Amplification, copy number variation, expression and association of non-synonymous SNP of bovine beta-defensin 129 gene with distinct fertility of cattle bull

Subhash Solanki

National Dairy Research Institute <https://orcid.org/0000-0002-5989-6732>

Poonam Kashyap

National Dairy Research Institute

Ashutosh vats¹

National Dairy Research Institute

Vijay Kumar

National Institute of Immunology

Martina Pukhrambam

National Dairy Research Institute

Sachinandan De

National Dairy Research Institute

Rakesh Kumar

National Dairy Research Institute

Tirtha Kumar Datta* (✉ tirthadatta@gmail.com)

National Dairy Research Institute <https://orcid.org/0000-0003-2426-8901>

Research Article

Keywords: Bovine, male reproduction, polymorphism, beta-defensin, sperm, and fertility

Posted Date: July 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-744570/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The male reproductive specific class-A β -defensins are adsorbed on sperm surface and enrich sperm functioning thus considered vital for maintaining male fertility. The primate DEFB129 play role in sperm maturation, motility, and fertilization but its contribution to bovine fertility is still unexplored.

Method: RLM-RACE and RT-qPCR approaches were used to characterize and expression analysis of Indian cattle BBD129 gene. The polymorphism analysis of the BBD129 gene was done by PCR, sequencing, and absolute RT-qPCR on sperm gDNA from distinct fertility cattle bulls. Bioinformatic analysis was performed to understand the structural and functional implications of SNP on BBD129 protein.

Results: The complete coding sequence of the BBD129 gene consists of 582 bp mRNA including UTRs and conserves all beta-defensin-like characteristics. Sequencing results revealed two conserved non-synonymous T169G (rs378737321, S57A) and A329G (rs383285978, N110S) SNPs in the functional protein-coding exon. Based on SNP position and linkage, BBD129 gene haplotypes were categorized into four groups: TA haplotype (169T & 329A), GA haplotype (T169G polymorphism), TG haplotype (A329G polymorphism), and GG haplotype (when T169G & A329G polymorphisms present together). The frequencies distributions of BBD129 haplotypes in the high fertile group (n=105 clones) were: TA (71.42%), GA (1.90%), TG (2.8%), and GG (24.76%), while in the low fertile group of bulls, the frequencies distributions of observed BBD129 haplotypes (n=149 clones) were: TA (36.24%), GA (0%), TG (2.68%), and GG (61.07%). The distributions of TA haplotype were majorly distributed in bulls with a high conception rate ($P=0.5256$) while double mutated GG haplotype was significantly more abundant in bulls with a lower conception rate ($P=0.0001$). BBD129 exist as a single-copy gene in the bovine genome and found higher expression in the corpus-epididymis region. Bioinformatic analyses found nsSNPs as neutral and non-deleterious but their structural-distorter could result in altered mRNA secondary structure, protein conformations decreased protein stability, and compromised biological functionalities. The polymorphisms resulted in altered O-glycosylations (deletion S57A and insertion N110S) and an increase in phosphorylations (52T-Threonine and 110S-Serine) post-translational-modifications.

Conclusion: BBD129 gene polymorphism could be associated with the fertility performance of cattle bulls.

Introduction

The molecular mechanism of reproduction is a complex process involving thousands of genes encode for proteins or glycoproteins and any defect or mutation in these genes directly or indirectly influences the process of gametogenesis, transportation of sperm to the egg, fertilization, and embryo formation [1–3]. In the last decade, advances in molecular-omics (genomics, transcriptomics, and proteomics) technologies have exceedingly improved the identification of fertility associated markers for selecting the best breeding domestic animals [4–7]. The testicular spermatozoa are non-motile and do not have the fertilizing ability, therefore, the complete sperm maturation takes place in the epididymis where they are exposed to a consortium of molecules such as lipids, organic ions, energy components, enzymes, and glycoproteins secretions from the epididymis epithelial cells in order to attain maturity for fertilization [8–9]. These uptakes of epididymal glycoproteins on sperm surface protect sperm from premature capacitation, acrosomal reaction and help sperm to cross the hostile female reproductive tract (FRT) barriers including cervical mucus passage, uterine immune evasion, OECs binding, and zona interactions [2, 8, 10–13]. The presence of an essential class of glycosylated beta-defensin proteins is one of the crucial modifications of sperm surface during epididymal maturation [9].

β -defensin (BD) is a most ancient class of defensin, it contains C1-C5, C2-C4, C3-C6 disulfide linkages [14–16] that give rise to the typical BD motif with defensin fold containing three antiparallel beta-strands [17–19]. Most of the reported BD genes are encoded by two exons and show a conserved signal sequence. The signal peptide is encoded by the first exon and cleaved off to give a mature functional peptide is encoded by the second exon. Amino acids exposed on the surface of functional proteins are conserved at cysteine positions suggesting the stabilization of core configuration of protein structure for their conserved antimicrobial functions in the species-specific adaptation during the evolution [20–22]. Although originally defensins were intact, during the course of evolution, β -defensin genes have undergone duplications and non-synonymous mutations giving rise to different clusters with region-specific functions and also have gained their reproductive specific functions [14, 23]. Out of many, class-A BDs (including DEFB126 and DEFB129) have shown age-specific and sex-specific expressions signifying their pleiotropic physiological relevance in humans [24], macaque [10–11, 25], rat [26–27], bovine [28], ovine [18], porcine [29], and equine [30]. In primates, the heavily glycosylated DEFB126 and rat DEFB22 (orthologs) are coated on the sperm surface which facilitates sperm functioning, and releasing of these BDs during

capacitation events in the FRT are essential for the sperm-zona interactions [25]. The polymorphisms in the BDs gene lead to decreased milk contents [31–32], increased round cells in semen [33], compromised sperm ability to cross mucus penetration ability [2, 34], the binding with OECs [10, 35], and egg binding [2]. The addition of recombinant BDs (r-BDs) to the deficient spermatozoa, improved their antimicrobial activity, motility [36, 37], OECs, and zona binding abilities [2; 10]. The buffalo beta-defensin 129 (BuBD129) gene has emerged as a novel orthologous of primate DEFB126 in bovine bearing maximum expression and long high potential glycosylated tail than other members of class-A beta-defensins (CA-BDs) [15, 23]. Rat DEFB129 or human DEFB116 expressed in the epididymis binds to the sperm surface chemokine receptor CCR6 and affects the sperm motility by influencing the calcium ion influx [38].

In spite of their importance, however, the genomic characterization of most of the β -defensin genes across the species including bovine remains incomplete and predictive in nature. Keeping in mind all importance of CA-BDs in male reproduction, the present study aimed to determine and characterization of complete architecture of the BBD129 gene sequence including UTR regions derived from random MRT of matured bull and also, determination of BBD129 mRNA expression patterns in the MRT. Finally to relate associations of BBD129 gene polymorphisms with high fertile and low fertile cattle bulls bearing distinct conception rates based on the first and second artificial inseminations data.

Materials And Methods

Cattle male reproductive tissue sample collection and RLM-RACE assay to amplify BBD129 mRNA

Sample collection and processing: The matured male reproductive tract (MRT) of Indian cattle (n=3 biological replicates, and age 3 to 5 years old) were collected from the Kolkata cattle abattoir, Kolkata, India. MRT samples were collected and washed with 1X PBS saline buffer solution and to maintain RNA quality, samples were frozen in the dry ice within 10-15 min of slaughter. In the laboratory, MRT tissues viz. Rete testes (RT), seminiferous tubule (ST), caput, corpus, cauda, and vas-deferens (supplementary fig 1) were dissected with a sterile scalpel into the 3 to 5 mm size pieces under the sterile condition and dipped into RNA later solution (Sigma, USA) and stored into the liquid nitrogen for further experimentations.

RNA extraction and DNase I treatment

Based on prior literature information [23, 25], the caudal tissues were selected for the total RNA extraction in cross-bred cattle. Total RNA extraction was done by TRIzol (Qiazol, USA) methodology (supplementary file 1). To remove genomic DNA contamination, DNase I (ThermoFisher, catalog # 18047019) treatment was done as per company manufactured protocol. RNA quantification was done by NanoDrop ND-1000 spectrophotometer (Thermo NanoDrop Technologies, USA). The absorbance ratio A260/280 and A260/230 were approximate 2.0 for samples used in this work (supplementary fig 2). The quality of the total RNA was analyzed on non-denaturing (1.2%) agarose (Lonza, Switzerland).

RLM-RACE primer designing strategy, cDNA synthesis and RACE PCR

To amplify the cattle BBD129 mRNA sequence, two sets of primers were designed from the coding region of the BBD129 gene as suggested by the company procedure. The BBD129 mRNA sequence of *Bos taurus* cattle was retrieved from the ENSEMBL database (transcript ID: DEFB129-201 ENSBTAT00000064705.2). All primers were designed by using NCBI primer-Blast tool [39] and the primers parameters were checked by Oligocal [40]. The sequences of primers are given in supplementary table 1.

RLM-RACE kit (Ambion, catalog #1700, USA) was used for the amplification of 5' and 3' ends of BBD129 mRNA and the procedure has been described in supplementary file 1. **Reverse transcription:** SuperScript™ IV First-Strand Synthesis kit (ThermoFisher Scientific, catalog #18091050, USA) was used for cDNA synthesis as per the company manufacturer's protocol. The 5' adapter-ligated RNA (10 μ l) was used as a template and negative Tobacco Acid Pyrophosphatase control and non-template control (NTC) were run to check any gDNA contamination. Briefly, 200 ng of total RNA was mixed with random hexamer & oligo₆(dT) primers (1 μ l each), and DEPC-treated water into a sterile 200 μ l PCR tube followed by incubation at 65°C for 5 minutes and immediate ice incubation for 2 minutes. After that, the 5X SuperScript IV reaction buffer, RNase inhibitor, dNTP mix, and RNase inhibitor, SuperScript IV reverse transcriptase was added to make final volume 20 μ l PCR reaction followed by incubation at 55°C for 30 min in a PCR thermal cycler (BioMetra T-Professional Basic Gradient PCR Machine, Germany). The reverse transcriptase was inactivated as per the kit suggested protocol (heating at 90°C for 10 min). **5' RLM-RACE PCR:** The 5' RLM-RACE outer and inner PCR were performed using primers sets

mentioned in supplementary table 1. A 50µl of PCR reaction composition was consisting a 5' adapter-ligated cDNA template, 5X Phusion HF buffer, dNTPs, primers, Phusion DNA polymerase, and nuclease-free water. The PCR thermal profile was: initial denaturation (98°C, for 30 sec) denaturation (98°C for 30 sec), annealing (63°C for 30 sec) and extension (72°C for 45 seconds) followed by 35 cycles and a final extension at 72°C for 10 minutes. Two negative control reactions viz. negative Tobacco Acid Pyrophosphatase treated cDNA and non-template reaction were run to confirm full-length RNA integrity and gDNA contamination, respectively. The amplified products were run on 2% agarose gel. **3' RACE PCR:** Total RNA extraction and reverse transcription reaction were done as described above. The 3' RACE adapter (5'-GCGAGCACAGAATTAATACGACTCACTATAGGT12VN-3') was used as primer in reverse transcription. The 3' RLM-RACE outer and inner PCR were performed similarly to the 5' RLM-RACE PCR. One cDNA PCR was also run with cDNA primers with PCR reaction composition and thermal profile as mentioned in 5' RACE PCR for amplification of the internal region of BBD129 mRNA. The list of primers used in this study is given in supplementary table 1.

Cloning and sequencing of amplified BBD129 mRNA.

Cloning: The cloning was done by using CloneJET 1.2 kit (ThermoFisher Scientific, #K1231). Ligation reactions of 20 µl volume was consisting pJET1.2 blunt cloning vector, T4 DNA ligase, T4 DNA ligase buffer, PCR eluted 3' and 5' RLM-RACE products, and nuclease-free water. Reactions were mixed gently followed by incubation at room temperature (RT) for 1 hour and subsequently used for the bacterial transformation. **Transformation:** The competent XL-1 Blue strain of E.coli cells (50 µl) was taken from -80°C and thawed on ice for 15 minutes. A 10 µl of ligated reaction was added to competent XL-1 Blue E.coli cells and incubated on ice for 30 min after that cells were challenged for heat transformation at 48°C in water-bath for 90 sec followed by immediate ice incubation for 10 min. A 900 µl of Lysogeny broth (LB) media (Hi-Media, India) and 100 µl of 1M glucose (Sigma-Aldrich, USA) was added to the tube and incubated at 37°C for 1 hour in a shaker incubator followed by plating on LB agar medium containing ampicillin antibiotic (50 µg/ml) and plates were incubated at 37°C for overnight. **Colony PCR confirmation:** Overnight grown colonies were processed for colony PCR. The colony PCR reaction compositions and thermal profile were similar as mentioned in the RLM-RACE PCR. The positive clones processed further for plasmid isolation. **Plasmid isolation and plasmid PCR:** It was performed manually by the alkaline lysis method as mentioned in supplementary file 1. The quantity and quality of isolated plasmids were evaluated by NanoDrop ND1000 spectrophotometer and 1% agarose gel electrophoresis (AGE).

Characterization of cattle BBD129 gene and phylogenetic analysis

Genomic characterization: Sequencing result were BLAST (BLASTn and BLASTp) against cattle genome (*Bos taurus x Bos indicus*) on NCBI (UOA_Brahman_1 GCF_003369695.1) and USSC-Genome Browser (Apr. 2018 ARS-UCD1.2/bosTau9). **Cysteine bond pairing prediction** was done by DiANNA Web server (41). **Secondary structure prediction** was done by PSIPRED and SOPMA server [42-43]. **Signal peptide prediction** was done by Signal IP 5.0 server [44]. **Biological function prediction** was done by Argot2 server [45]. The complete sequence of cattle BBD129 gene was submitted on NCBI-BankIt [46].

Evolutionary phylogenetic tree analysis: The protein sequences of the DEFB129 gene from ruminants and non-ruminants mammalian species were retrieved from the NCBI database by using the "beta-defensin 129" keyword (supplementary file 2). To find the evolutionary relationships between cattle BBD129 (Bovidae family) identified by RLM-RACE and its neighbor a family, a phylogenetic evolutionary analysis was performed by the Maximum Likelihood method and Tamura-Nei model by using the MEGA 10 bioinformatics tool [47]. This analysis involved 29 protein sequences of DEFB129 genes. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed.

Amplification of BBD129 gene from sperm gDNA

Collection of frozen semen and sperm genomic DNA extraction

Sample collection: Eleven distinct fertility cattle bulls were selected on the basis of their first and second artificial inseminations conception rate (CR) data available at the Artificial Breeding Research Center, NDRI, India. Cattle bulls with CR less than 31% were classified in the low fertile group and bulls with CR with more than 50% were classified in the high fertile group (Table 1). Frozen semen samples of selected cattle bulls of distinct fertility (5 bulls from high and six from low) were collected into liquid nitrogen.

Genomic DNA isolation: Frozen semen samples were thawed by immersing them into the pre-warmed 37°C distilled water for 30 sec and contents were collected into the 15ml falcon tubes containing 8ml of washing buffer-A (150mM NaCl and 10mM EDTA) and centrifuged at 800g for 10 min. Pellets were dissolved into the 300 µl of lysis buffer-B (100mM TRIS-Cl, 10mM EDTA, 0.5M NaCl and

1% SDS) and 100µl of 1M DTT (Dithiothreitol) followed by incubation for 15 min at RT. After that, 100µl of Proteinase-K enzyme (0.2 mg/ml) was added to the tube followed by overnight incubation at 55°C in the dry bath. The sperm gDNA was isolated by the Phenol:Chloroform: Isoamyl alcohol (PCI) method [48]. The quality and quantity of gDNA were measured by NanoDrop ND1000 spectrophotometer and 1% AGE.

DEFB129 genomic DNA amplification and cloning

Two sets of primers were designed by using the NCBI primer tool to amplify both the exons of the BBD129 gene. PCR reactions of 50µl contained 2µl of each primers [exon 1 (Forward = 5'GAGGTCTTCCTTCTTGTTACTGA3', Reverse = 5'AACCCAGATCCAAGGCTTCTT-C3') and BBD129 exon 2 (Forward = 5'CCATCAACGTC-CCTAACCACT3', Reverse = 5'CTGTGAGGCTCCTGTGAGAAA3')], dNTPs, Phusion buffer 5X, Phusion high fidelity polymerase, genomic DNA template and remaining nuclease-free water. PCR thermal reaction profile was: initial denaturation at 98°C for 5 min, denaturation 98°C for 30 sec, annealing 58.2°C for BBD129 exon 1 and 60°C for BBD129 exon 2, and polymerization at 72°C for 40 sec. The 35 cycles were run followed by a final extension at 72°C for 10 min. NTC was included as a negative control to check nucleic acid contaminations. **Bacterial cloning and PCR** was done as mentioned above (Cloning and sequencing of amplified BBD129 mRNA section) with the following modifications: exon 1 primer (Forward = 5'GAGGTCTTCCTTCTTGTTACTGA3', Reverse = 5'AACCCAGATCCAAGGCTTCTT-C3'), exon 2 primers (Forward = 5'CCATCAACGTC-CCTAACCACT3', Reverse = 5'CTGTGAGGCTCCTGTGAGAAA3') and annealing (58.2°C for BBD129 exon 1 and 60°C for BBD129 exon 2). Positive plasmids were processed for sequencing.

Absolute quantification of BBD129 gene copy number

The protocols and procedures used for the optimization of primer concentration, gDNA template, and generation of the standard curve are provided in supplementary file 1. To determine the absolute copy number of BBD129 gene in genomic DNA of distinct fertility cattle bulls, the RT-qPCR reaction was set up with the following components and thermal profile: SYBR (5 µl), forward primer (1.5µl), reverse primer (1.5 µl), nuclease-free water, and gDNA template (2µl of 1.25ng/µl). The thermal profile was: initial denaturation (95°C for 3 min), denaturation (95°C for 10 sec), annealing (49°C for 15 sec), extension (72°C for 15 min), and repeated run for 35 cycles. After the amplification, a melting peak analysis with a temperature gradient of 0.5°C/sec from 65°C to 95°C was performed.

The expressional analysis of Indian cattle BBD129 by RT-qPCR assay

Reverse transcription and Primer designing

Total RNA extraction from frozen cattle MRT tissues and cDNA synthesis was done as mentioned above (RNA extraction & DNase I treatment and RLM-RACE primer designing strategy, cDNA synthesis, and RACE PCR sections). The primers were designed from our RLM-RACE BBD129 product (supplementary table 1). GAPDH (glyceraldehyde-3-phosphate dehydrogenase) and eEF-2 (eukaryotic elongation factor 2) reference genes were chosen based on the literature [18, 49]. RT-qPCR optimization and validation of reference genes are provided in supplementary file 1.

Annealing temperature (AT) optimization was done by the conventional gradient PCR. The best AT was selected based on the brightest and sharpest band obtained on 2% agarose gel. PCR chemical composition involves cDNA template, dNTPs, gene-specific primers, Taq DNA polymerase, reaction buffer, and Nuclease-free water in a 25µl reaction volume. The PCR thermal profile was: Initial denaturation at 95°C for 5 minutes, denaturation at 95°C for 30 seconds, annealing-variable between 45-55°C C for 30 seconds, extension at 72°C for 15 seconds and a final extension for 5 minutes at 72°C.

The selection of best primer concentration was done by RT-qPCR assay with the following reaction components: SYBR (5 µl), forward primer (0.25µl/0.5µl/1µl/1.5µl), reverse primer (0.25µl/0.5µl/1µl/1.5µl), nuclease-free water, and cDNA template (2µl of 10ng/µl). The following RT-qPCR thermal profile was used: initial denaturation (95°C for 3 min), denaturation (95°C for 10 sec), annealing (49°C for 15 sec, extension (72°C for 15 min), and 35 cycles were run on Bio-Rad CFX96 Maestro qPCR machine by using SYBR Green Master Mix (Bio-Rad, USA). After the amplification, a melting peak analysis with a temperature gradient of 0.5°C/sec from 65°C to 95°C was performed to ensure specificity and primer-dimer formation. RT-qPCR reaction efficiency was done by using a serial dilution by using BBD129 gene-specific primers for the MRT tissues. RT-qPCR efficiency was calculated according to the $E = 10(-1/\text{slope})$ [50].

Cattle BBD129 RT-qPCR assay

The relative expression of the cattle BBD129 gene was done as duplicate in a 10 µl reaction volume with RT-qPCR thermal profile: initial denaturation (95°C for 3 min), denaturation (95°C for 10 sec), annealing (49°C for 15 sec, extension (72°C for 15 min), and 35 cycles were run. After the amplification, a melting peak analysis with a temperature gradient of 0.5°C/sec from 65°C to 95°C was performed to ensure specificity and primer-dimer formation. A non-template control was run to cross-confirmation of reaction component nucleic acid contamination.

Bioinformatics analysis of cattle BBD129 polymorphism

BLASTn and BLASTp alignment analysis was used for the nucleotide and protein search alignment, respectively [51].

Prediction of functional impacts of nsSNPs on mRNA secondary structure

To predict mRNA secondary structure, FASTA formats of native and mutated mRNA were submitted to the RNAfold server [52]. It provides three structural information i) conventional secondary structure graph, ii) Dot plot, iii) Mountain plot. A mountain plot is an x-y graph and shows minimum free energy structural curve, centroid structural curve, partition function curve, and positional entropy resulting from the pair probabilities.

Prediction of functional impacts of nsSNPs on protein stability and functionalities

PROVEN (Protein Variation Effect Analyzer) [53], SIFT (Sorting Intolerant from Tolerant) [54], Polyphen-I (Polymorphism Phenotyping-I) & Polyphen-2 [29], SNPs & GO [55], and PhD-SNP [56] are sequenced and SVM based bio-computational tools used for prediction of the impact of observed amino acid substitutions (S57A & N110S) on BBD129 protein structural information and functions. The impact may neutral (tolerant) and disease-causing (non-tolerant) depending on the alignment and physicochemical properties of amino acids. Polyphen algorithm score denotes substitution benign if the score is <0.4; possibly damaging (0.4-0.9), and probably damaging (>0.9). SIFT alignment score denotes deleterious (<0.05) and neutral (>0.05). PROVEN algorithm score denotes deleterious (<-2.5) and neutral (>-2.5) [57]; SNAP2 predict the disease-related variants [58]; I-Mutant2.0 predict protein stability depending on SVM and ProTherm databases, helped to evaluate AAS free energy changes values (DDG = DGmutant – DG wild type) [56, 59]; MAPP (multivariate analysis of protein polymorphisms) was used for the interpretation of missense variant causing deleterious effect [60]. PSIPRED and SOPMA servers were used for the prediction of the impact of nsSNPs on protein secondary structure. Biological function prediction was done by the Argot2 server [45]. The translated amino acid sequences of native and mutated BBD129 gene were submitted above servers.

Prediction of impact of snSNPs on protein physiochemical properties and post-translational modifications

ProtParam [61] and Mupro [62] bioinformatics tools were used to computes various physicochemical properties such as molecular weight, theoretical pI, amino-acids composition, atomic compositions, aliphatic index, instability, extinction coefficient, protein stability, and estimated half-life. **Post-translational modifications:** i) O-glycosylation site prediction was done by using NetOGlyc 4 server [63] and GlycoEP standard predictor under the composition profile of pattern [64]. ii) N-glycosylation site prediction was done by using NetNGlyc 1.0 server [65] which is ANN (artificial neural networks) based that examine the sequence motif of Asn-Xaa-Ser/Thr amino-acid sequence. iii) Phosphorylation site prediction was done by using NetPhos 3.1 server [66] which works based on neural networks. The translated amino acid sequences of native and mutated BBD129 gene were submitted above servers.

Statistical analysis

GraphPad Prism 5.0 [67] unpaired T-test was used for the analysis of distributions of SNPs and absolute quantification of copy number variations in the distinct fertility bulls. A P-value <0.05 was set to be statistically significant. The geometrical mean sample Cq (cycle of quantification) values for cross-bred BBD129 transcript in triplicate samples and relative expression ratio of BBD129 gene expression was computed using the 2^{-ΔΔCt} (Threshold cycle) method [68] and implemented by the GraphPad Prism 5.0 software. GAPDH and eEF2 were used as reference genes in the expression experiments. The BBD129 gene expression in the seminiferous tubule region of MRT was used to normalize Cq values obtained from other MRT regions. The differential BBD129 gene expression level was analyzed by ANOVA (Tukey posthoc correction test) in GraphPad Prism 5.0.

Complete coding BBD129 gene in Indian cattle

RLM-RACE was performed on the corpus-testicular RNA for the complete amplification of bovine BBD129 mRNA. The 5' RLM-RACE inner PCR had given a band of approximately 190bp long product while 3' RACE PCR had given an approximately 300 bp long amplification product (supplementary fig 3). The BBD129 5', 3', and cDNA PCR sequences obtained were aligned to each other to generate a complete coding sequence of BBD129 mRNA. The FASTA sequences of amplified produced are provided in supplementary file 3. The BBD129 gene is located on chromosome 13 between the nucleotide coordinates chr13: 22660780-22799850 (NCBI: UOA_Brahman_1 GCF_003369695.1 and USSC: Apr. 2018 ARS-UCD1.2/bosTau9) and flanked by the LOC11390270 gene on upstream and the C13H20orf96 gene on the downstream region. BBD129 gene having two exons separated by a ~1.6 kb non-coding intron. The first exon code for the 34 bp long 5' UTR and signal peptide while the second exon code for functional protein and 71 bp long 3' UTR regions (Fig 1). Gene ORF starting from nucleotide 47th and end with 560th nucleotide carrying a 511bp long coding mRNA which codes for 171 amino acids long peptide protein. The detail of genomic characterization was submitted to NCBI with accession number 2448405 and is provided in figure 1 and supplementary table 1. Cattle BBD129 conserve cysteine pairing rule (C1-C5, C2-C4, and C3-C6) and secondary structure with a conserved N-terminal signal peptide of 18 amino acids. The beta-defensins have the properties to exogenously secretion and signal peptide cleavage to give mature functional protein. Bovine BBD129 protein bear protease cleavage site between the amino acids 19 and 20 viz. VNT-EY and has been predicted with a cleavage probability of 0.998 (supplementary fig 4). BBD129 was predicted for conserved three beta-sheets secondary structure, immune defensive roles, efficiency to bind with CCR6 chemokine receptor and lipopolysaccharides binding, and extensive post-translation modifications especially O-glycosylations and phosphorylations (fig 1 and supplementary table 1).

Figure 1

The multiple sequence alignment (MSA) and phylogenetic analyses of the BBD129 gene across the twenty-nine mammalian species indicate a strong evolutionary relationship between the Indian cattle BBD129 gene and the Bovidae family BBD129 (fig 2; supplementary fig 5). The Camelidae and Equidae families show relationships next to the Bovidae family. MSA analysis found Indian cattle are structurally similar to *Bos taurus* except for few amino acids.

Figure 2

Association of cattle BBD129 sequence variations with bull fertility

The BBD129 gene was amplified from sperm gDNA of distinct fertility bulls and the full length of exon first (507bp) and exon second (607bp) was achieved by amplification of intronic flanking regions of exons (Fig 3). To determine the polymorphisms in the BBD129 gene in the distinct fertility bulls, we sequenced 254 clones from 11 bulls. There was no polymorphism observed in the exon first. However, in exon second, we observed two missense or non-synonymous SNPs at positions 169th (T169G cDNA position, rs378737321) and 329th (A329G cDNA position, rs383285978) in the gDNA from distinct fertility bulls (Fig 4; Supplementary fig 6A). After translating nucleotide to amino acid sequence, we have also observed substitutions at the amino acids level. In the BBD129 protein, the serine amino acid at the 57th position is replaced by alanine (S57A/rs378737321) while asparagine amino acid is replaced by serine at the 110th position (N110S/rs383285978) (Fig 4; supplementary fig 6B). Based on polymorphism position and linkage, BBD129 haplotypes were categorized into four groups: TA haplotype (169T & 329A), GA haplotype (T169G polymorphism), TG haplotype (A329G polymorphism), and GG haplotype (when T169G & A329G polymorphisms present together). The frequencies distributions of haplotypes in the high fertile group (n=105 clones) were: TA (71.42%), GA (1.90%), TG (2.8%), and GG (24.76%), while in the case of low fertile group bulls, the frequencies distributions of observed haplotypes (n=149 clones) were: TA (36.24%), GA (0%), TG (2.68%), and GG (61.07%) (Fig 5; Table 1). There was no significant difference in TA haplotype distribution in the high fertile and low fertile groups (P=0.5256) but there was a 35.15% difference in BBD129 TA haplotype in both the group of high fertile and low fertile bulls. There was a significant difference in GG haplotypes frequency distribution in the low fertile bulls as compared to high fertile bulls (P=0.0002). The distribution of double mutated GG haplotype was found an association with bull's low conception rate (table 1).

Table 1

Figure 3

Figure 4

Figure 5

Copy number variation in BBD129 gene in distinct fertility cattle bulls

In order to investigate the fertility-related association in the copy number of the bovine beta-defensin 129 (BBD129) gene, we examined the genomic copy number of the BBD129 gene in two groups of distinct fertility cross-bred bulls (n=11). Standard curve-based quantification of the BBD129 gene was carried out by absolute real-time quantitative PCR. The known concentration (copy number) of the plasmids carrying the gene of BBD129 fragment was used to construct the standard curve for the BBD129 gene by plotting the log concentration values against the crossing point (Cq) values (Supplementary fig 7 & 8). Crossing point (Cq) values were determined from the RT-qPCR run of equal concentration of genomic DNA (1.25 ng/ul) of distinct fertility categorized bulls. The universal bovine CSN2 (casein beta) gene was used as a single-copy number reference gene. The absolute Cq values and their log transformation for BBD129 gene and CSN2 control are provided in supplementary table 1. It was observed that the mean copy numbers of CSN2 and BBD129 genes did not vary significantly between bulls of distinct fertility (Fig 6) suggesting a single copy of BBD129 gene in the bovine genome and figure 10B suggesting no significant difference of BBD129 gene CNV in high fertile and low fertile bulls.

Figure 6:

Expression of beta-defensin BBD129 gene in male reproductive tract

The relative distribution of BBD129 expression was analyzed in adult cattle MRT. The expression dynamics of BBD129 were slightly different from the other mammalian species reported earlier, the higher expression of BBD129 was observed in the corpus segment of the epididymis as compared to others MRT tissues (viz. ST, RT, caput, cauda, and VD) (Fig 7). The corpus regions show 14.2 fold higher expression of the BBD129 gene than the normalizer rete testis region. The order of reducing expression of BBD129 expression was corpus, cauda, VD, RT, and caput region. The details of mean difference and P-value have provided in supplementary table 1.

Figure 7

The possible impact of nsSNPs on BBD129 protein

The possible impact of nsSNPs on mRNA structure

Polymorphism in the BBD129 gene can alter base-pairing probability in the mRNA secondary structure (Supplementary fig 9). The optimal mfe of native BBD129 mRNA (no polymorphism) was predicted at -93.77 kcal/mol while in double mutated BBD129 minimum free energy was -98.20 kcal/mol. The free energy of the native BBD129 thermodynamic ensemble was free -102.28 kcal/mol and the ensemble diversity value was 142.05 while in the case of the double mutated BBD129 thermodynamic ensemble free energy was -107.88 kcal/mol and ensemble diversity value was 119.47. The minimum free energy of native BBD129 centroid secondary structure was 46.40 kcal/mol while in double mutated the value was predicted 61.78 kcal/mol. The MFE structural curve, the thermodynamic ensemble of RNA structural curve, the centroid structural curve, and entropy changes in the mutated BBD129 mRNA can be seen in the mountain plot (Fig 8

Figure: 8

The possible impact of nsSNPs on protein stability

There are many tools used for the predictions of the deleterious effect of observed nsSNPs on BBD129 protein. Polyphen-I, PROVEN, SNPs & GO PhD-SNP, Meta-SNP, and PredictSNP have been predicted as the neutral and non-deleterious effects of observed SNPs on BBD129 protein. SNPs to be benign in both HumDiv and HumVar models analyzing. The deleterious effective score of observed rs378737321 SNP was 0.002 with 0.99 sensitivity and 0.30 specificity and the deleterious effective score of rs383285978 SNP was predicted as 0.004 with 0.97 sensitivity and 0.59 specificity (supplementary fig 10). Similarly, SIFT, SNAP, and MAPP predicted S57A mutation as deleterious or disease-causing polymorphism and N110S mutation as neutral. I-Mutant2.0 server predicted both nsSNPs as structure distorters. The result details of the above bioinformatics servers are provided in table 2.

Table 2

Possible impact of nsSNPs on physiochemical properties of BBD129 protein

Comparative ProtParam results between native BBD129 protein sequence and double mutated protein sequence have shown major alterations in the molecular weight, polar and non-polar amino acids, instability index, aliphatic index, and GRAVY (supplementary table 1). PSI-PRED and SOPMA servers predicted alterations in the protein secondary structure especially in the helix and coils (Fig 4). SOPMA predicted native protein sequence with 33.77 % alpha-helix, 1.32 % beta-turn, 10.60 % extended strand, and 54.30 % random coil, however, in the mutated BBD129 protein the alpha helix increase by 5.3 % while the beta-turn and random coil regions decrease about 1.32 % and 3.97 %, respectively. Alterations in BBD129 protein stability of missense variants were examined by MUpro software. MUpro predicted that both the substitutions negatively affect the protein stability of BBD129 protein. The substitution of serine to alanine amino acid at 57th position (S57A_rs378737321) decreased the protein stability by confidence score -0.83258649 and -0.998569279497182 in Support Vector Machine and Neural Network Machine, respectively. The substitution of asparagine to serine amino acid at 110th position (N110S_rs378737321) decreased the protein stability by confidence score -0.63001616 and -0.65131013529855 in Support Vector Machine and Neural Network Machine, respectively (supplementary table 1).

The possible impact of nsSNPs on BBD129 protein post-translation modification

NetOglyc 4.0 and GlycoEP servers predicted one deletion of the O-glycosylation site at the 57th amino acid position and one insertion of new O-glycosylation site at the 110th amino acid position in the double mutated BBD129 protein (Fig 4; supplementary table 1). SNPs have no effect on N-glycosylation. NetPhos 3.1 predicted fifteen sites for phosphorylations in the BBD129 TA haplotype protein (supplementary table 1) while two possible additional phosphorylation sites were predicted in the double mutated BBD129 GG haplotype protein. The motif CKKKTCCIR (52nd amino acid) was predicted as a potential candidate for threonine phosphorylation with 0.545/0.503 scores for Phosphokinase-G/Phosphokinase-A enzymes. Another motif IKSASAFK (110th amino acid) was predicted as a new potential candidate for serine phosphorylation for the PKC enzyme with a 0.617 score. These new phosphorylation sites were the result of S57A and N110S substitutions, respectively (Fig 4 and supplementary table 1).

The possible impact of nsSNPs on BBD129 protein biological functions

As predicted above that nsSNPs decrease protein stability and influence the protein secondary structure and post-translation modifications, here we have predicted possible impacts observed nsSNPs on the biological functioning of BBD129 protein by Argot 2 server. The prediction found double mutated BBD129 protein has decreased scores for all predicted biological functions compared to native BBD129 protein suggesting nsSNPs negatively impact the protein biological functioning (Table 3).

Table 3

The summarized results of bioinformatics tools used for the predictions of various physiochemical and structural changes in the mutated BBD129 are provided in the table 4.

Table 4

Discussion

The reproductive-specific class-A BDs have got many attractions because of their importance in the fertilization events, however, insufficiency of their complete information and characterizations limit their usages. In the present work, we have characterized the bovine β -defensin 129 gene in Indian cattle and showed its association with bull fertility performances. The complete coding sequence of BBD129 mRNA was amplified from the corpus-testicular region and expression analysis was performed to explore its region-specific roles. The sequence variations analyses were performed by PCR amplification and absolute quantification of BBD129 from sperm genomic DNA of distinct fertility characterized cattle bulls. The bio-computational analyses of non-synonymous polymorphism on the biological functioning of BBD129 protein found an association with cattle bull conception rate percentage.

The testicular sperm are not able to fertilize an oocyte as they need post-gonadal maturations to gain fertilizing ability [69, 70]. The complete maturation of testicular spermatozoa takes place in the epididymal environment where they are exposed to a plethora of molecules including highly negatively charged cysteine-rich glycosylated beta-defensin antimicrobial peptides secreted from the epididymis epithelial cells and coating of these glycoproteins or exogenously addition of epididymal proteins (e.g. BDs) improve fertilization potential of sperm [71–73]. The addition of prokaryotically expressed bovine r-BBD126 protein to non-motile caput-epididymal spermatozoa improves their motility and mucus penetration ability but lacks the ability to improve fertility [74–75]. This failure to improve sperm fertilizing ability could be due to the prokaryotic expression of bovine BBD126 because prokaryotes lack a

post-translational modification system and DEFB126 or its orthologs are known for their post-translational modifications (glycosylations) [23, 71]. Another reason could be the absence of uncharacterized complete coding sequences or predicted sequences of BDs available on the public databases [76]. RACE methodology has been used to amplify capped and tailed mRNAs [77–78], therefore accurately determining the 5' capped end and 3' polyadenylation ends [79–80]. *Bos taurus* BBD129 gene was characterized as two exons and with 50 bp UTR region at both the end of the gene [81]. Herein, we amplified the 5' and 3' end of cattle BBD129 mRNA and revealed strong conservation in the coding region (UOA_Brahman_1 GCF_003369695.1). Bioinformatic analysis revealed its location on chromosome 13 and was observed to retain all the essential characteristics as reported in beta-defensins. The MSA and phylogenetic analysis found its strong conservation in the Bovidae family and other mammalian families. The MSA of DEFB129 across the mammalian species found structural variations in the protein-coding region. The protein structure of BBD129 is almost similar in the Bovidae family excepts very fewer places and the variations increase as phylogenetic distance increase.

In the bovine, studies are largely focused on genomic SNPs associated with milk production, susceptibility to immune function, and parasite resistance [82]. The non-synonymous mutations (CC/AA and CT/AC) genotypes of cattle BNBD4 and DEFB103 are associated with higher milk, fat, proteins, lactose, dry matter contents, somatic cell counts, and resistance to mastitis [72, 75, 83–84]. The genotypic analysis found that the human DEFB126 gene is highly polymorphic bearing 77 known SNPs, out of these, rs140685149 and rs11467497 are associated with human infertility affecting sperm surface glycans, motility (rs11467497), higher round cells in semen, cervical mucus penetration ability [10–11, 33–34, 74]. In *Bos taurus*, the BBD129 gene has been reported for the twelve missense or non-synonymous SNPs in the genome [81]. In this study, we observed two conserved non-synonymous SNPs (T169G and A329G) in the Indian cattle genome suggesting the conservation nature of snSNPs. These non-synonymous nucleotide polymorphisms also substitute non-synonymous amino acids (S57A and N110S, respectively) in translated BBD129 protein. In Indian bovine species, the publically available predicted BBD129 gene has conserved 57A and 110S, however, in our study, RLM-RACE BBD129 amplification found 57S and 110N as reported in *Bos taurus* [81]. The sequencing revealed Indian cattle bulls are producing a heterogeneous population of spermatozoa bearing different haplotypes of the BBD129 gene. The BBD129 native TA haplotype (169T & 329A) and double mutated GG haplotype (169G & 329G) were found majorly distributed in the Indian cattle genome. The TA haplotype is primarily distributed in the group of high fertile bulls with a difference of 35.18 % to low fertile bulls while the major percentage of heterogeneous sperm population from the low fertile bulls found with double mutated GG haplotype and this was significantly associated with bulls of low conception rate ($P = 0.0001$).

The polymorphism and duplication in the β -defensin genes lead to their selective environment expressions in the reproductive organs signifying their role in different manners other than traditional antimicrobial activity [81, 85–86]. The variations in HBD2, DEFB4, DEFB103-07, PRR23D1, and SPAG11 gene copy number (< 4) are associated with altered defensin expressions and biological functions [86]. β -casein gene present as a single copy in the bovine genome [87] and in our study, it was used as a single-copy reference gene. The standard curve analysis has found that there was no duplication in the BBD129 gene and it has been retained as a single-copy gene in the genome of contrasting fertility populations of Indian cattle bulls.

The distributions of the beta-defensins in the MRT across the species are similar suggesting their conserved expression pattern and their traditional antimicrobial activity in the male reproductive organs depicting their importance in the reproductive processes [14, 18, 29, 88–89]. *Bos taurus* and *Bubalus bubalis*, newly emerged primate DEFB126 ortholog viz. BBD129 shows region-specific and sex-specific expression in the healthy MRT suggests its pleiotropic activities, in addition to traditional antimicrobial importance [23, 81]. The DEFB129 is anticipated to play a vital physiological role in sperm maturation via membrane modulations and providing motility [14, 27, 90–91]. Human epididymal DEF129 protein bind to the sperm surface chemokine receptor CCR6 and influence the calcium ion influx causing sperm hypermotility. The exogenous additions of different concentrations of recombinant-BDs to deficient spermatozoa significantly increase their motility, viability, and antimicrobial activity [36, 75, 92]. To date, there is no published report on the expression of the Indian cattle BBD129 gene in the male reproductive tract. The expression pattern of BBD129 mRNA in Indian cattle MRT show higher expression in the corpus-epididymis region with a region-specific expression as found in other mammals [23, 81, 91, 93]. Interestingly, expression of the BBD129 gene initiates from the rete testes in Indian cattle and had maximum expression in the middle corpus region of the epididymis, thereafter the expression decreases suggests that as the sperm acquire its BBD129 coat largely in the corpus region of the epididymis. The MRT expression pattern of the BBD129 gene suggests its region-specific roles in sperm maturation, protection and makes sperm fertilizable. The cross-bred cattle-yak epididymal transcriptomics study reported an association of BBD129 gene down-expression with male sterility [94]. It is a well-understood phenomenon that sperm epididymal coat proteins follow the last-in, first-out sequence in the female reproductive tract in order to execute proper sperm

function. The current finding of intense expression of cattle BBD129 gene in healthy matured MRT tissues clearly indicates the region-specific abundance of BBD129 protein destined for uptake on the sperm surface.

The emerging bio-computational analyses have shortened the issues of expensive in-vitro experimentations and high throughput technologies cost. In this study, bioinformatic analyses underpinned the possible causes of nsSNPs which could alter the structure and functions of BBD129 protein. The polymorphisms could lead to abrupt mRNA synthesis and conformational distortion of protein or non-functional proteins. The bovine BNBD-4 SNPs (CT2239) affect gene functioning by incorrect intron splicing results in a premature stop codon and unstable/aberrant mRNA production [95]. The non-synonymous genomic variation (C483A, rs378652941) in the TMEM95 gene introduces a premature stop codon that leads to mRNA decay and also, reported that this altered codon resides within the transmembrane domain of TMEM95 most likely resulting in a disturbed anchorage of the truncated protein on the sperm plasma membrane [96]. In this study, the mountain curve plot analysis of minimum free energy, centroid secondary structure, partition function, thermodynamic ensemble, and entropy is an easy view of the effects of nsSNP on BBD129 mRNA. The possible impact of observed BBD129 polymorphism on mRNA found that they are making more complex BBD129 mRNA secondary structure by altering the base pairing which could affect the transcription or translation processes and may affect BBD129 protein abundance.

The mutation Q47R in PATE1 protein causes structural damages in protein leads to less bio-availability to the sperm membrane and results in reduced sperm motility associated with phosphorylations. [97–99]. Bovine SPAG11 gene polymorphisms are associated with higher ejaculate volumes, higher sperm concentrations, higher fresh sperm motility, and higher post-thaw cryopreserved sperm motility [100]. The serine-to-alanine S191A in progesterone receptors causes de-phosphorylation leads to reduced female fertility (litters size) and uterine growth [101–102]. The bioinformatic servers SIFT, SNAP, SNPs & GO, and MAPP analyses revealed amino acid S57A substitution as disease-causing and amino acid N110S substitution as neutral suggesting possible reason behind lower observations as single mutated BBD129 haplotypes in the Indian cattle genome. While Polyphen, PROVEN, PredictSNP, and PhD-SNP analyses revealed both observed linked nsSNPs (S57A & N110S together) as non-deleterious or non-disease-causing effects in nature and that could be a reason for still persistent double mutated BBD129 haplotypes in the Indian cattle genome. I-Mutant 2.0 suggested both the substitutions as protein structural distortion or protein stability decreasing mutations. ProtParam, Argot 2, MUpro, SOPMA, and PSI-PRED analyses found that nsSNPs disrupting the protein stability and altering protein physicochemical properties of BBD129 lead to the compromised biological functions. The substitutions of S57A and N110S in functional coding regions of the BBD129 protein found possible damaging effects on the hydrophobicity, net positive charge, aliphatic contents, molecular weight, and causing protein instability.

Capacitation and acrosome reactions are associated with increased serine, tyrosine, and threonine amino-acids phosphorylations and kinase activities [103–104]. The epididymal secretions and seminal plasma proteins are potent de-capacitation factors that maintain the premature capacitation and acrosomal integrity of the sperm in the epididymis and FRT before the conference with the zona pellucida membrane [104–106]. The polymorphism in BBD129 increases threonine and serine phosphorylations which may trigger the sperm hyperactive motility, early capacitation, early acrosomal reaction or may attract immune cells [107–108]. The adsorption of primate DEFB126 on the sperm surface provides a thick glycocalyx which facilitates numbers of sperm functions and polymorphisms in the DEFB126 cause abnormal glycosylations leads to sub-fertility [10–11, 34, 71]. In buffalo, the abundance of glycans on the spermatozoa from distinct fertility bull affects the neutrophil phagocytosis and NETosis process [109]. Similarly in our study, glycosylation prediction revealed nsSNPs causing positional alterations in the O-glycosylations pattern on BBD129 protein. All prediction results suggest snSNPs could result in a substantial change in functional attributes of BBD129 protein or sperm functioning [11, 22, 34] and their abundance in the group of low fertilizing bulls could be a possible marker for bovine fertility.

Conclusion

The present study clearly suggests a close association between polymorphisms of the BBD129 gene with bull's conception rate. Maximum expression of the BBD129 gene in the corpus epididymis region might, however, be of interest for its uptake on the sperm surface. Full-length of BBD129 in Indian cattle was absolute in consonant with *Bos taurus* sequence reflecting the conserved functionality of the protein. A detailed insight into the SNPs proves the non-synonymous polymorphisms in the BBD129 gene are convincingly related to the low conception rate of Indian cattle bulls. Further, the change in functional characteristics of BBD129 protein due to SNPs indicating the altered structural and conformational topology of BBD129 protein could be responsible for the distinct fertilizing potential of sperm. However, further study on large sample size and in-vitro validation assay needs to be taken up

for assessing the altered fertilization of sperm and other reproductive pleiotropic functions that emerged due to SNPs in BBD129 of Indian cattle.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests

Funding

BMGF project, Project code [INV-008501 OPP1154401](#)

Authors' contributions

The individual contributions of authors to the manuscript should be specified in this section

SSS has written the manuscript, PK and MP have contributed in cloning, AV has contributed real time PCR analysis, VB has contributed in the bioinformatic analysis, SD, RK and TKD has contributed scientific ideas, reading and editing the manuscript.

Acknowledgements

I gratefully acknowledge Dr. Mukesh Bhakat (Principal Scientist, ABRC, NDRI, Karnal, Haryana, India, 132001), for samples availability.

Authors' information

Given in separate file

Footnotes

No footnotes

We certify that the current project carries approval of the Institutional Animal Ethics Committee (IAEC) of National Dairy Research Institute, Karnal India.

Abbreviations

RLM-RACE= RNA ligase mediated Rapid amplification of cDNA end

RT-qPCR= Reverse Transcription quantitative Polymerase chain reaction

BBD129= Bovine Beta-defensin 129

PCR=Polymerase Chain reaction

SNP=Single Nucleotide polymorphism

UTR= Untranslated region

DEFB/BD = Beta Defensin

FRT= Female reproductive Tract

OEC= Oviduct epithelial cell

CCR=chemokine receptor

PBS=Phosphate buffer saline

NTC=Non-template control

DTT=Diethyl pyrocarbonate

MRT= Male reproductive Tract

dNTP=dioxy-nucleotide triphosphate

LB= Lysogeny broth

CA-BD= Class-A Beta-Defensin

CR=Conception rate

PCI= Phenol:Chloroform:Isoamylalcohol

CNV= Copy number variation

PROVEN=Protein Variation Effect Analyze

SIFT= Sorting Intolerant from Tolerant

Polyphen-I= Polymorphism Phenotyping-I

MAPP= Multivariate analysis of protein polymorphisms

MSA= Multiple sequence alignment

CSN2=Casein beta

BNBD= Bovine neutrophil beta-defensin

BBD= Bovine beta defensin

HF= High fertile

LF= Low fertile

References

1. Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nature medicine*. 2008 Nov;14(11):1197-213.
2. Tollner TL, Bevins CL, Cherr GN. Multifunctional glycoprotein DEFB126—a curious story of defensin-clad spermatozoa. *Nature Reviews Urology*. 2012 Jul;9(7):365-75.
3. Fujihara Y, Noda T, Kobayashi K, Oji A, Kobayashi S, Matsumura T, Larasati T, Oura S, Kojima-Kita K, Yu Z, Matzuk MM. Identification of multiple male reproductive tract-specific proteins that regulate sperm migration through the oviduct in mice. *Proceedings of the National Academy of Sciences*. 2019 Sep 10;116(37):18498-506.

4. Oehninger S, Franken DR, Ombelet W. Sperm functional tests. *Fertility and sterility*. 2014 Dec 1;102(6):1528-33.
5. Han Y, Peñagaricano F. Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC genetics*. 2016 Dec;17(1):1-1.
6. Nani JP, Rezende FM, Peñagaricano F. Predicting male fertility in dairy cattle using markers with large effect and functional annotation data. *BMC genomics*. 2019 Dec;20(1):1-0.
7. Butler ML, Bormann JM, Weaber RL, Grieger DM, Rolf MM. Selection for bull fertility: a review. *Translational Animal Science*. 2020 Jan;4(1):423-41.
8. Schröter, S., Osterhoff, C., McArdle, W., & Ivell, R. (1999). The glycocalyx of the sperm surface. *Human reproduction update*, 5(4), 302-313.
9. Rickard JP, Pool KR, Druart X, de Graaf SP. The fate of spermatozoa in the female reproductive tract: a comparative review. *Theriogenology*. 2019 Oct 1;137:104-12.
10. Tollner TL, Yudin AI, Tarantal AF, Treece CA, Overstreet JW, Cherr GN. Beta-defensin 126 on the surface of macaque sperm mediates attachment of sperm to oviductal epithelia. *Biology of reproduction*. 2008 Mar 1;78(3):400-12..
11. Tollner TL, Yudin AI, Treece CA, Overstreet JW, Cherr GN. Macaque sperm coating protein DEFB126 facilitates sperm penetration of cervical mucus. *Human reproduction*. 2008 Nov 1;23(11):2523-34.
12. Narciandi F, Fernandez-Fuertes B, Khairulzaman I, Jahns H, King D, Finlay EK, Mok KH, Fair S, Lonergan P, Farrelly CO, Meade KG. Sperm-coating beta-defensin 126 is a dissociation-resistant dimer produced by epididymal epithelium in the bovine reproductive tract. *Biology of Reproduction*. 2016 Dec 1;95(6):121-.
13. Cannarella R, Condorelli RA, Mongioi LM, La Vignera S, Calogero AE. Molecular biology of spermatogenesis: novel targets of apparently idiopathic male infertility. *International journal of molecular sciences*. 2020 Jan;21(5):1728.
14. Meade KG, Cormican P, Narciandi F, Lloyd A, O'Farrelly C. Bovine β -defensin gene family: opportunities to improve animal health?. *Physiological genomics*. 2014 Jan 1;46(1):17-28.
15. Zhang ZY, Zhang HM, Li DS, Xiong TY, Fang SG. Characterization of the β -defensin genes in giant panda. *Scientific reports*. 2018 Aug 17;8(1):1-4.
16. Solanki SS, Singh P, Kashyap P, Sansi MS, Ali SA. Promising role of defensins peptides as therapeutics to combat against viral infection. *Microbial Pathogenesis*. 2021 Apr 29;104930.
17. Selsted ME, Tang YQ, Morris WL, McGuire PA, Novotny MJ, Smith W, Henschen AH, Cullor JS. Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils. *Journal of Biological Chemistry*. 1993 Mar 25;268(9):6641-8.
18. Hall TJ, McQuillan C, Finlay EK, O'Farrelly C, Fair S, Meade KG. Comparative genomic identification and validation of β -defensin genes in the *Ovis aries* genome. *BMC genomics*. 2017 Dec;18(1):1-1.
19. Bagnicka E, Strzałkowska N, Jóźwik A, Krzyżewski J, Horbańczuk J, Zwierzchowski L. Expression and polymorphism of defensins in farm animals. *Acta Biochimica Polonica*. 2010 Dec 6;57(4).
20. Morrison GM, Semple CA, Kilanowski FM, Hill RE, Dorin JR. Signal sequence conservation and mature peptide divergence within subgroups of the murine β -defensin gene family. *Molecular Biology and Evolution*. 2003 Mar 1;20(3):460-70.
21. Beckloff N, Diamond G. Computational analysis suggests beta-defensins are processed to mature peptides by signal peptidase. *Protein and peptide letters*. 2008 Jun 1;15(5):536-40.
22. Narciandi F, Lloyd A, Meade KG, O'Farrelly C. A novel subclass of bovine β -defensins links reproduction and immunology. *Reproduction, Fertility and Development*. 2014 Aug 7;26(6):769-77.
23. Batra V, Maheshwarappa A, Dagar K, Kumar S, Soni A, Kumaresan A, Kumar R, Datta TK. Unusual interplay of contrasting selective pressures on β -defensin genes implicated in male fertility of the Buffalo (*Bubalus bubalis*). *BMC evolutionary biology*. 2019 Dec;19(1):1-9.
24. Tollner TL, Venners SA, Hollox EJ, Yudin AI, Liu X, Tang G, Xing H, Kays RJ, Lau T, Overstreet JW, Xu X. A Common Mutation in DEFB126 Causes Impaired Sperm Function and Subfertility. *Science Translational Medicine*. 2011 July;3(92):92ra65
25. Tollner TL, Yudin AI, Treece CA, Overstreet JW, Cherr GN. Macaque sperm release ESP13. 2 and PSP94 during capacitation: The absence of ESP13. 2 is linked to sperm-zona recognition and binding. *Molecular Reproduction and Development: Incorporating Gamete Research*. 2004 Nov;69(3):325-37.

26. Jia HP, Mills JN, Barahmand-Pour F, Nishimura D, Mallampali RK, Wang G, Wiles K, Tack BF, Bevins CL, McCray Jr PB. Molecular cloning and characterization of rat genes encoding homologues of human β -defensins. *Infection and immunity*. 1999 Sep 1;67(9):4827-33.
27. Johnston DS, Turner TT, Finger JN, Owtscharuk TL, Kopf GS, Jelinsky SA. Identification of epididymis-specific transcripts in the mouse and rat by transcriptional profiling. *Asian journal of andrology*. 2007 Jul;9(4):522-7.
28. Luenser K, Ludwig A. Variability and evolution of bovine β -defensin genes. *Genes & Immunity*. 2005 Mar;6(2):115-22.
29. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012 Oct;7(10):e46688.
30. Johnson GP, Lloyd AT, O'Farrelly C, Meade KG, Fair S. Comparative genomic identification and expression profiling of a novel β -defensin gene cluster in the equine reproductive tract. *Reproduction, Fertility and Development*. 2016 Sep 8;28(10):1499-508.
31. Bagnicka E, Strzałkowska N, Flisikowski K, Szreder T, Jóźwik A, Prusak B, Krzyżewski J, Zwierzchowski L. The polymorphism in the β 4-defensin gene and its association with production and somatic cell count in Holstein-Friesian cows. *Journal of Animal Breeding and Genetics*. 2007 Jun;124(3):150-6.
32. Bagnicka E, Strzałkowska N, Szreder T, Prusak B, Jóźwik A, Kościuczuk E, Krzyżewski J, Zwierzchowski L. A/C polymorphism in the β -4 defensin gene and its association with phenotypic and breeding values of milk production traits in Polish-Friesian cows. *Animal Science Papers and Reports*. 2008;26(4):239-50.
33. Duan S, Shi C, Chen G, Zheng JF, Wu B, Diao H, Ji L, Gu Y, Xin A, Wu Y, Zhou W. Another functional frame-shift polymorphism of DEFB 126 (rs11467497) associated with male infertility. *Journal of cellular and molecular medicine*. 2015 May;19(5):1077-84.
34. Yudin AI, Treece CA, Tollner TL, Overstreet JW, Cherr GN. The carbohydrate structure of DEFB126, the major component of the cynomolgus Macaque sperm plasma membrane glycocalyx. *The Journal of membrane biology*. 2005 Oct;207(3):119-29.
35. Sangeeta K, Yenugu S. Male reproductive tract antimicrobial expression in the extremes of ages of rats. *Gene*. 2019 Aug 20;710:218-32.
36. Diao R, Fok KL, Chen H, Yu MK, Duan Y, Chung CM, Li Z, Wu H, Li Z, Zhang H, Ji Z. Deficient human β -defensin 1 underlies male infertility associated with poor sperm motility and genital tract infection. *Science translational medicine*. 2014 Aug 13;6(249):249ra108-.
37. Zupin L, Polesello V, Martinelli M, Luppi S, Giolo E, Zito G, Romano F, Segat L, Crovella S, Ricci G. Human β -defensin 1 in follicular fluid and semen: impact on fertility. *Journal of assisted reproduction and genetics*. 2019 Apr;36(4):787-97.
38. Diao R, Cai X, Liu L, Yang L, Duan Y, Cai Z, Gui Y, Mou L. In vitro chemokine (CC motif) receptor 6-dependent non-inflammatory chemotaxis during spermatogenesis. *Biological research*. 2018;51.
39. NCBI-BLAST primer designing tool. <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>. Accessed 28 August 2018.
40. Oligocalculator to check primer's parameters. <http://biotools.nubic.northwestern.edu/OligoCalc.html>. Accessed 28 August 2018.
41. Ferrè F, Clote P. DiANNA: a web server for disulfide connectivity prediction. *Nucleic acids research*. 2005 Jul 1;33(suppl_2):W230-2.
42. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics*. 1995 Dec 1;11(6):681-4.
43. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics*. 2000 Apr 1;16(4):404-5.
44. Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *Journal of molecular biology*. 2004 Jul 16;340(4):783-95.
45. Falda M, Toppo S, Pescarolo A, Lavezzo E, Di Camillo B, Facchinetti A, Cilia E, Velasco R, Fontana P. Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC bioinformatics*. 2012 Dec;13(4):1-9.
46. GeneBank nucleotide submission of complete coding BBD129 gene. <https://www.ncbi.nlm.nih.gov/WebSub/?form=history&tool=genbank>. Accessed 12 April 2021.
47. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in bioinformatics*. 2008 Jul 1;9(4):299-306.
48. Zhang BW, Li M, Ma LC, Wei FW. A widely applicable protocol for DNA isolation from fecal samples. *Biochemical Genetics*. 2006 Dec 1;44(11-12):494.

49. Hruz T, Wyss M, Docquier M, Pfaffl MW, Masanetz S, Borghi L, Verbrugghe P, Kalaydjieva L, Bleuler S, Laule O, Descombes P. RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC genomics*. 2011 Dec;12(1):1-4.
50. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular detection and quantification*. 2015 Mar 1;3:9-16.
51. BLAST BBD129 sequencing result. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>. Accessed 30 January 2021.
52. Hofacker IL. R NA Secondary Structure Analysis Using the Vienna RNA Package. *Current protocols in bioinformatics*. 2009 Jun;26(1):12-2.
53. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015 Aug 15;31(16):2745-7.
54. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*. 2003 Jul 1;31(13):3812-4.
55. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC genomics*. 2013 May;14(3):1-7.
56. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006 Nov 15;22(22):2729-34.
57. Mohammad T, Amir M, Prasad K, Batra S, Kumar V, Hussain A, Rehman MT, AlAjmi MF, Hassan MI. Impact of amino acid substitution in the kinase domain of Bruton tyrosine kinase and its association with X-linked agammaglobulinemia. *International Journal of Biological Macromolecules*. 2020 Dec 1;164:2399-408.
58. Bendl J, Musil M, Štourač J, Zendulka J, Damborský J, Brezovský J. PredictSNP2: a unified platform for accurately evaluating SNP effects by exploiting the different characteristics of variants in distinct genomic regions. *PLoS computational biology*. 2016 May 25;12(5):e1004962.
59. Rotimi SO, Peter O, Oguntade O, Rotimi OA. In silico analysis of the functional non-synonymous single nucleotide polymorphisms in the human CYP27B1 gene. *Egyptian Journal of Medical Human Genetics*. 2018;19(4):367-78.
60. Chao EC, Velasquez JL, Witherspoon MS, Rozek LS, Peel D, Ng P, Gruber SB, Watson P, Rennert G, Anton-Culver H, Lynch H. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms–mismatch repair (MAPP-MMR). *Human mutation*. 2008 Jun;29(6):852-60.
61. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*. 2005:571-607.
62. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics*. 2006 Mar 1;62(4):1125-32.
63. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO journal*. 2013 May 15;32(10):1478-88.
64. Chauhan JS, Rao A, Raghava GP. In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. *PloS one*. 2013 Jun 28;8(6):e67008.
65. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *InPac Symp Biocomput* 2001 Dec 12 (Vol. 7, pp. 310-22).
66. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology*. 1999 Dec 17;294(5):1351-62.
67. GraphPad Prism 5.0. www.graphpad.com. Accessed 7 January 2021.
68. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta CT$ method. *methods*. 2001 Dec 1;25(4):402-8.
69. Yanagimachi R. Mechanisms of fertilization in mammals. *In Fertilization and embryonic development in vitro* 1981 (pp. 81-182). Springer, Boston, MA.
70. Tecle E, Gagneux P. Sugar-coated sperm: unraveling the functions of the mammalian sperm glycocalyx. *Molecular reproduction and development*. 2015 Sep;82(9):635-50.

71. Yudin AI, Generao SE, Tollner TL, Treece CA, Overstreet JW, Cherr GN. Beta-defensin 126 on the cell surface protects sperm from immunorecognition and binding of anti-sperm antibodies. *Biology of reproduction*. 2005 Dec 1;73(6):1243-52.
72. Dorin JR. Novel phenotype of mouse spermatozoa following deletion of nine β -defensin genes. *Asian journal of andrology*. 2015 Sep;17(5):716.
73. Kalita DJ, Barkakati J, Rajkhowa D, Hazarika M and Kumar A. Molecular characterization of exon-2 of buffalo mammary defensin for exploring its potency for synthesis of novel antimicrobial agents. *Int J Chem Stud*. 2019;7(3), 1868-1871.
74. Fernandez-Fuertes B, Narciandi F, O'Farrelly C, Kelly AK, Fair S, Meade KG, Lonergan P. Cauda epididymis-specific beta-defensin 126 promotes sperm motility but not fertilizing ability in cattle. *Biology of reproduction*. 2016 Dec 1;95(6):122-.
75. Khayamabed R, Tavalaei M, Taherian SS, Nasr-Esfahani MH. Effect of recombinant β -defensin 1 protein on human sperm motility and viability. *Andrologia*. 2020 Feb;52(1):e13455.
76. NCBI Bos taurus DEFB129 gene. https://www.ncbi.nlm.nih.gov/nuccore/XM_005215086.4. Accessed 16 July 2021.
77. Watanabe T, Sadamoto H, Aonuma H. Molecular basis of the dopaminergic system in the cricket *Gryllus bimaculatus*. *Invertebrate neuroscience*. 2013 Dec 1;13(2):107-23.
78. Xu XF, Li J, Cao YX, Chen DW, Zhang ZG, He XJ, Ji DM, Chen BL. Differential expression of long noncoding RNAs in human cumulus cells related to embryo developmental potential: a microarray analysis. *Reproductive Sciences*. 2015 Jun;22(6):672-8.
79. Liu X, Gorovsky MA. Mapping the 5' and 3' ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic acids research*. 1993 Oct 25;21(21):4954-60.
80. Dholpuria S, Kumar M, Kumar S, Sarwalia P, Rajput S, Kumar R, De S, Datta TK. Differential expression of newly identified long intergenic non-coding RNAs in buffalo oocytes indicating their possible role in maturation and embryonic development. *Journal of cellular biochemistry*. 2017 Jul;118(7):1712-21.
81. Narciandi F, Lloyd AT, Chapwanya A, O'Farrelly C, Meade KG. Reproductive tissue-specific expression profiling and genetic variation across a 19 gene bovine β -defensin cluster. *Immunogenetics*. 2011 Oct;63(10):641-51.
82. Chen Y, Luo Y, Carleton H, Timme R, Melka D, Muruvanda T, Wang C, Kastanis G, Katz LS, Turner L, Fritzinger A. Whole genome and core genome multilocus sequence typing and single nucleotide polymorphism analyses of *Listeria monocytogenes* isolates associated with an outbreak linked to cheese, United States, 2013. *Applied and Environmental Microbiology*. 2017 May 26;83(15):e00633-17.
83. Ryan LK, Rhodes J, Bhat M, Diamond G. Expression of β -defensin genes in bovine alveolar macrophages. *Infection and immunity*. 1998 Feb 1;66(2):878-81.
84. Revay T, Quach AT, Maignel L, Sullivan B, King WA. Copy number variations in high and low fertility breeding boars. *BMC genomics*. 2015 Dec;16(1):1-2.
85. Candille SI, Kaelin CB, Cattanaach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS. A β -defensin mutation causes black coat color in domestic dogs. *Science*. 2007 Nov 30;318(5855):1418-23.
86. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinisch W, Teml A, Schwab M, Lichter P, Radlwimmer B. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *The American Journal of Human Genetics*. 2006 Sep 1;79(3):439-48.
87. UCSC genome browser. <https://genome.ucsc.edu>. Accessed 18 February 2021
88. Hollox EJ, Davies J, Griesenbach U, Burgess J, Alton EW, Armour JA. Beta-defensin genomic copy number is not a modifier locus for cystic fibrosis. *Journal of negative results in biomedicine*. 2005 Dec;4(1):1-5.
89. Gurao A, Kashyap SK, Singh R. β -defensins: An innate defense for bovine mastitis. *Veterinary world*. 2017 Aug;10(8):990.
90. Jelinsky SA, Turner TT, Bang HJ, Finger JN, Solarz MK, Wilson E, Brown EL, Kopf GS, Johnston DS. The rat epididymal transcriptome: comparison of segmental gene expression in the rat and mouse epididymides. *Biology of reproduction*. 2007 Apr 1;76(4):561-70.
91. Ribeiro CM, Silva EJ, Hinton BT, Avellar MC. β -defensins and the epididymis: contrasting influences of prenatal, postnatal, and adult scenarios. *Asian journal of andrology*. 2016 Mar;18(2):323.
92. Barabas N, Röhr J, Holler E, Hehlhans T. Beta-defensins activate macrophages and synergize in pro-inflammatory cytokine expression induced by TLR ligands. *Immunobiology*. 2013 Jul 1;218(7):1005-11.

93. Tu J, Li D, Li Q, Zhang L, Zhu Q, Gaur U, Fan X, Xu H, Yao Y, Zhao X, Yang M. Molecular evolutionary analysis of β -defensin peptides in vertebrates. *Evolutionary Bioinformatics*. 2015 Jan;11:EBO-S25580.
94. Zhao W, Ahmed S, Ahmed S, Yangliu Y, Wang H, Cai X. Analysis of long non-coding RNAs in epididymis of cattleyak associated with male infertility. *Theriogenology*. 2021 Jan 15;160:61-71.
95. Goroohi Z, Sharifiyazdi H, Mirzaei A. Association between beta defensin gene polymorphism and clinical endometritis in dairy cows. *Comparative Clinical Pathology*. 2019 Apr;28(2):377-82.
96. Pausch H, Kölle S, Wurmser C, Schwarzenbacher H, Emmerling R, Jansen S, Trottmann M, Fuerst C, Götz KU, Fries R. A nonsense mutation in TMEM95 encoding a nondescript transmembrane protein causes idiopathic male subfertility in cattle. *PLoS genetics*. 2014 Jan 2;10(1):e1004044.
97. Soler-García AA, Maitra R, Kumar V, Ise T, Nagata S, Beers R, Bera TK, Pastan I. The PATE gene is expressed in the accessory tissues of the human male genital tract and encodes a secreted sperm-associated protein. *Reproduction*. 2005 Apr 1;129(4):515-24.
98. Chan CC, Shui HA, Wu CH, Wang CY, Sun GH, Chen HM, Wu GJ. Motility and protein phosphorylation in healthy and asthenozoospermic sperm. *Journal of proteome research*. 2009 Nov 6;8(11):5382-6.
99. Zhang S, Wang QM, Ding XP, Wang T, Mu XM, Chen ZY. Association of polymorphisms in PATE1 gene with idiopathic asthenozoospermia in Sichuan, China. *Journal of reproductive immunology*. 2016 Nov 1;118:54-60.
100. Liu X, Ju Z, Wang L, Zhang Y, Huang J, Li Q, Li J, Zhong J, An L, Wang C. Six novel single-nucleotide polymorphisms in SPAG11 gene and their association with sperm quality traits in Chinese Holstein bulls. *Animal reproduction science*. 2011 Nov 1;129(1-2):14-21.
101. Lydon JP, DeMayo FJ, Funk CR, Mani SK, Hughes AR, Montgomery CA, Shyamala G, Conneely OM, O'Malley BW. Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. *Genes & development*. 1995 Sep 15;9(18):2266-78.
102. Grimm SL, Ward RD, Obr AE, Franco HL, Fernandez-Valdivia R, Kim JS, Roberts JM, Jeong JW, DeMayo FJ, Lydon JP, Edwards DP. A role for site-specific phosphorylation of mouse progesterone receptor at serine 191 in vivo. *Molecular Endocrinology*. 2014 Dec 1;28(12):2025-37.
103. Bravo MM, Aparicio IM, Garcia-Herreros M, Gil MC, Pena FJ, Garcia-Marin LJ. Changes in tyrosine phosphorylation associated with true capacitation and capacitation-like state in boar spermatozoa. *Molecular Reproduction and Development: Incorporating Gamete Research*. 2005 May;71(1):88-96..
104. Grasa P, Colas C, Gallego M, Monteagudo L, Muino-Blanco T, Cebrián-Pérez JA. Changes in content and localization of proteins phosphorylated at tyrosine, serine and threonine residues during ram sperm capacitation and acrosome reaction. *Reproduction*. 2009 Apr 1;137(4):655.
105. Nixon B, MacIntyre DA, Mitchell LA, Gibbs GM, O'Bryan M, Aitken RJ. The identification of mouse sperm-surface-associated proteins and characterization of their ability to act as decapacitation factors. *Biology of Reproduction*. 2006 Feb 1;74(2):275-87.
106. Khatun M, Kaur S, Kanchan CS. Subfertility problems leading to disposal of breeding bulls. *Asian-Australasian journal of animal sciences*. 2013 Mar;26(3):303.
107. Caballero-Campo P, Buffone MG, Benencia F, Conejo-García JR, Rinaudo PF, Gerton GL. A role for the chemokine receptor CCR6 in mammalian sperm motility and chemotaxis. *Journal of cellular physiology*. 2014 Jan;229(1):68-78.
108. Duan YG, Wehry UP, Buhren BA, Schruppf H, Oláh P, Bünnemann E, Yu CF, Chen SJ, Müller A, Hirchenhain J, Lierop A. CCL20-CCR6 axis directs sperm–oocyte interaction and its dysregulation correlates/associates with male infertility. *Biology of reproduction*. 2020 Aug 21;103(3):630-42.
109. Batra V, Dagar K, Nayak S, Kumaresan A, Kumar R, Datta TK. A higher abundance of O-linked glycans confers a selective advantage to high fertile buffalo spermatozoa for immune-evasion from neutrophils. *Frontiers in immunology*. 2020 Aug 28;11:1928.

Tables

Table 1: The details of basis of selection of cattle bulls, number of clones sequenced and distribution of SNP in distinct fertility bulls

Cattle Bulls No.	Conception rate (%)	BBD129 clones	BBD129_W Haplotype (169T/329A)	rs378737321 & rs383285978 Haplotype (169G/329G)	rs378737321 Haplotype (169G)	rs383285978 Haplotype (329G)	Gene Copy/ genome	Remark
5788HF	50.75 %	18	72.22%	11%	0%	1.6%	1	High fertile bulls Number of clones sequenced = 105 TA haplotype = 71.42 % GG haplotype = 24.76 % GA haplotype = 1.90 % TG haplotype = 2.8 %
6555HF	55.13 %	23	65%	34%	0%	0%	1	
4873HF	54.65 %	17	82.3%	23.5%	0%	0%	1	
6479HF	57.30 %	25	64%	24%	12%	0%	1	
6251HF	57.81 %	22	77%	23%	0%	0%	1	
7205LF	29.52 %	22	13.63%	86.36%	0%	0%	1	Low fertile bulls Number of clones sequenced = 149 TA haplotype = 22.33 GG haplotype = 75.72 TG haplotype = 1.94
7141LF	30.17 %	22	45%	45%	0%	9%	1	
6568LF	28.36 %	21	38%	61%	0%	0%	1	
5450LF	27.19 %	19	0%	100%	0%	0%	1	
7063LF	31.79 %	19	10.52%	89.48%	0%	0%	1	
7178LF	30.85%	46	67.39%	28.26%	0%	4.34%	1	

Table 2: Bioinformatics tools used for prediction of impact of SNPs on BBD129 protein

Tools	Mutation S57A (rs378737321) effect	Mutation N110S (rs383285978) effect
PhD-SNP	Neutral 0.118	Neutral 0.031
SIFT	Disease 0.030	Neutral 1.000
SNAP	Disease 0.515	Neutral 0.305
Meta-SNP	Neutral 0.163	Neutral 0.037
I-Mutant2.0	Stability decrease	Stability decrease
SNPs & GO	Neutral 0.118	Neutral 0.031
PredictSNP	Neutral 66%	Neutral 83%
MAPP	Disease 66%	Neutral 85%
Polyphen-I	Neutral 67%	Neutral 67%
Polyphen-2	Neutral 63%	Neutral 72%
PROVEN	Neutral -7.75	Neutral -0.024

Table 3: List of comparison of biological function between native BBD129 and double mutated BBD129 proteins predicted by Argot 2.5 server

GO ID	Name	Information Content	BBD129 TA haplotype Score	Double mutated BBD129 GG haplotype Score	Difference
GO:0005576	Extracellular region	6.5520748	5664.02	5617.75	- 46.27
GO:0009986	Cell surface	9.67	3005.11	2963.85	- 41.26
GO:0006952	Defense response	5.707	2774.83	2750.72	- 24.11
GO:0045087	Innate immune response	7.54	2242.19	2216.76	- 25.43
GO:0042742	Defense response to bacterium	11.843	3521.8	3481.86	- 39.94

Table 4: List of bioinformatics tools used for the prediction of SNPs effects on different possible functions or physiochemical properties of BBD129 protein

Tools	What it Predicts	Our Observation	Biological logic by which it supports our hypothesis
MUpro tool	Prediction of Protein Stability Changes for Single-Site Mutations from Sequence	Decrease in protein stability & $\Delta \Delta G$ energy	Disturbance in functional structure formation (less availability, reduced antibacterial activity)
NetPhos tool	Predicts serine, threonine or tyrosine phosphorylation sites	Increase and frame shift in phosphorylation	Early induction of acrosome reactions i.e. Phagocytosis invitation
NetOGlyc 4	O-glycosylation predictions	Frame shift	Alteration in glycosylation Pattern (problem in cumulus penetration - infertility)
Argot2	Function prediction (gene ontology)	Decrease in score (same GO ID hit)	Decrease in their desired functions (mal-functioning, reduced antibacterial activity)
PSI-PRED	Predicts secondary structure, disorder, and phi/psi dihedral angles of amino acids	Changes in secondary structure, disorder	Disturbance in formation of functional protein structure (mis-folding less availability)
ProtParam	Instability index, thermostability, no. aa, MW, PI, charge, atoms, hydrophobicity, % aa, ext. coefficient	Decrease = atoms, mw	Influence antibacterial activity
		Decrease = Instability & thermostability	
O-Gly GlycoEP	O-linked glycosylation	Frame shift	Alteration in glycosylation Pattern
NetNGlyc	N-linked glycosylation	No change	Alteration in glycosylation Pattern
SOPMA	Protein secondary structure prediction	Change in % of alpha helix, beta turn, extended strand, and random coil	Improve 3D structure prediction

Figures

Figure 1

RLM-RACE amplification of BBD129 mRNA: BBD129 complete coding sequence and its genomic location on cattle chromosome 13. BBD129 protein maintains the cysteine pairing rule.

Tree scale: 0.1

Figure 2

Phylogenetic tree analysis: it was performed by the Maximum Likelihood method and Tamura-Nei model by using the MEGA 10 bioinformatics tool. It indicates a strong evolutionary relationship between the RLM-RACE amplified BBD129 gene and the Bovidae family BBD129

Figure 3

gDNA amplification of cattle BB129 gene. A) BBD129 exon 1 gDNA PCR amplification from distinct fertility bulls. B) BBD129 exon 2 gDNA PCR amplification from distinct fertility

Figure 4

Bioinformatic analysis: the location of non-synonymous BBD129 SNPs (Red colour in dotted amino acid), effect of SNPs on BBD129 protein structure and location of post-translational modifications.

Figure 5

Distribution of BBD129 TA haplotype and GG haplotype in distinct fertility cattle bulls.

Figure 6

: Absolute RT-qPCR analysis: distribution of BBD129 gene copy number in distinct fertility cattle bulls.

Figure 7

RT-qPCR expression analysis was done to know the abundance of BBD129 mRNA in different parts of healthy matured cattle bull.

Figure 8

Mountain graph: RNAfold server to predict the possible impact of nsSNPs on mRNA secondary structure. Mountain plot has shown double mutated BBD129 mRNA get more complex than native BBD129 mRNA. SNPs are influencing the base pairing between single-stranded mRNA.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.png](#)
- [SupplementaryFigure2.jpg](#)
- [SupplementaryFigure3.png](#)
- [SupplementaryFigure4.png](#)
- [supplementaryFigure5.png](#)
- [supplementaryFigure6.png](#)
- [SupplementaryFigure7.png](#)
- [SupplementaryFigure8.png](#)
- [supplementaryFigure9.tif](#)
- [supplementaryFigure10.png](#)
- [supplementaryFile1.docx](#)
- [supplementaryfile2.docx](#)
- [supplementaryfile3.docx](#)
- [supplementaryTable1.docx](#)