind J Vet Sci and Biotech (2024): 10.48165/ijvsbt.20.1.20

Introduction

Microarrays were once prevalent, but with the emergence of high-throughput sequencing technologies, it became feasible to conduct in-depth examinations of the biological machinery (Özbek *et al*., 2021). The concealed information at various levels such as DNA, RNA and proteins could be scrutinized. Researchers have employed next-generation sequencing as a tool to sequence both DNA and RNA and by comparing with the available *Bos taurus* genome as a reference from sources like NCBI and Ensembl variation present can be understood (Whitacre *et al*., 2015). As time has progressed, the cost associated with sequencing has decreased, rendering this approach more accessible to researchers seeking to understand the variations within a specific scenario (Phillips *et al*., 2014). It is worth noting that, at a specific moment, a particular set of genes undergo expression, and it is also known that sperm is transcriptionally inactive (Ren *et al*., 2017). Therefore, RNA required for post-fertilization functions and to transfer genomic DNA from the male counterpart is packaged during spermatogenesis (Sahoo *et al*., 2021).

An earlier investigation has been undertaken, elucidating the transcriptome profile within spermatozoa (Prakash *et al*., 2020). Additionally, an examination of the differential expression of transcripts between highly fertile and sub-fertile bulls has been conducted (Saraf *et al*., 2021). Given that fertility is a pivotal trait crucial to the selection of superior breeding stock, it has been deemed a pivotal determinant influencing breeding programs (Haskell *et al*., 2014). RNA-Seq generates extensive large-scale data, necessitating a meaningful biological interpretation (Deshpande *et al*., 2023). To facilitate data analysis, multiple bioinformatics tools are available (Ramya *et al*., 2021). Mere acquisition of markers through analysis lacks significance. It is crucial to investigate single nucleotide polymorphisms originating from the carrier genome, as these variants may be of great importance (Sinha *et al*., 2022). The observed variations can be linked to previously reported fertility quantitative trait loci (QTLs). Therefore, the current investigation was oriented towards devising a methodological strategy for detecting variations in two sample groups (high and low fertile) by alignment to the reference genome and determining read depth at variant positions, subsequently establishing a scientific approach for linking these variations with fertility quantitative trait loci (QTL). This systematic pipeline holds the potential to facilitate

[1]Jain University (Deemed-to-be University), JC Road, Bengaluru - 560069, Karnataka, India.

[2]Theriogenology Laboratory, Southern Regional Station of ICAR-National Dairy Research Institute, Bengaluru - 560030, Karnataka, India.

Corresponding Author: Dr. Aranganathan V, Jain University (Deemed-to-be University), JC Road, Bengaluru - 560069, Karnataka, India, e-mail: v.aranganathan@jainuniversity.ac.in

How to cite this article: Sinha, M. K., Kumaresan, A., Patil, S., & Aranganathan, V. (2024). Crossbred Bull Fertility Prediction Using a Transcriptome-based SNP Methodology. Ind J Vet Sci and Biotech. 20(1), 97-101.

Source of support: Nil

Conflict of interest: No potential conflict of interest was reported by the author(s).

Submitted 19/10/2023 **Accepted** 29/11/2023 **Published** 10/01/2024

the development of predictive markers for bull fertility in sperm analysis.

Materials and Methods

The present investigation was undertaken at Jain University Bengaluru, with experiments conducted at the Theriogenology Laboratory of the NDRI, Bengaluru.

Sample Source, RNA Extraction and Sequencing

Twelve experimental bulls, with Holstein Friesian crossbred genetics comprising 50-75 % exotic inheritance were chosen from a pool of 50 bulls through an assessment of their established fertility status. These bulls are consistently employed in an artificial breeding program. Their categorization into high or low fertility (HF or LF, n=6 in each group) was determined according to their respective field conception rates (CR). Bulls exhibiting a CR Mean+1 standard deviation (SD) were categorized as high-fertile (HF), while those with a CR Mean-1 standard deviation (SD) were classified as low-fertile (LF).

For RNA extraction, semen was purified using Percoll gradient (90-45 % discontinuous) centrifugation to eliminate the semen extender and epithelial cell contamination. Total RNA was extracted from frozen sperm using TRIzol (Ambion, Thermo Fisher Scientific, United States) (Parthipan *et al*., 2015) with minor modifications. Initiating from an initial concentration of 50-100 ng of total RNA extracted from each crossbred bull spermatozoa sample, the synthesis of cDNA was performed, utilizing a blend of oligo (dT) and random hexamers. This process was executed with the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, United States, Catalog number K1622), adhering to the manufacturer's guidelines to attain a final volume of 20 µL. Two representative cDNA samples were generated by mixing equimolar quantities of sperm cDNA from three HF and LF bulls each, and a comparable process was employed to produce two representative cDNA samples from six High- and six low-fertility bulls. Following this, all these samples underwent analysis utilizing the Illumina Nextseq-500 sequencing system. Utilizing 1 µg of total RNA, mRNA enrichment was accomplished through the NEB Magnetic mRNA Isolation Kit (Illumina, USA). Subsequently, the transcriptome library was constructed employing the NEB ultraII RNA library prep kit (Illumina, USA) and subjected to sequencing on the Illumina Nextseq-500 (Illumina, USA). This process yielded a paired-end sequencing format of 2 × 75 base pairs, resulting in the production of approximately 25 million raw reads per sequenced sample.

Read Processing, Alignment and Variant Identification of Bull Sperm

The initial stage following the acquisition of the raw reads entailed subjecting them to quality control analysis through FastQC (developed by Babraham Bioinformatics, UK). Subsequently, the reads underwent processing via Cutadapt (v2.8), a tool employed for the removal of adapter sequences and low-quality bases.

Several tools capable of performing splice mappings, such as TopHat2, STAR, and HISAT2, are available. Among these, HISAT2 (v2.2.1) was employed for mapping against the reference genome (ARS-UCD1.2 – GCA_002263795.2).

To perform variant identification analysis, we quantified the number of mutational events using 'bcftool v1.10.' Subsequently, we applied 'VCFtools v0.1.8' to filter the identified variants, retaining those with a minimum Phred score (Q30) and a minimum coverage of 20 at each base depth. The 'varFilter' function was utilized to generate the filtered VCF file (Danecek *et al*., 2011). These identified filtered variants were subjected to annotation through the utilization of the VEP (Ensembl Variant Effect Predictor) and SnpEff program. The SNPs we uncovered were then compared against the reference bovine genome with the assistance of the general feature file (gff) database, and their functional effects were predicted using the VEP and SnpEff tool (Cingolani *et al*., 2012).

Fertility QTL and Identification of SNP Markers Associated with Fertility

The CattleQTL db, available at https://www.animalgenome.org/, encompasses numerous traits, among which the reproductive traits are noteworthy. Within the realm of reproductive traits, the fertility trait exhibits a comprehensive array of 39 distinct trait attributes. The chromosomal locations corresponding to the QTL location were discerned in the total set of identified SNPs (comprising two samples each of high fertility and low fertility), as indicated in Table 1. Consequently, the total number of variants residing within the fertility QTL was determined using SNP mapping to QTL location.

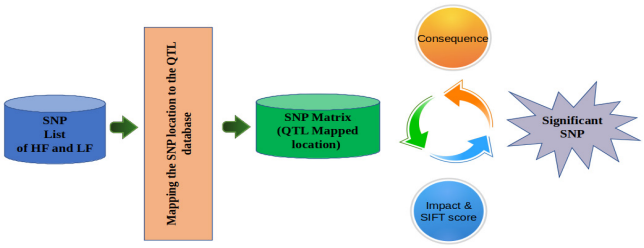

Fig. 1: A methodology for screening through a vast array of SNPs and pinpointing the ones that hold significance to bull fertility.

After individually mapping SNPs for each sample, our objective was to construct a matrix encompassing all four sequenced samples. This procedure proves invaluable in the identification of fertility-line-specific markers for the breed. Subsequently, the matrix was subjected to filtration to isolate SNPs exhibiting mutations in either the HF or LF group (Table 2). This matrix encompasses crucial data pertaining to the mutational effects. Markers were singled out based on whether they exhibited mutations either in both HF or LF samples, and an assessment was made regarding the deleterious or missense effects of these mutations.

Figure 1 depicts the workflow for identifying significant SNPs associated with fertility.

Table 1: After the identification of the SNP, a few entries to observe results (in HF2 sample) when mapping to the QTL database

Chromosome	SNP position	Reference	Allele	rsID Existing variation	Gene Symbol	QTL Start	QTL End	QTL Description
9	32439	T	C	rs209066715	-	27249	208910	Interval from first to last insemination QTL (5006)
20	517237	G	A	rs379225179	SLIT3	477285	650843	Interval to first estrus after calving QTL (5033)
29	766955	G	A	rs42677052	HEPHL1	615557	45287410	Gestation length QTL (5371)
29	1015628	C	T	rs475076470	MED17	615557	45287410	Gestation length QTL (5371)
29	1130078	T	C	rs472509021	CEP295	615557	45287410	Gestation length QTL (5371)
29	2586776	C	T	rs441275872	FAT3	615557	45287410	Gestation length QTL (5371)

Table 2: The matrix contains data related to the four samples, including information about the existing rsID, the consequences of mutations, the mutation's impact, and the SIFT score

Chromosome	Position	Reference	HF1 variant	HF2 variant	LF1 variant	LF2 variant	Symbol	QTL Start	QTL End	QTLDescription	Existing Variant	Consequence	Impact	SIFT
6	22016268	C	--	--	T	T	UBE2D3	14901095	38506182	Calving ease QTL (10759)	rs137527222	Missense variant	Moderate	deleterious low confidence (0.01)
10	21388664	G	C	C	--	--	AP1G2	20879144	51750015	Calving ease QTL (10871)	rs110741547	Missense variant	Moderate	Tolerated (0.55)
1	92137645	G	T	T	--	--	NAALADL2	50246903	102544700	Conception rate QTL (3439)	rs437514718	Missense variant	Moderate	Tolerated (1)
3	89391909	A	G	G	--	--	FYB2	80738997	91082682	Dystocia QTL (30798)	rs526757711	Missense variant	Moderate	Tolerated (0.16)
8	85570327	T	C	C	--	--	-	62616650	104596066	Calving ease QTL (4657)	rs43289335	Missense variant	Moderate	Tolerated low Confidence (0.26)
29	6424919	C	G	G	--	--	TYR	615557	45287410	Gestation length QTL (5371)	rs451256530	Missense variant	Moderate	Tolerated (0.11)

Result And Discussion

In the present study, we evaluated the genetic variation profile of crossbred bull spermatozoa and detected distinct positional variant profiles between high- and low-fertile crossbred bulls. We utilized a specific method to study spermatozoa and to identify SNPs as potential fertility-line-specific markers and the subsequent results along with possible explanations are discussed below. Different types of spermatozoal RNAs including rRNA, mRNA and both large and small non-coding RNAs are present. By employing RNA-seq, a high-throughput sequencing method, it is possible to generate comprehensive transcriptomic profiles of sperm cells (Jodar *et al*., 2013). Furthermore, it is known that sperm carry various RNAs to the oocyte during fertilization, which likely influences fertilization, embryo development, the phenotype of the offspring and possibly future generations. Previous research reports are available where the differential expression was studied using RNA-Seq data (Prakash *et al*., 2021). As a result, RNA plays a crucial role in studying various groups and is associated with sperm function. High throughput sequencing resulted from Illumina Next Seq-500 RNA sequencing gave the raw reads of ~29 million from the high fertile semen sample and ~28.5 million in the low-fertile sample. After processing raw reads processed reads of ~26 million in high-fertile populations and ~23 million in low-fertile populations were obtained. On mapping against the reference genome, a total of ~77k SNPs were detected in crossbred bull spermatozoa samples, with a minimum read depth of 20 reads for each SNP detection and a minimum read quality score of q30. Additionally, there were ~42k and ~34k SNP variations found in HF and LF bulls, respectively.

A high number of SNPs are achieved through sequencing, and the many associations of SNPs with male fertility have already been studied. Genes like MTHFR, FASLG and BRCA2 contain variants, and it is evident that they are associated with male infertility, suggesting that SNPs in certain genes are a risk factor in azoospermia (Ghadrkhomi *et al*., 2021). Similarly, many genomic and exome analyses have been carried out in relation to fertility to screen for the maximum number of SNPs. The association of fertility traits with SNPs was made possible through Genome-Wide Association Studies (GWAS), which showed that the potential SNPs had non-synonymous mutation effects (Greither *et al*., 2022). The Cattle Quantitative Trait Locus Database (CattleQTLdb) contains curetted data from association studies (Hu *et al*., 2021). The locus information obtained from CattleQTLdb is of great importance, as many numbers of SNPs have been obtained, and obtaining relevant SNPs from a list of approximately 77k, mapping to fertility cattle QTL, we found that ~23k SNPs mapped. Thus, by using high-quality data and mapping to reference genome and then subsequently mapping the SNPs to fertility QTL ~25k SNPs were seen in the HF group and ~21K SNPs in LF group (Table 1).

A novel approach was employed to identify significant SNPs that may vary between two groups. SNPs were mapped to QTLs in the four samples from the two groups and an SNP matrix was generated (Table 2). The effect of mutations is a key factor in identifying significant SNPs, along with the QTL association. Occasionally, mutations in the genome are observed. For example, in the region of BTA27 (29.7 - 39.7 Mb), five missense mutations were identified, and low-fertility bulls carried a deleterious mutation (rs109302554, SIFT = 0.01) in the PLEKHA2 gene were reported previously (Abdollahi-Arpanahi *et al*., 2021).

In the present analysis, a total of 6 significant fertility line-specific markers were obtained (Table 2). The attributes that were considered for acquiring these SNPs were Consequence, IMPACT and SIFT score (using the VEP tool). Out of the six fertility line-specific markers, one missense mutation, BTA6:22016268 C>T, showed an identical effect in both LF samples and is related to calving ease. Calving ease is highly significant, as it determines the number of unassisted births a bull will produce when mated to heifers (Dhakal *et al*., 2013). Five fertility line-specific markers were observed in the HF sample. The QTLs observed in these five markers were identified as associated with gestation length, calving ease, dystocia and conception rate. These fertility QTLs and observed mutations are certainly associated with reproductive performance and the mutation effects from these SNPs incidents could be a positive and beneficial factor. They might be reasons for high fertility.

Conclusion

The present study analyzes transcriptome data of spermatozoa using a novel approach to identify significant variants in the genes that regulate fertility. The research indicates that examining SNP, fertility QTL and the impacts of mutations collectively and the markers identified, *i.e.*, UBE2D3, AP1G2, NAALADL2, FYB2 and TYR could be potential markers and bring a paradigm shift in the crossbred bull fertility assessment.

Acknowledgements

The authors thank the Director, ICAR-National Dairy Research Institute, India for providing the necessary facilities for carrying out this research. The authors acknowledge the help from BIOSTICA SERVICES Pvt. Ltd., Bengaluru, India for sequencing data formal analysis.

References

Abdollahi-Arpanahi, R., Pacheco, H.A., & Peñagaricano, F. (2021). Targeted sequencing reveals candidate causal variants for dairy bull subfertility. *Animal Genetics, 52(4)*, 509-513.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, *SnpEff Fly, 6(2)*, 80-92.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics, 27(15)*, 2156-2158.

Deshpande, D., Chhugani, K., Chang, Y., Karlsberg, A., Loeffler, C., Zhang, J., Muszyńska, A., Munteanu, V., Yang, H., Rotman, J., Tao, L., Balliu, B., Tseng, E., Eskin, E., Zhao, F., Mohammadi, P., P. Łabaj, P., & Mangul, S. (2023). RNA-seq data science: From raw data to effective interpretation. *Frontiers in Genetics, 14,* 997383.

Dhakal, K., Maltecca, C., Cassady, J., Baloche, G., Williams, C., & Washburn, S. (2013). Calf birth weight, gestation length, calving ease, and neonatal calf mortality in Holstein, Jersey, and crossbred cows in a pasture system. *Journal of Dairy Science, 96(1)*, 690-698.

Ghadrkhomi, E., Angaji, S.A., Khosravi, M., & Mashayekhi, M.R. (2021). Association of novel single nucleotide polymorphisms of genes involved in cell functions with male infertility: A study of male cases in northwest Iran. *Journal of Reproduction & Infertility, 22(4)*, 258.

Greither, T., Behre, H.M., & Herlyn, H. (2022). Genome-wide association screening determines peripheral players in male fertility maintenance. *International Journal of Molecular Sciences, 24(1)*, 524.

Haskell, M.J., Simm, G., & Turner, S.P. (2014). Genetic selection for temperament traits in dairy and beef cattle. *Frontiers in Genetics, 5*, 368.

Hu, Z.L., Park, C.A., & Reecy, J.M. (2021). Bringing the Animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. *Nucleic Acids Research, 50(D1)*, D956 - D961.

Jodar, M., Selvaraju, S., Sendler, E., Diamond, M.P., & Krawetz, S.A. (2013). The presence, role and clinical use of spermatozoal RNAs. *Human Reproduction Update, 19(6)*, 604-624.

Özbek, M., Hitit, M., Kaya, A., Jousan, F.D., & Memili, E. (2021). Sperm functional genome associated with bull fertility. *Frontiers in Veterinary Science, 8*, 610888.

Parthipan, S., Selvaraju, S., Somashekar, L., Kolte, A.P., Arangasamy, A., & Ravindra, J.P. (2015). Spermatozoa input concentrations and RNA isolation methods on RNA yield and quality in bull (*Bos taurus*). *Analytical Biochemistry, 482*, 32-39.

Phillips, K.A., Trosman, J.R., Kelley, R.K., Pletcher, M.J., Douglas, M.P., & Weldon, C.B. (2014). Genomic sequencing: Assessing the health care system, policy, and big-data implications. *Health Affairs, 33(7)*, 1246-1253.

Prakash, M.A., Kumaresan, A., Ebenezer Samuel King, J.P., Nag, P., Sharma, A., Sinha, M.K., Kamaraj, E., & Datta, T.K. (2021). Comparative transcriptomic analysis of spermatozoa from high- and low-fertile crossbred bulls: Implications for fertility prediction. *Frontiers in Cell and Developmental Biology, 9,* 647717.

Prakash, M.A., Kumaresan, A., Sinha, M.K., Kamaraj, E., Mohanty, T.K., Datta, T.K., & Morrell, J.M. (2020). RNA-Seq analysis reveals functionally relevant coding and non-coding RNAs in crossbred bull spermatozoa. *Animal Reproduction Science, 222*, 106621.

Ramya, L., Swathi, D., Archana, S.S., Lavanya, M., Parthipan, S., & Selvaraju, S. (2021). Establishment of bioinformatics pipeline for deciphering the biological complexities of fragmented sperm transcriptome. *Analytical Biochemistry, 620*, 1141410.

Ren, X., Chen, X., Wang, Z., & Wang, D. (2017). Is transcription in sperm stationary or dynamic? *Journal of Reproduction and Development, 63*(5), 439-443.

Sahoo, B., Choudhary, R.K., Sharma, P., Choudhary, S., & Gupta, M.K. (2021). Significance and relevance of spermatozoal RNAs to male fertility in livestock. *Frontiers in Genetics, 12,* 768196.

Saraf, K.K., Kumaresan, A., Sinha, M.K., & Datta, T.K. (2021). Spermatozoal transcripts associated with oxidative stress and mitochondrial membrane potential differ between high- and low-fertile crossbred bulls. *Andrologia, 53(5)*, e14029

Sinha, M.K., Kumaresan, A., Rao Talluri, T., Ebenezer Samuel King, J.P., Prakash, M.A., Nag, P., Paul, N., Raval, K., Kamaraj, E., & V.A. (2022). Single nucleotide polymorphisms cumulating to genetic variation for fertility in crossbred (*Bos taurus* × *Bos indicus*) bull spermatozoa. *Animal Biotechnology, 34(7)*, 2875-2886.

Whitacre, L.K., Tizioto, P.C., Kim, J., Sonstegard, T.S., Schroeder, S.G., Alexander, L.J., Medrano, J.F., Schnabel, R.D., Taylor, J.F., & Decker, J.E. (2015). What's in your next-generation sequence data? An exploration of unmapped DNA and RNA sequence reads from the bovine reference individual. *BMC Genomics, 16(1)*, 1-7.