



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/225893/>

Version: Published Version

---

**Article:**

Batra, V., Maheshwarappa, A., Dagar, K. et al. (2019) Unusual interplay of contrasting selective pressures on  $\beta$ -defensin genes implicated in male fertility of the Buffalo (*Bubalus bubalis*). *BMC Evolutionary Biology*, 19 (1). 214. ISSN: 1471-2148

<https://doi.org/10.1186/s12862-019-1535-8>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

RESEARCH ARTICLE

Open Access



# Unusual interplay of contrasting selective pressures on $\beta$ -defensin genes implicated in male fertility of the Buffalo (*Bubalus bubalis*)

Vipul Batra<sup>1</sup>, Avinash Maheshwarappa<sup>1</sup>, Komal Dagar<sup>1</sup>, Sandeep Kumar<sup>1</sup>, Apoorva Soni<sup>1</sup>, A. Kumaresan<sup>2</sup>, Rakesh Kumar<sup>1</sup> and T. K. Datta<sup>1\*</sup>

## Abstract

**Background:** The buffalo, despite its superior milk-producing ability, suffers from reproductive limitations that constrain its lifetime productivity. Male sub-fertility, manifested as low conception rates (CRs), is a major concern in buffaloes. The epididymal sperm surface-binding proteins which participate in the sperm surface remodelling (SSR) events affect the survival and performance of the spermatozoa in the female reproductive tract (FRT). A mutation in an epididymal secreted protein, beta-defensin 126 (DEFB-126/BD-126), a class-A beta-defensin (CA-BD), resulted in decreased CRs in human cohorts across the globe. To better understand the role of CA-BDs in buffalo reproduction, this study aimed to identify the BD genes for characterization of the selection pressure(s) acting on them, and to identify the most abundant CA-BD transcript in the buffalo male reproductive tract (MRT) for predicting its reproductive functional significance.

**Results:** Despite the low protein sequence homology with their orthologs, the CA-BDs have maintained the molecular framework and the structural core vital to their biological functions. Their coding-sequences in ruminants revealed evidence of pervasive purifying and episodic diversifying selection pressures. The buffalo CA-BD genes were expressed in the major reproductive and non-reproductive tissues exhibiting spatial variations. The Buffalo BD-129 (BuBD-129) was the most abundant and the longest CA-BD in the distal-MRT segments and was predicted to be heavily O-glycosylated.

**Conclusions:** The maintenance of the structural core, despite the sequence divergence, indicated the conservation of the molecular functions of the CA-BDs. The expression of the buffalo CA-BDs in both the distal-MRT segments and non-reproductive tissues indicate the retention the primordial microbicidal activity, which was also predicted by in silico sequence analyses. However, the observed spatial variations in their expression across the MRT hint at their region-specific roles. Their comparison across mammalian species revealed a pattern in which the various CA-BDs appeared to follow dissimilar evolutionary paths. This pattern appears to maintain only the highly efficacious CA-BD alleles and diversify their functional repertoire in the ruminants. Our preliminary results and analyses indicated that BuBD-129 could be the functional ortholog of the primate DEFB-126. Further studies are warranted to assess its molecular functions to elucidate its role in immunity, reproduction and fertility.

**Keywords:** Reproduction, Buffalo, Male fertility, Expression, Beta-defensin, Orthologs, Paralogs, Purifying and diversifying selection

\* Correspondence: [tirthadatta@gmail.com](mailto:tirthadatta@gmail.com)

<sup>1</sup>Animal Genomics Lab, National Dairy Research Institute, Karnal 132001, India  
Full list of author information is available at the end of the article



## Background

Beta-defensins (BDs), first discovered in *Bos taurus* [1], are highly conserved innate effector molecules present ubiquitously from the prokaryotes to the higher eukaryotes. This class of host defence peptides (HDPs) possesses antimicrobial activities against a variety of microbes ranging from the virus [2] and bacteria [3] to fungi [4]. The distinctive attributes of BDs are their relatively small size, the occurrence of six cysteine residues forming the three characteristic disulfide linkages, the presence of a defensin fold, and abundance of basic residues [5]. The disulfide bonds are formed between particular cysteines, C<sub>1-5</sub>, C<sub>2-4</sub> and C<sub>3-6</sub>, and are required for CCR2-CCR6 (chemokine receptors) mediated chemotactic activities [6] usually displayed by these antimicrobial peptides (AMPs). The BD gene structure is usually comprised of two exons; exon 1 (E1), and exon 2 (E2), encoding for the signal and the mature peptide, respectively [1]. The similarities among the BD orthologs are apparent only at the structural level of their genes and/or their products, but not at the sequence level of their translated gene products. Many of the BDs are constitutively expressed in accordance with their pleiotropic functions [7] while others show 'up-regulation' during infection and in the presence of pathogens [8].

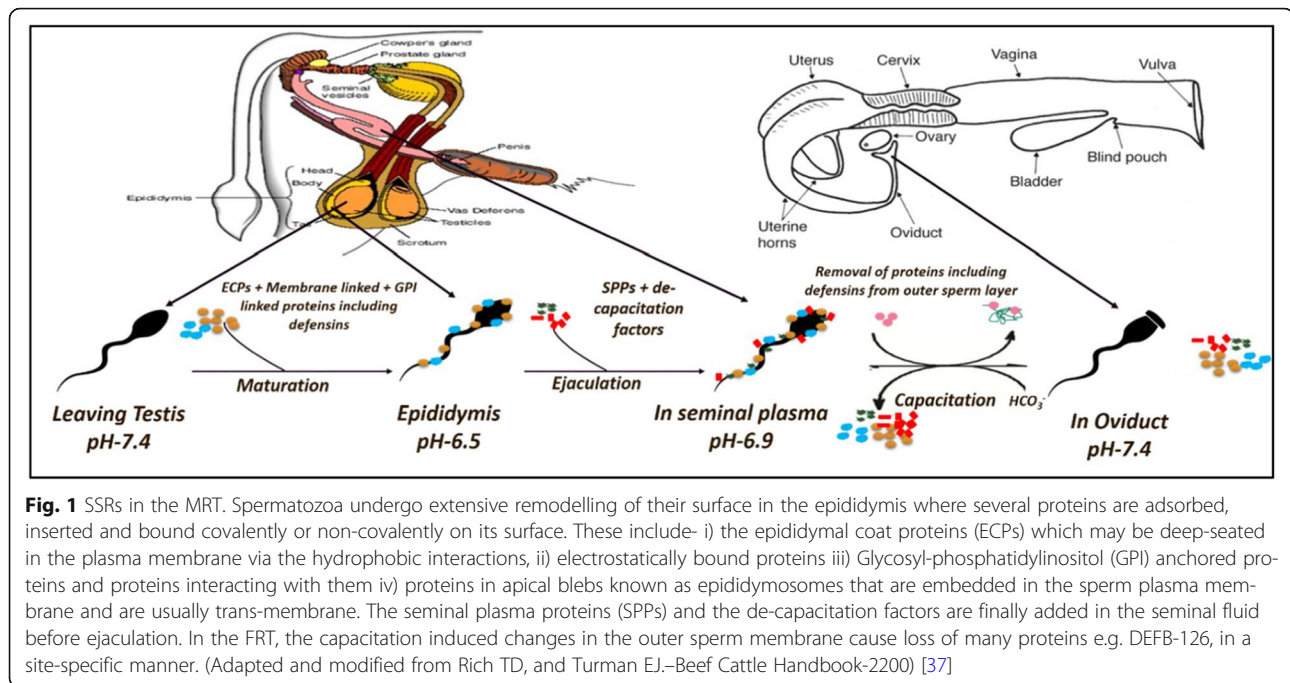
The BDs were hitherto known only as the innate effectors, the AMPs. Nonetheless, their emerging pleiotropic roles in immunomodulation [9], cancer [10], wound healing [11], cell migration [12], angiogenesis [13] and male reproduction [14, 15] are continually being reported. These AMPs are regulated majorly by the androgens and they link the innate and the adaptive immune systems by initiating crosstalk between these two arms of immunity [16]. They also have the ability to modulate the immune response towards pro-inflammatory Th1 or anti-inflammatory Th2 responses by modulating downstream signalling cascades of the chemokine receptors [17–19], and [20]. As previously mentioned, both the constitutive and the inducible expression of BDs has been reported [21, 22], and [23]. Some BDs are subject to positive regulation while others are negatively regulated [24, 25], and [16]. A number of BDs possess positive immunomodulatory properties, whereas others have negative modulatory roles [26, 27]. In addition to the orthologs, the paralogs also exist for BDs [28]. The single nucleotide polymorphism (SNP) genotypes of BDs have been shown to be associated with milk and meat yield [29, 30]. Further, their copy numbers are now well known to affect numerous traits, including biotic resistance, colour, yield etc. [31, 32].

The pleiotropic roles of the BDs in male fertility and reproduction have gained considerable attention since Li et al. [33] reported for the first time about the role of a BD, Bin1b, in reproduction and reproductive tract (MRT) host defence. This BD was found to be present in the proximal-MRT segments where it assists the spermatozoa

in their maturation and storage. In addition, it inhibits the invasion of microorganisms in the neighbouring gamete production site, the testes. Numerous reports have emerged about the preferential and spatiotemporal regulation of BD expression in the MRT [34, 35], and [36]. A number of secretagogues, like Bin1b and DEFB-126, are applied as a peripheral coat onto the sperm surface during its passage through the epididymis. Such a coat assists the spermatozoa in the FRT to ultimately fertilize the egg (Fig. 1).

The secretagogues from the distal segments of the MRT include the products of a specific BD cluster (e.g. hBD-126 on human chromosome 20 and macaque DEFB-126 on chromosome 10), which are secreted and then adsorbed on the surface of the traversing spermatozoa [38]. These specific BDs were named class-A to demarcate them from other BD family members based on their chromosomal cluster, tail lengths, and their roles in male reproductive functions [30]. The functional roles played by these HDPs, especially the CA-BDs, in reproduction and fertility are diverse and equally complex [39, 40]. The CA-BD family members (e.g. DEFB-126) have been demonstrated to render the spermatozoa un-capacitated before they reach the site of fertilization [41]. Further, they assist the spermatozoa to surmount the arduous barriers in the FRT, such as cervical mucus [42], immune response by the PMNs and the macrophages in the uterus and other segments of the FRT [43]. These AMPs are also required for acquiring sperm motility in the epididymis prior to capacitation [14] and for the formation of the oviductal sperm reservoir [44]. The primate CA-BDs have also been demonstrated to assist the spermatozoa to capacitate and to finally attain their ultimate goal of fertilizing the ovum by playing a crucial role in the capacitation process and sperm-zona recognition during fertilization [45, 46]. Molecules with such disparate and critical roles need to be further explored and characterized for unveiling the complete diversity of their pleiotropic functions, especially in the reproductive processes.

The buffalo was considered as a model for this study due to its economic importance in the developing countries. Although it is a premier dairy animal with superior milk-producing ability, it suffers from certain reproductive limitations that constrain its lifetime productivity. Male sub-fertility which is often manifested as low CRs is a major concern in buffalo. Interestingly, a mutation in defensin 126 (DEFB-126) was found to be responsible for decreased CRs in human cohorts across the globe [38]. However, the role of BDs in buffalo reproduction and reproductive tract defence is unknown. Therefore, the objectives of the present work were to characterize the buffalo BD gene clusters and their products with special reference to CA-BDs, to understand the selection pressure(s) on these genes, to identify the most highly expressed CA-BD gene in the buffalo MRT, and to predict its reproductive functional significance.



## Results

### In silico sequence analyses of the buffalo CA-BDs

Thirty BD genes, including 15 novel and the six CA-BDs, were mined from the buffalo reference sequence (RefSeq) assembly (UOA\_WB\_1) based on their size, sequence, gene structure and cluster similarities with that of cattle BD orthologs. Similarly, the BD genes were mined from the RefSeq assemblies of other ruminant, pseudo-ruminant and non-ruminant species (Additional files 1 and 2). The BDs in buffalo were found to be located in four clusters on the chromosome No. 1, 2, 3 and 14. The standardized gene nomenclature was used for naming the identified buffalo BDs as BuBDs on the basis of their orthologous relationship with that of the cattle and the considered mammalian species. The size of the class-A BuBD cluster was found to be approximately 120 kb on the chromosome 14 *vis-a-vis* 320 kb on the chromosome 13 of the cattle [30, 48]. Moreover, all the retrieved BDs exhibited strong evidence of synteny. The six class-A BuBDs which were considered for further analyses were found to be oriented in either 5′-3′ (plus) or 3′-5′ (minus) direction (e.g. BuBD-128). In addition, other classical AMPs like the Lingual Antimicrobial Peptide (LAP), Tracheal Antimicrobial Peptide (TAP), Bovine neutrophil BDs (BNBDs) and Enteric BD (EBD), were also retrieved from the buffalo RefSeq assembly. Further analysis of these classical AMPs was excluded as they are known to be well-characterized as AMPs [47],

The translated gene products of the identified BuBD cluster on chromosome 14 (i.e. CA-BDs) were subject to in silico sequence analyses, which revealed the conservation of the key BD molecular framework characterised

by the six cysteine motif and the hallmark spacing between the cysteines similar to their identified orthologs in other mammalian species (Additional file 3). Interestingly, the multiple sequence alignment (MSA) of class-A BuBDs revealed the presence of unique gene-specific motifs (GSMs), which were either present in all the considered species or only ruminant species (Additional file 3). For example, the ERY motif in BuBD-125 and QVNT motif in BuBD-129 are GSMs present in all the species, whereas SSCIPL motif in BuBD-125 and MMQT motif in BuBD-129 are GSMs limited to the ruminants. The effect of substitution with other amino acids at each site also determined that the first residue of the BD peptide, the six cysteines and their immediate neighbouring residues were the most crucial residues to the function of class-A BuBDs. Thus, the strongest signals (score > 50) were produced for their predicted effect by the SNAP2 tool (Additional file 3). Further, all the class-A BuBDs were found to contain abundant positively-charged amino acids, lysine (K) and arginine (R), and polar residues like threonine (T) and serine (S) (Additional file 3). The loops were found to be the major secondary structural elements of the class-A BuBDs, as predicted by PROFsec. Most of the region, of the class-A BuBD proteins was predicted to be exposed and hence solvent accessible (Additional file 3). Nevertheless, the beta-strands and the alpha helices were also predicted, albeit in lesser proportions. In addition to the above secondary structural elements, the Meta disorder predictor identified the presence of long regions with no regular secondary structure (NORS) i.e. Seventy or more consecutive surface residues without either the beta strands

or the helices in all the class-A BuBDs (Additional file 3). As expected, the putative 3D-structures of their translated gene products were found to contain loops in a major region. The 3D-structures of the modelled buffalo CA-BDs also contained the characteristic three  $\beta$  strands in the anti-parallel beta-sheet arrangement and a flanking partial helix at the N-terminus (Fig. 2). All the modelled class-A BuBDs exhibited a high similarity in their 3D-structures. The BuBDs appear to be secreted in response to either biotic or abiotic stimuli and exhibit binding and regulatory activities with relatively high reliability score as predicted by the Metastudent tool (Additional file 3). The ProfSIS and someNA tools predicted BuBD-125 to be involved in nucleotide binding between the amino-acid residues 32–37, whereas three BuBDs viz. BuBD-125, BuBD-127, BuBD-128 were predicted to be involved in protein-protein interactions between the amino acids 32–40, 32–47 and 28–35, respectively (Additional file 3). Furthermore, all the class-A BuBDs were predicted as AMPs by using a support vector machine (SVM) classifier computational strategy. The primate DEFB-126 and its murine ortholog are heavily glycosylated proteins possessing nine and 13 O-glycosylation sites respectively [30, 49]. O-glycosylation is a crucial post translational modification required for the pleiotropic functions performed by the CA-BDs e.g. cervical mucus penetration (CMP). However, amongst the class-A BuBDs, BuBD-129 was predicted to be heavily glycosylated, possess ten O-glycosylation and three N-glycosylation sites and BuBD-125 was predicted to contain seven O-glycosylation sites (Table 1). Intriguingly BuBD-126, the sequence ortholog of DEFB-126 had only two O-

glycosylation sites and a trans-membrane helix as predicted by the NetOGlyc 4.0 and TMHMM server *ver.* 2.0, respectively.

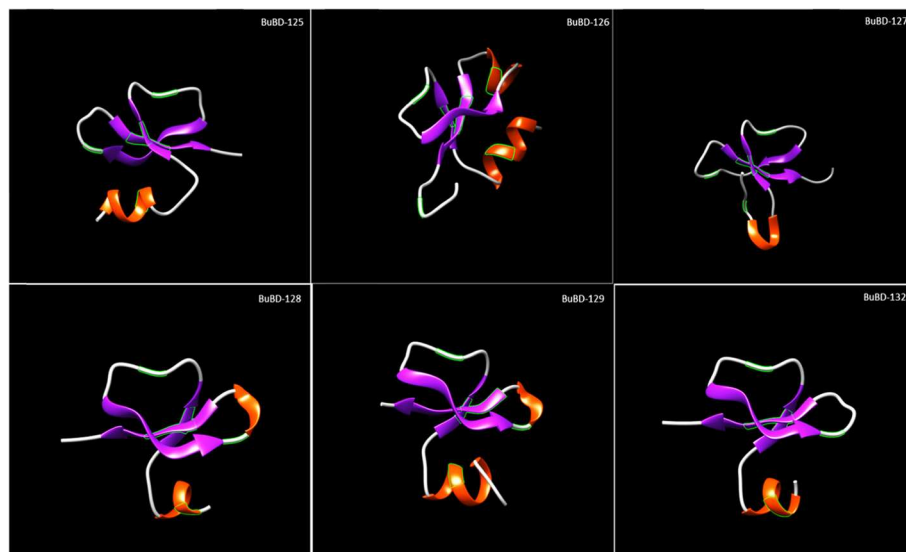
## Evolutionary analyses

### Phylogenetic tree

The consensus tree indicated a strong evolutionary relationship between the buffalo CA-BD genes and their orthologs in other ruminant, pseudo-ruminant and non-ruminant species (Fig. 3). These defensin family members in buffalo have maintained a clear 1:1 orthologous relationship with BDs in their respective orthologous families, and they were found to be clustered accordingly. Interestingly, the BuBDs-125, 126, and 127 sub-clustered with ovine or caprine branches, whereas the BuBD-128, 129 and 132 segregated from *Bovidae* sub-sub cluster. Moreover, BuBD-128, 129 and 132 appeared to be more evolutionary related to the indicine than the taurine cattle.

### Selection pressure characterization

Various statistical models and tools in the HyPhy (Hypothesis testing using Phylogenies) package determined that episodic and pervasive contrasting selective pressures acting on the BD genes have expanded and maintained the BD functional repertoire in the buffalo and in the other ruminants. This temporal orchestration of the contrasting selective pressures has endowed the BDs with functional diversity and the ability to preserve the novel acquired functions.



**Fig. 2** Putative 3D structures of BuBDs. The class-A BuBDs were modelled on the available structures of BDs, in the protein data bank (PDB), using the SWISS-MODEL server. BuBD-125,126 and 132 were modelled on human beta defensin 6, hBD6 (2lwl.1.a), BuBD-127 on hBD4 (5ki9.1.A), BuBD-128 on hBD1 (1kj5.1.a) and BuBD-129 on the bovine neutrophil defensin BNBD (1bnb.1.a) All the defensins possess the characteristic  $\beta$ -fold despite the low homology amongst orthologs



test, at  $P = 0.05$ ). No recombination breakpoints were found in other CA-BDs. The selection pressure was subsequently characterized at the branch and the site levels by selecting the buffalo or all true ruminants ‘a priori’ as the foreground branches for analyses (Table 2 and 3). The aBSREL (adaptive branch-site random effects likelihood) tool didn’t reveal any evidence of the episodic selective pressure acting on the exon1 (E1), exon2 (E2) and the whole-CDSs either in the ruminants or the buffalo. However, RELAX revealed an ineffable intensification of positive selection only on the whole-CDS of BuBD-125 and E2 of BuBD-127 in the buffalo and other true ruminants despite the fact that no episodic selective pressure was found at any of the branches of the CA-BDs. The MEME (mixed effects model of evolution) tool detected the strong episodic site-level diversifying selection acting on the E2 and the whole-CDS of the ruminant and buffalo BD-125, 126, 128 and 129. The BD-125 and 129 had 13 and 6 codon sites under such selective pressures both in the ruminants and the buffalo (at  $P = 0.05$ ). No episodic diversifying selection was, however revealed in the E1 of any of the ruminant or buffalo BDs. The BUSTED (branch-site unrestricted statistical test for episodic diversification) tool revealed the gene-wide evidence of episodic diversifying selection only in whole-CDS of ruminant BD-127 ( $P = 0.026$ ) and the E1 of BuBD-126 ( $P = 0.006$ ).

Taken together, ample evidence of site-level episodic diversifying selection in the CA-BDs was revealed in the ruminant and buffalo CA-BD gene elements. Nonetheless, as determined by the fixed effect likelihood (FEL) tool, the E1 of ruminant and buffalo BD-126 had more sites under pervasive diversifying selection, while the E1 of ruminant BD-132 was under purifying selection. In ruminants, the E2 of BD-125 and 132 had four and three codon sites respectively, under pervasive purifying selection (at  $P = 0.05$ ) in contrast to the buffalo where BuBD-126 was the only CA-BD under pervasive purifying selection. Additionally, all the ruminant CA-BD CDSs, except the ruminant BD-127, had at-least one codon site under pervasive purifying selection (at  $P = 0.05$ ). However, two codon sites of the ruminant BD-125 and no sites of the BD-132 were under pervasive diversifying selection (at  $P = 0.05$ ). Likewise, the E1 of BuBD-126 (3 codon sites) and 132 (1 codon site) and the E2 of BuBDs-125 (four codon sites) and BuBD-126, 127 & 129 (one codon site each) revealed the evidence of site-level pervasive diversifying selection ( $P = 0.05$ ).

**Analysis of selection pressure on the CA-BD orthologs by molecular evolutionary genetic analyses (MEGA-X)**

The MEGA-X software detected the evidence of only purifying selection acting on CA-BDs, which appeared to

**Table 2** Compiled results of the various statistical tools, used to characterize the selection pressure in the CA-BD genes amongst the ruminants selected ‘a priori’ as the foreground branches, as implemented on the datamonkey server, a graphic user interface (GUI) for HyPhy (Hypothesis testing using Phylogenies). NS = non-significant, xPyN (for FEL) = x codon sites under positive and y under negative selective pressure respectively, RuBD = Ruminant beta-defensin

	HyPhy- a priori Ruminants																	
	aBSREL			RELAX			FEL			MEME			BUSTED			GARD		
Defensin	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene
RuBD-125	No Sites	No Sites	No Sites	NS	NS	P=0.006	No Sites	3P4N	2P4N	No evidence	13P	4P	No evidence	No evidence	No evidence	NS	NS	NS
RuBD-126	No Sites	No Sites	No Sites	NS	NS	NS	2P1N	1P0N	2P1N	No evidence	3P	2P	No evidence	No evidence	No evidence	NS	NS	NS
RuBD-127	No Sites	No Sites	No Sites	NS	p=0.012	NS	1P0N	4P0N	4P0N	No evidence	0P	0P	No evidence	No evidence	p=0.026	NS	NS	NS
RuBD-128	No Sites	No Sites	No Sites	NS	NS	NS	2P0N	2P1N	2P1N	No evidence	1P	1P	No evidence	No evidence	No evidence	NS	NS	NS
RuBD-129	No Sites	No Sites	No Sites	NS	NS	NS	0P0N	5P0N	3P1N	No evidence	6P	6P	No evidence	No evidence	No evidence	NS	NS	NS
RuBD-132	No Sites	No Sites	No Sites	NS	NS	NS	0P2N	1P3N	0P2N	No evidence	0P	0P	No evidence	No evidence	No evidence	NS	NS	NS

**Table 3** Compiled results of the various statistical tools, used to characterize the selection pressure in class-A beta-defensin genes in the buffalo selected ‘a priori’ as the foreground branch, as implemented on the datamonkey server, a GUI for HyPhy (Hypothesis testing using Phylogenies). NS = non-significant, xPyN (for FEL) = x codon sites under positive and y under negative selective pressure respectively, BuBD = Buffalo beta-defensin

HyPhy- a priori Buffalo																		
Defensin	aBSREL			RELAX			FEL			MEME			BUSTED			GARD		
	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene
BuBD-125	No Sites	No Sites	No Sites	NS	NS	P=0.006	0	4.0	4.0	No evidence	13P	4P	No evidence	No evidence	No evidence	NS	NS	NS
BuBD-126	No Sites	No Sites	No Sites	NS	NS	NS	3.0	1.0	1.1	No evidence	3P	2P	P=0.006	No evidence	No evidence	NS	NS	NS
BuBD-127	No Sites	No Sites	No Sites	NS	P=0.012	NS	0.0	1.0	1.0	No evidence	0P	0P	No evidence	No evidence	No evidence	NS	N	NS
BuBD-128	No Sites	No Sites	No Sites	NS	NS	NS	0.0	0.0	0.0	No evidence	1P	1P	No evidence	No evidence	No evidence	NS	NS	NS
BuBD-129	No Sites	No Sites	No Sites	NS	NS	NS	0.0	2.0	1.0	No evidence	6P	6P	No evidence	No evidence	No evidence	NS	NS	NS
BuBD-132	No Sites	No Sites	No Sites	NS	NS	NS	1.0	0.0	0.0	No evidence	0P	0P	No evidence	No evidence	No evidence	NS	NS	NS

be the major force in shaping the BD evolution in ruminants under dissimilar substitution patterns. The substitution patterns of the sequences were found to conform to homogeneity in the E1 of all the CA-BDs (Table 4) across the selected mammalian species. This suggested that these sequences have evolved with the same pattern of substitution, as adjudged from the extent of differences in base composition biases between sequences. However, the test of the homogeneity showed a significant ( $P < 0.05$ ) difference in the substitution patterns of the E2 and whole-CDSs among all the CA-BDs, except for BD-128 and 132.

The probability values from the one-tailed Fisher’s exact test of neutrality were non-significant for the sequence pairs of both the exons as well as the whole-CDSs of all the CA-BDs indicating the possibility of negative or positive selection on these gene elements. However, the codon-based test for analysis between the sequences revealed no evidence of positive selection ( $P > 0.05$ ) on the E1, E2 and the whole-CDSs. Nevertheless, the codon-based tests considering all the sequence pairs together determined a significant excess of synonymous substitution per site (dS) over non-synonymous substitutions per site (dN), ( $dS >$

**Table 4** Compiled results of the various tests characterizing the selective pressures in the CA-BD genes in buffalo, as implemented in MEGA-X. NS = non-significant, S = significant, Pur. sel. = Purifying selection, RuBD = Ruminant beta-defensins

MEGA-X															
Defensin	Test of Homogeneity			Z test for pos. selection			Z test for neg. selection			Overall Z			Tajima		
	E1	E2	Whole Gene	E1	E2	Whole Gene	E1	E2	Whole Gene	E1	E2	Whole Gene	E1	E2	Whole Gene
RuBD-125	NS	S	S	NS	NS	NS	BuBD vs OBD; Ruminants vs monogastrics	BuBD vs all ruminants & rabbit;	BuBD vs all ruminants except sheep & rabbit;	Pur. sel.	Pur. sel.	Pur. sel.	-1.2973	0.5607	-0.807852
RuBD-126	NS	NS	S for All ruminants vs EBD	NS	NS	NS	All ruminant BDs except BuBD vs EBD;	BosBDs vs GBD; GBD vs YBD	BosBDs vs CBD & EBD; BuBD vs CBD,	NS	NS	Pur. sel.	-0.5961	-0.788	-0.743053
RuBD-127	NS	S for all ruminants vs. monogastrics	S for All ruminants vs OcBD	NS	NS	NS	All ruminant BDs except BuBD vs LBD;	All ruminants BDs vs LBD, HBD & OcBD	All ruminants BDs vs LBD, HBD & OcBD ;	Pur. sel.	Pur. sel.	P	-1.5709	0.0100	-0.07224
RuBD-128	NS	NS	NS	NS	NS	NS	All ruminant BDs vs CBD;	ll ruminants BDs vs LBD	BuBD & BosBDs vs monogastrics	Pur. sel.	Pur. sel.	Pur. sel.	-0.3261	0.2871	0.251051
RuBD-129	NS	S for All ruminants vs OcBD, FeBD, OvBD	S for All ruminants vs OcBD, FeBD, OvBD	NS	NS	NS	All ruminant BDs vs LBD	All ruminants BDs vs LBD, CaBD, CBD, EBD & HBD	All ruminants BDs vs others’BDs	Pur. sel.	Pur. sel.	Pur. sel.	-0.4017	0.2059	-0.201798
RuBD-132	NS	NS	NS	NS	NS	NS	All ruminant BDs vs OBD	BuBD&HBD vs OvBD	BuBD, BIBIBD & HIBD vs OvBD	NS	NS	NS	-2.4151	1.2294	-1.4

$dN_p = 0.01$ ) in most of the E2 and whole CDSs of all the considered CA-BDs, except BD-132. Tajima's D was found to be negative for whole-CDSs of most of the CA-BDs, except for the BD-128. Thus, the substitution patterns acting on BDs appear to be non-neutral and such dissimilar substitution patterns are indicative of directional selection.

**Analysis of selection pressure on the CA-BD paralogs by HyPhy**

HyPhy analyses indicated that a strong pervading purifying selection interwoven with the episodes of diversifying selection has acted on the CDSs of the buffalo paralogs. As observed in the orthologs, aBSREL analysis (Table 5) was not able to detect any evidence of episodic selective pressures on any of the branches of the exons and the whole-CDSs of the 30 BuBD paralogs. However, MEME detected episodic diversifying selective pressures across the E2 and the whole-CDS, but not in the E1 of these paralogs. The presence of gene-wide incidences of diversifying selection acting on the E2 ( $P = 0.023$ ) and whole-CDSs ( $P = 0.000$ ) of BuBD paralogs was revealed by the BUSTED tool. However, the results were non-significant if the class-A BDs were selected 'a priori' as foreground branches to be tested.

The FEL results revealed that the E1 of all BuBD paralogs had one codon site, while the E1 of the CA-BDs had two codon sites under pervasive purifying selection (at  $P = 0.05$ ). Similarly, the E2 of all BuBDs had 16 codon sites and the E2 of class-A BuBDs had 17 codon sites under pervasive purifying selection (at  $P = 0.05$ ). Thirteen codon sites were determined under pervasive purifying selection in comparison to a single codon site under pervasive diversifying selection in the whole-CDSs of the BuBD paralogs (at  $P = 0.05$ ). The class-A BuBDs also exhibited a similar pattern with 10 codon sites under the pervasive purifying and 2 codon sites under the pervasive diversifying selection (at  $P = 0.05$ ).

**Expression dynamics of the CA-BD genes across the MRT and other tissues**

The relative expression profiles of the selected panel of class-A BuBD genes were generated using qRT-PCR, in the five different segments of the MRT viz. Rete testis, the caput, corpus and, the cauda of epididymis and vas deferens (Additional file 3) and six non-reproductive tissues viz. Brain, heart, spleen, bladder, liver and lung from buffalo bulls, and two reproductive tissues from the FRT, the uterus and ovary. The GAPDH (glyceraldehyde-3-phosphate dehydrogenase) and eEF-2 (eukaryotic elongation factor) were employed as the reference genes, and the rete testis was considered as the calibrator tissue since the expression of the BDs, especially the CA-BDs is known to be absent in the rete-testis. The expression of BuBD-126 and a novel candidate BuBD-129 was found to be the highest in the distal segments of the MRT. A large variability in the expression levels in the MRT was observed among the biological replicates. In contrast to the cattle, the buffalo non-reproductive tissues actively expressed the CA-BDs although without much inter-animal variation.

**Expression of the class a BuBDs in the MRT**

A high spatial and inter-animal variability was observed in the gene expression pattern of class-A BuBDs across the MRT. The highest expression of the class A BuBDs was found in the epididymis specifically the corpus epididymis, which also exhibited highest inter-animal variability in the expression levels. Primarily, a site-specific expression pattern of the class-A BuBDs was observed across all the segments of the MRT starting from the caput epididymis. The mean expression of BuBD-125 in the corpus epididymis was higher compared with the expression observed in either the caput ( $P > 0.05$ ) or in the distal segments of MRT, the cauda ( $P < 0.01$ ) and the vas deferens ( $P < 0.001$ ). Interestingly, the ortholog of DEFB-126 in buffalo i.e. BuBD-126 was found to be

**Table 5** Compiled results of the various tools, used to characterize the selection pressures in BuBD paralogs of the buffalo in exploratory mode and 'a priori' with the CA-BD genes selected as the foreground, as implemented on the datamonkey server, a GUI for HyPhy. NS = non-significant, xPyN (for FEL) = x codon sites under positive and y under negative selective pressures respectively, BuBD = buffalo beta-defensin

HyPhy-Buffalo Paralogs																		
	aBSREL			RELAX			FEL			MEME			BUSTED			GARD		
	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene	E1	E2	Whole gene
BuBDs	Exploratory -No sites, Class-A genes: No sites	Exploratory -No sites, Class-A genes: No sites	Exploratory- No sites, Class-A genes: No sites	NS	NS	NS	Explorator y-0P1N, Class A genes: 0P2N,	Exploratory- 1P16N; Class-A genes: 0P17N,	Exploratory- 1P13N; Class-A genes: 2P10N	0	1P	2P	Exploratory- No sites, Class-A genes: No sites	Exploratory- P=0.023; Class-A genes: NS	Exploratory- P=0.000; Class-A genes: NS	NC	NS	NS

preferentially expressed in the corpus and cauda epididymis and decreased later in the VD ( $P > 0.05$ ). The expression level of BuBD-126 was significantly higher in the caudal region of the epididymis as compared to the caput epididymis ( $P < 0.05$ ). On the contrary, BuBD-127 was preferentially expressed in the caput epididymis and diminishes thereafter. Its level of expression in the distal segments of MRT i.e. cauda and vas deferens was not significantly different from each other. For BuBD-128 the proximal regions i.e. the caput ( $P < 0.01$ ) and the corpus epididymis were the preferred sites of expression, after which its expression decreased significantly in the cauda ( $P < 0.05$ ) and VD ( $P < 0.01$ ) when compared to corpus epididymis. The expression of BuBD-129 followed the expression dynamics of BuBD-126 with peak expression in the corpus epididymis ( $P < 0.001$ ) and maintaining this elevated expression in the distal MRT segments. The coefficient of variation values indicated that the corpus epididymis exhibited the highest inter-animal variability in the expression levels of the BuBD-129 across the MRT. The BuBD-126 and BuBD-129 were the two class-A BuBDs with the highest expression levels in distal regions viz. the cauda epididymis and the vas deferens. BuBD-132 initially followed a similar trend in expression, however, its expression levels decreased in the distal MRT segments. As evident from the heatmap (Additional file 3) the relatively proximal parts of the MRT viz. caput and corpus epididymides lied in one cluster and the distal parts of the MRT cauda and vas deferens in a different cluster. It indicated that the mean expression profiles for proximal regions of the MRT were similar to each other like that of distal regions. However, the mean expression profiles of the proximal and distal region were very different from each other.

#### Expression of the class-a BuBDs in other tissues

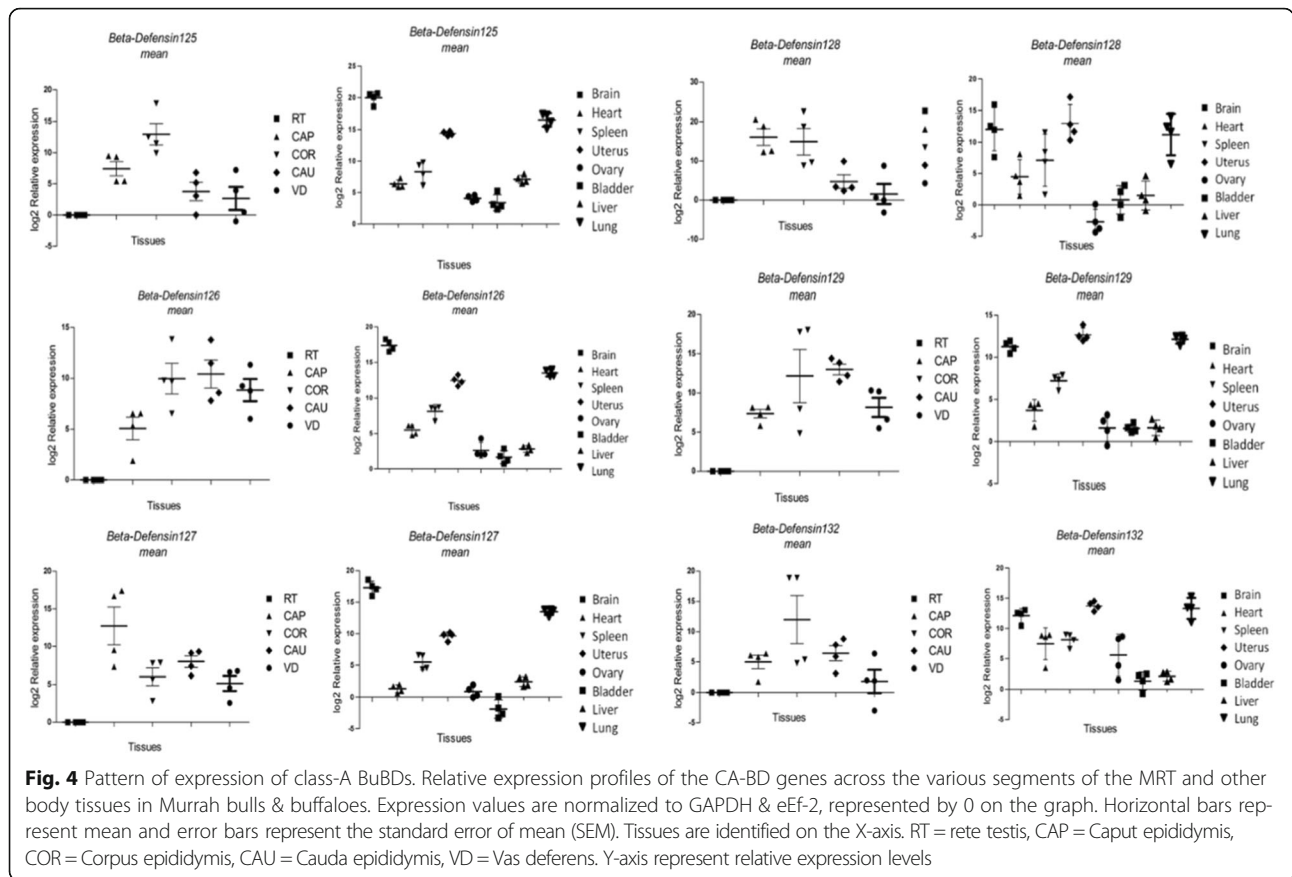
The expression of the six CA-BDs was validated not only in the male reproductive tissues but also in non-reproductive tissues as well as female reproductive tissues of the buffalo. The variability among the biological replicates in the expression levels of BuBDs was comparatively lesser in the non-reproductive tissues than that of the male reproductive tissues. However, the disparate patterns of preferential spatial expression of BuBDs were present even in the non-reproductive tissues with the highest expression observed in the brain (Fig. 4, and Additional file 3). Particularly, the highest expression of BuBD-125, 126 and 127 ( $P < 0.001$ ) was observed in the brain followed by the lung. Similarly, the expression of BuBD-128, 129 and 132 ( $P < 0.01$ ) was also elevated in the brain followed by the uterus and the lung. The ovary and the bladder were the two tissues exhibiting the least levels of the expression, while an intermediate expression level was observed in the uterus for most of the class-A BuBDs. These observations surprisingly

provided a clue that in contrast to the cattle, the buffalo female reproductive tissues actively expressed CA-BDs.

#### Discussion

The present study was undertaken to identify the cattle BD orthologs for characterization and evaluation of their sequence, gene ontology and physicochemical, structural characteristics of the translated products of class A BuBDs. In addition, the selection pressure characterization and the expression analysis of class-A BuBDs in the MRT and other body tissues were also performed to predict their reproductive functional significance. Thirty cattle BD orthologs, including the CA-BDs, were mined from the genome assemblies of the ruminants, including buffalo, pseudo-ruminants and non-ruminants. Despite the low sequence homology, the buffalo CA-BDs possess an analogous residue profile to their orthologs, thus have maintained the structural core which is required for their functions. The investigated CA-BDs were found to possess a similar molecular framework and secondary structural elements as reported in other BDs. Among CA-BDs, the BuBD-129 appears to be the novel functional ortholog of the primate or rodent BD-126. The diversity in the BD arsenal of buffalo immune system appears to be a result of contrasting selection pressures acting on their gene elements at different timescales.

The ensembles of microorganisms present in and on multicellular animals and their bodily fluids vary according to the physiochemical milieu in which they flourish [50, 51]. These pathogens drive the molecular evolution of their animal hosts by antagonistic co-evolution [52]. For example, the ruminants developed sophisticated immune mechanisms by their co-evolution with the rumen microflora [7]. Accordingly, a positive correlation exists between the, dynamic microbial insults in the niches inhabited by the host animals and the number BD genes found in a species [53]. We have mined 30 BD sequences (RNA & CDSs) in the buffalo, including 15 novel BDs, along with classical AMPs (LAP, TAP, EBD, etc.) that represents a large expansion of the buffalo BD repertoire comparable to that of the cattle. Our results indicated a rapid evolution of these genes in the buffalo through duplications usually followed by their diversification of their sequence and function, as reported in the cattle and other mammalian species [54]. This swift-evolution, considering their pleiotropic roles and the 3-D structural similarities despite the sequence dissimilarities, indicate diverse immune and non-immune functions of the buffalo CA-BDs. The class-A BuBDs considered in this study were predicted to possess antimicrobial activity by the SVM classifier. This indicates the putative conservation of the primordial antimicrobial function of the BDs in spite of attaining novel additional functions after episodic/pervasive diversifying selection [39, 55]. This prediction is underpinned by our



finding that the CA-BDs were found to be expressed in a variety of non-reproductive tissues of the buffalo apart from the MRT.

All the class-A BuBDs were found to retain the canonical six cysteine motif, “X2–10CX5–6[G/A] XCX3–4CX9–13CX4–7CCXN”, in addition to a characteristic sequence feature the G-X-C motif which is required for their proper folding [56]. The class-A BuBDs contain ample basic residues such as lysine and arginine, which appears to help them in interacting with the negatively charged plasma membranes of microorganisms by electrostatic interactions [57]. The amino-acid compositional bias, in terms of the reduced share of the order-promoting residues and the prevalence of polar amino acids, substantiates the prediction of loops in the major region of these peptides which must render a substantial surface area of these BDs accessible to the surrounding solvent. The presence of the six-cysteines in a conserved molecular framework and the predicted presence of three β-strands in the anti-parallel arrangement suggest that the information for proper folding indeed exists in the primary sequence of the class-A BuBDs [58]. The three disulfide bonds stabilize the structural properties of the characteristic ‘β-fold’, which is vital to their molecular function/s. The increasing dissimilarity between the translated CA-BD gene products with the increasing evolutionary

distance has had a comparatively lesser effect on the 3-D structure of buffalo CA-BDs. This explains why the ‘hallmark 3-D structure’ of the BDs was predicted to tolerate substitutions with residues of diverse physicochemical properties at most positions. However, not at the sites which are required for proper folding [59] e.g. the six cysteines and their neighbouring sites, as observed from the SNAP2 results for class-A BuBDs.

The focus of the recent studies has shifted from the canonical antimicrobial role played by these HDPs to their pleiotropic functions, for example, in the male reproductive processes where they assist the sperm to ultimately fertilize the oocyte. The major SSR events, which are necessary not only for the survival of the sperm in the FRT (e.g. immune-evasion) but also for the capacitation and fertilization with the oocyte, take place in the epididymis. The presence of the buffalo CA-BD transcripts in the MRT indicate their secretion in the epididymal lumen of the buffalo. Furthermore, the glycan modifying enzymes e.g. glycosidase and glycosyltransferase are considered critical for the SSR events. Their family members transform the glycocalyx of the molecules involved in fertilization-specific activities which are also known to embellish the sperm-surface e.g. addition of sialic acid to DEFB-126 [38, 34], and [60]. The

spermatozoa require DEFB-126 to penetrate the cervical mucus [42], to evade the immune response mounted in the FRT [43], and to form the oviductal sperm reservoir [14, 61]. Moreover, a small deletion (2 nucleotides) in the DEFB126 gene that modifies its O-glycan structure and thus the sperm surface was found to be implicated in subfertility in the human cohorts [62]. The terminal sialic acids present abundantly on DEFB-126 not only delay, but also assist the sperm in the capacitation process as and when required in the FRT [63]. Intriguingly, a novel CA-BD, BuBD-129 was predicted to be the most heavily glycosylated CA-BD family member rather the BD-126 as reported in primates and rodents [42]. It is thus legitimate to hypothesize that BuBD-129 could be the functional ortholog of the BD-126 in humans (Primates) and mice (Rodentia) primarily because of its preferential expression in the distal segments of MRT, a high potential of O- and N-glycosylation in its translated gene products, its gene length, chromosome cluster, extended tail as reported in the DEFB-126.

The BDs are primarily secreted by the epithelial cells of the gut, skin, airway, mouth, kidney, nose, eyes and mammary glands. However, the CA-BDs have been identified majorly in the epididymal luminal proteome that binds the sperm surface [45, 48, 64–67], and [68]. The buffalo CA-BDs were predicted to be secreted in the extracellular space, possibly into the epididymal lumen of the distal segments of the MRT. This was substantiated by the presence of the polar, hydrophilic amino acid residues, prediction of the loops as the major, secondary structural elements and the expression of their genes in the major non-reproductive tissues. Recently, most of the CA-BD genes (BBD125, 126, 127 and 128) have been characterized along with their genetic variants, which have now been implicated in the cattle bull sperm function [40].

In the present study, the comparison of CA-BD genes across the selected mammalian species revealed an evolutionary pattern, in which the different BD genes appeared to follow different evolutionary paths, albeit at different evolutionary time scales. Selection pressure characterization of the CA-BD genes revealed ample evidence of episodic diversifying selection in the ruminants, including the buffalo. Many immune-related genes have revealed evidence of directional or diversifying selection as they have to counter a variety of pathogenic fauna and gut flora which are evolving at high rates of genetic drift and shift [7, 69–71], and [72]. Afferent molecules, which are involved in sensing and recognizing the pathogens, are required to distinguish and thus help to eliminate a variety of pathogens. Such molecules have to adapt to the dynamic pathogenic fauna and microflora in their niches. Therefore, the afferent molecules are found to be under diversifying selection pressure. A relatively strong episodic positive selection pressure was

found to have swept across the E2 and whole CDSs of the selected ruminants and the buffalo CA-BD genes. On the other hand, most ruminant CA-BDs were also found to contain at least one codon site under pervasive purifying selection. Efferent molecules which are involved in eliminating infections can't afford to lose their specialized function, thus they are under purifying selection pressure [73]. The buffalo and ruminant CA-BDs considered in this study revealed evidence of episodic diversifying and pervasive purifying selection pressure on the E2 and the whole-CDSs which appears to be the driving force for sequence and functional divergence of the CA-BDs in the buffalo and possibly other ruminants. The CA-BD genes are now known to be implicated in male fertility in a number of species where they assist the spermatozoa to achieve their ultimate goal of fertilization [7, 30, 42, 46, 62, 63], and [74]. In fact, the evolutionary driver for innate immune genes could either be positive or negative selection, which depends upon their functional properties in the immune cascade [75]. The evidence of purifying selective pressure in the CA-BD gene elements also implies that the natural selection has had selected the most efficacious alleles because low tolerance for new variants is a common strategy for genes involved in the efferent arm of immunity [76]. Nevertheless, the evidence of diversifying selection pressure on these innate effector genes can't be ignored because of its tendency to widen the spectrum of functional activity of these genes due to high rates of genetic drift, a signature of the innate immune system [77].

The observed patterns of selection of BD genes observed in this study, could have been moulded by the microbiome with which the given immune genes had co-evolved. Besides, factors like demographic features, and the microbial load of the host might also have important implications in deciding the direction and type of the selection [78]. Our results thus revealed that both the purifying [39, 55] and [78] and the diversifying selection [79, 80] have acted on the CA-BD genes, however, at different time scales. Although the major pervasive selection force in maintaining the crucial CA-BD alleles in ruminants appears to be the purifying selection, the evolutionary age of the BD genes should also be considered [81]. This is because of the duplication events, which appear to initiate a period of relaxed selection, where one of the duplicates maintains the wild type gene function and the other duplicate is free to explore a new functional space. Thus, if a new function is acquired then one of the duplicates will subsequently be maintained by purifying selection [82], as observed in the BuBD paralogs. It may also be possible that a BD gene earlier subject to positive selection may not be subject to such selective forces and vice versa due to the environmental dynamicity [79, 83]. The purifying selection force acting on the mature peptide (E2) and the whole CDSs of the ruminant CA-BD

genes is expected to drive a low allele diversity as observed in some species [84, 85], while the diversifying selection force appears to have produced the diversity in the functional repertoire of these pleiotropic AMPs [79, 86]. Considering the contrasting selective forces, the CA-BDs appear to have evolved under an intricate balance of contrasting selective pressures and restraints, possibly a putative compromise between utilitarian selection (e.g. role in innate immunity) and sexual selection (e.g. role in reproductive processes).

The expression of class-A BuBDs, hitherto, has known to be restricted to the MRT and to be age-specific [7, 30], and [75]. However, constitutive expression of class-A BuBDs was observed in the buffalo non-reproductive and reproductive tissues in the absence of any visible infection. This indicates that BuBDs aren't just evolutionary relics of their orthologs, but they are fully-fledged functional genes that are transcribed and probably translated too. Our observations differ from what has been reported in the cattle [7, 30], and [75] where their expression, including the BBD-126 was found to be restricted only to the MRT. Contrarily, some members of the CA-BD family, have also been reported in non-reproductive tissues of humans [65, 87], pig [88], rabbit [89] and even in the equine females [90]. The expression of class-A BuBDs in non-reproductive tissues apart from MRT tissues points to their active antimicrobial roles, which means that these defensins are still maintaining the primordial antimicrobial activity even though having gained new additional functions by diversifying selection. Further functional assays are required to validate their molecular functions. The expression of these genes in MRT had been reported to be absent in the MRT of immature cow bulls [30]. Interestingly, the protein product of the BBD-126 was later shown to be present in the testes of immature cattle bulls despite the absence of any transcription [91]. Further investigation is required to ascertain whether the protein BBD-126 is translocated from other tissues to the testes of immature cow bulls. As discussed earlier, until recently, the expression of the CA-BDs has been ascribed to specific sex, age and tissue. An interesting question that arises is "why is this specificity not observed in the buffalo?" Is it because of the differential composition of the ensemble of flora in the gut of the buffalo? Though the answer remains enigmatic, a clear picture has emerged, that the buffalo is as different from the cattle as are other ruminants with reference to the CA-BD genes. Our observations on the expression analysis of CA-BDs in the MRT indicated that the observed spatial variation in their expression point towards the variable biological functions at different sites [69, 91], and [92]. The spatial variation observed in the expression of these CA-BDs could be explained by the fact that the different BDs target only the region-specific pathogens in accordance with their anatomical distributions [56, 82]. The distal segments of the MRT

(the cauda epididymis and the V.D.) are the regions that usually exhibit the highest expression levels of the CA-BDs [34, 35]. A similar trend was observed for the buffalo CA-BDs, especially for a novel candidate BuBD-129 in addition to the BuBD-126, which also had abundant transcripts in the distal segments of the MRT. This unusual expression pattern of the class-A BDs in the buffalo organs such as the heart, brain, spleen, and the epididymis indicated that BuBDs appeared to have evolved in coherence with the hypothesis of 'Niche adaptation' [7]. It is also likely that the observed high basal-level expression represents a response, which is induced in the context of colonization (tolerance to self and co-inhabitants), or any invisible infection thus being an immune advantage [50]. Infections or pathogenic insults are an important force driving the evolution of defensins and other immune genes. The high and the variable expression of the class-A BDs in the buffalo MRT without any visible infection could also be indicative of their putative role in assisting the sperm to achieve their goal of fertilization, which has also been established to be implicated in bull fertility of cattle [40].

## Conclusion

The buffalo possesses a diverse catalogue of BD genes, which have maintained a direct orthologous relationship with several ruminant and non-ruminant species. The increasing dissimilarity between sequences with increasing evolutionary distance has had a comparatively lesser effect on the protein structure of class-A BuBDs. A complex interplay of contrasting selective forces acting on different evolutionary timescales has raised a diverse repertoire of the BD arsenal in the buffalo immune system. The expression pattern of class-A BuBDs observed in this study suggested that these molecules maintain an infection-free environment not only in the reproductive but also in the non-reproductive parts of the buffalo. The quantitative variation in the expression patterns in various tissues signifies that the site-specific expression pattern may be attributed to the differential functions performed by the class-A BuBDs in accordance with their anatomical distributions. A study is warranted to assay the antimicrobial and other pleiotropic activities of these defensins in the MRT or on the sperm-surface. As the annotation of the buffalo genome is presently in the primitive stage, a fully sequenced and annotated buffalo genome may lead to the thorough characterization of the BuBDs, the discovery of additional orthologs and elucidation of their functional attributes.

## Methods

### Experimental design

The RNA sequences and complete CDSs of *Bos taurus* BDs (retrieved from assembly ARS-UCD 1.2) were used to mine their sequence orthologs in the latest,

chromosome level-RefSeq assemblies of buffalo and 15 other species including true-ruminants, pseudo-ruminants and non-ruminants. The retrieved CDSs of the six CA-BDs were translated and subject to various in silico tools to evaluate their physicochemical properties and some post-translational modifications and to perform the sequence ontologies of the CA-BDs. The CA-BDs, known to play crucial roles in reproduction, were further used for evolutionary (selection pressure characterization), structural and expression pattern analyses in  $n = 4$  biological replicates in five segments of the MRT (Rete testis, caput, corpus & cauda epididymis, and vas deferens) and eight other body tissues (Brain, heart, spleen, bladder, uterus, ovary, lung and liver).

#### In silico sequence retrieval & analyses

Fifty-seven beta-defensins (BDs), maximum in any ruminant species, have been reported in *Bos taurus* [48], therefore it was considered the reference-species for searching orthologs in buffalo and other selected mammalian species. The RNA/CDS sequences for *Bos taurus* BDs were mined from its latest RefSeq assembly (ARS-UCD 1.2). These RNA/CDS were subjected to BLASTN [93] against the latest RefSeq assembly of *Bubalus bubalis* (UOA\_WB\_1) and other ruminants and non-ruminant species. The sequences were obtained from GenBank [94], if not found in respective assemblies. The assemblies of 'chromosome level (if available) were used to determine the loci, genome coordinates, duplications, and orientation of the genes and gene elements.

The retrieved CDSs were translated using the ExPASy translate tool [95]. The sequence characteristics & ontology, secondary structure, composition, solvent accessibility etc. of the class-A BuBD proteins (Chromosome 13 defensin cluster) were predicted by the various tools as implemented on the Predict protein server [96]. The functional effects of various mutations/SNPs were predicted by SNAP2, a trained classifier that not only takes the sequence and variant features into account, but also the evolutionary information from an automatically generated MSA [97]. The protein disorder in the class-A BuBD proteins was predicted by Meta-disorder predictor method [98]. It is a combination of several orthogonal methods that captures many types of disorder, combining the output from various prediction methods with sequence profiles and other useful features such as predicted solvent accessibility, secondary structure and low complexity regions. The Gene Ontology analysis for the BuBDs was performed by Metastudent, which predicts the gene ontology (GO) terms for protein sequences through homology [99]. The presence of the Protein-Protein Interaction, PPI and Protein-Nucleotide Interaction, PNI was predicted through the interaction sites identified by profiSIS [100] and someNA [101] tools. The neural network predictions of mucin type GalNAc O-

glycosylation sites were produced using NetOGlyc 4.0 server [102] whereas N-glycosylation sites were predicted using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr *sequons* using NetNGlyc 1.0 server [103]. The predictions for the presence of a trans-membrane helix were made using TMHMM server v.2.0 [104]. Since the BDs are known to be HDPs, an in silico prediction of the presence of the antimicrobial activity was done using Support vector machine classifier algorithm available at the CAMP<sub>R3</sub> database [105]. The presence of the signature G-X-C motif was inspected manually in the CA-BDs.

#### Structure analysis

Very few structures of BDs exist in Protein Data Bank-PDB [106], although a search for the term 'beta-defensin' yields 74 results. However, only ten of them are mammalian defensins and the only ruminant BD listed is BNBD-1, the bovine neutrophil beta-defensin. Thus the class-A BuBDs could be modelled only on the existing BD structures in the PDB. The BuBD-125,126 and 132 were modelled on human BDs namely hBD6 (2lwl.1.a), BuBD-127 on hBD4 (5ki9.1.A) respectively, whereas the BuBD-128 was modelled on hBD1 (1kj5.1.a) and BuBD-129 on the bovine neutrophil defensin, BNBD (1bnb.1.a). The structural features of the CA-BDs were analysed using the folding patterns of these peptides as predicted by the SWISS-MODEL homology-modelling server [107]. The PDB files of the models were viewed with UCSF-Chimera *ver.* 1.13.1 [108]. The structural assessment of the models was done using Molprobit *ver.* 4.4 [109].

#### Evolutionary analyses: Orthologs and Paralogs

The retrieved class-A BuBDs were compared with their orthologs in other ruminants (*Bos taurus*, *Bos indicus*, *Brahman cattle*, *Bos mutus*, *Bison bison*, *Ovis aries*, *Capra hircus*, *Odocoileus texanus*, pseudo-ruminants (*Camelus dromedaries & bactrianus*), monogastric organisms (*Felis catus*, *Canis familiaris*, *Loxodonta africana*, *Oryctolagus cuniculus*) and primates (*Homo sapiens*). The respective CDSs were aligned using multiple sequence alignment (MSA) on MAFFT *ver.* 7.409 [110] and were visualized using Jalview 2.10.5 [111]. The MSAs obtained from MAFFT program were visually inspected and subsequently subjected to GBlocks filter on translator server [112] to check for ambiguous sequence regions. The cleaned alignments were used for further analysis.

#### Phylogenetic tree

The whole-CDSs of class-A BuBDs along with their exons viz. the E1 and the E2 were aligned using clustal in Mega X [113]. As a preliminary step for phylogenetic studies, the Maximum likelihood method (MLM) fit for

56 different amino-acid substitution models and 24 different nucleotide substitution models [113, 114] were obtained. The best model describing the substitution pattern was selected and was used for subsequent analyses. The evolutionary analysis for phylogenetic tree building was done by the MLM method in conjunction with the best-chosen model. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the analyzed taxa [115].

### Selection pressure characterization

As a pre-processing step for inferring selection pressure, all CDSs for CA-BD genes were checked for evidence of intragenic recombination using GARD [116]. The significant topological incongruence of recombination breakpoints was tested using the Kishino–Hasegawa test [117]. For characterizing the patterns of selection on these sequences, statistical models and tools from the HyPhy (Hypothesis testing using Phylogenies) package implemented in the accompanying datamonkey web-server [118] were used and *P*-values/posterior probabilities of 0.05 were considered significant. The selection pattern characterization was also performed using molecular evolutionary genetic analysis tool MEGA-X.

### Strategy 1: analysis of selection pressure on the CA-BD orthologs by HyPhy

The cleaned alignments for E1, E2 and whole-CDSs of class-A beta-defensin genes were used for the analysis on HyPhy. The testing of branches of interest viz. *Bubalus bubalis* or ruminants, for the episodic positive/purifying selection, was performed using aBSREL [119]. aBSREL is a better version of the common branch-site models and is a preferred approach for identifying positive selection at individual branches since it models both, site-level as well as branch-level  $\omega$  (dN/dS) heterogeneity. The shifts in the stringency of natural selection in the CA-BD genes were determined using a hypothesis testing framework RELAX [120]. The branch with buffalo or all true ruminants as taxa was a priori selected as the foreground *vis-à-vis* the background branches branch to be tested for section pressure characterization. Determination of pervasive (at a subset of branches) positive or purifying selection was done by FEL [121] which uses a maximum-likelihood (ML) approach to infer non-synonymous (dN) and synonymous (dS) substitution rates on a per-site basis. However, for determining episodic positive or purifying selection in the CA-BD sites, MEME was used [122]. For each site, MEME infers two  $\omega$  rate classes and corresponding weights representing the probability that the site evolves under each respective  $\omega$  rate class at a given branch. Lastly, the gene-wide test for positive selection in the CA-BDs by asking “whether a CA-BD gene has

experienced positive selection at any site on at least one of the branches” was performed by BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification) [123]. The parent branch with the buffalo or the ruminants as taxa were ‘a priori’ selected as the foreground branch to be tested. For each phylogenetic partition (i.e. the foreground and background branch sites), BUSTED fits a codon model with three rate classes, constrained as  $\omega_1 \leq \omega_2 \leq 1 \leq \omega_3$ . BUSTED simultaneously estimates the proportion of sites per partition belonging to each  $\omega$  class.

### Strategy 2: analysis of the selection pressure on the CA-BD orthologs by molecular evolutionary genetic analyses (MEGA-X)

Additionally, to test whether the sequences have evolved with a similar pattern of substitution, a Monte Carlo test (1000 replicates) was used to estimate the *p*-values for the disparity index test, indicating the differences in base composition biases between sequences [113, 124]. The codon-based Z-test of neutrality (dN = dS), positive selection (dN > dS) and negative selection (dN < dS) for analysis between sequences as well as averaging over all sequence pairs were conducted for assessment of selection pressures [113, 125]. The test statistic used for detecting the codons that have undergone positive selection was dN – dS, where dS is the number of synonymous substitutions per site (s/S) and dN is the number of non-synonymous substitutions per site (n/N). Moreover, the one-tailed Fisher’s exact test for neutrality was also conducted for both the exons E1 and E2 as well as for complete CDS sequences [113, 125], and [126]. To further corroborate the results the population genetic test statistic Tajima’s D was also calculated [113, 127].

### Analysis of selection pressure on the CA-BD paralogs by HyPhy

To characterize the driving force behind the functional diversification and maintenance of the buffalo BD genes, we tested the role of Darwinian selection. Thus, the retrieved CDSs of the paralogs from buffalo were subjected to the same tests used for the orthologs using the HyPhy package.

### Expression dynamics of the buffalo CA-BD genes across the MRT and other tissues

#### Tissue collection

The MRT and other body tissues were collected in the month of January of 2017 from a local abattoir, Delhi (28°37’40”N, 77°19’51”E) with the average temperature varying from 7°–10 °C. The five segments from the MRT viz. the caput, corpus and cauda epididymides, the vas deferens and the rete testes and eight body tissues viz. the brain, heart, spleen, uterus, ovary, bladder, liver and the lung were collected aseptically within 20 min of

slaughter to obtain 3 mm sized samples from healthy and mature Murrah buffalo bulls ( $n = 4$  biological replicates) aged between 3 and 4 years. The tissue samples were immediately stored in the RNA later solution (Ambion, USA) and afterwards stored in  $-80^{\circ}\text{C}$ . The maturity and age of bulls were confirmed by the presence of spermatozoa in rete testes giving it a milky white appearance and counting of teeth prior to slaughter respectively.

#### RNA isolation

Total RNA was isolated from the MRT tissues using TRI Reagent, RNA isolation reagent (Sigma-Aldrich, USA) and quantified using a NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). The  $A_{260/280}$  and  $A_{260/230}$  were close to 2.0 for all the samples used in the study. The quality of the extracted RNA was assessed by running 200 ng of the RNA (heated at  $70^{\circ}\text{C}$  for 1 min) in non-denaturing TAE buffered 1.2% agarose (Sigma-Aldrich, USA). The isolated RNA was given the DNase I (Thermo Scientific, USA) treatment, according to the manufacturer's instructions to get rid of gDNA contamination.

#### cDNA synthesis

RevertAid, H Minus first-strand cDNA synthesis kit (Thermo Scientific, USA) was used to convert  $2\ \mu\text{g}$  of RNA into cDNA as per manufacturer's instructions. Briefly,  $2\ \mu\text{g}$  of RNA was mixed with oligo-(dT)<sub>n</sub> and random hexamer primers and subsequently the nuclease-free water was added. The reaction mixture was incubated at  $65^{\circ}\text{C}$  for 5 min. Later, the reaction buffer, RNase inhibitor, dNTP mix and the RevertAid H minus M-MuLV reverse transcriptase were added to make final reaction volume of  $20\ \mu\text{l}$  and incubated at  $25^{\circ}\text{C}$  for 5 min, followed by incubation at  $42^{\circ}\text{C}$  for 1 h in a thermal cycler (Biometra, Analytik Jena, Germany). The reaction was terminated by heating the mixture at  $70^{\circ}\text{C}$  for 5 min. A reverse transcriptase negative control was prepared alongside for the assessment of residual genomic DNA contamination in the RNA sample.

#### Primer design

The primer designing tool at NCBI, the Primer-BLAST [128] was used to design primers for buffalo CA-BD genes viz. BuBD-125, 126, 127, 128, 129 and 132 and three reference genes viz. HPRT (Hypoxanthine-guanine phosphoribosyltransferase), eEF-2 (eukaryotic elongation factor 2) and GAPDH (glyceraldehyde-3-phosphate dehydrogenase) from conserved regions found in MSA (MAFFT *ver.7.409*) [110]. Intron-spanning primers, wherever possible, were designed. However, the HPRT was later excluded from the investigation as the  $C_q$ s obtained were greater than 38 in three out of five MRT tissue samples. The self-annealing sites, and the mismatches and the secondary structures in the primers were checked using

Oligo Calc [129]. The specificity of primers was again checked using the BLAST alignment tool [95] and in silico PCR [130, 131] was run for each set of primers prior to commercial synthesis (Sigma-Aldrich, USA).

#### RT-qPCR optimization

To ensure the quality of RT-qPCR and RT-qPCR data MIQE guidelines [132] were followed at every step, wherever possible (Additional file 4).

#### RT-qPCR assay

The relative quantification of the BD genes was done on a Roche LC-96 light cycle platform using the Maxima SYBR Green qPCR master mix (Fermentas, USA) in a  $10\ \mu\text{L}$  reaction mix. The thermal profile was  $95^{\circ}\text{C}$  for 10 min, 40 cycles consisting of denaturation at  $95^{\circ}\text{C}$  for 15 s, annealing at variable optimized temperatures for 20 s and extension at  $72^{\circ}\text{C}$  for 20 s, followed by the melt curve protocol with 10 s at  $95^{\circ}\text{C}$  and then 60 s each at  $0.5^{\circ}\text{C}$  increments between  $65^{\circ}\text{C}$  and  $95^{\circ}\text{C}$ . A no-template control (NTC) was run in each plate to confirm the absence of nucleic acid contamination. A melt curve analysis was performed to ensure a specific, unique product formation and to ascertain minimal primer-dimer formation. The RT-qPCR products were subjected to agarose gel electrophoresis for ensuring specific product formation.

#### RT-qPCR data and statistical analysis

The mean sample  $C_q$  (Cycle of quantification) values for different transcripts were calculated for duplicate samples and relative expression ratio of the beta-defensin gene expression was calculated using the  $\Delta\Delta C_t$  (Cycle threshold) method [133], compared with the average of the two reference genes, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and Eukaryotic elongation factor 2 (Eef2). Here,  $\Delta C_t$  is the difference in  $C_t$  between target and reference and  $\Delta\Delta C_t$  is the difference in  $\Delta C_t$  between all the samples and the calibrator, the sample with the lowest expression. The Rete testes were taken as the calibrator (control) tissue since the defensin genes are not known to be expressed in the rete-testes. They are known to start expressing from the caput epididymis continuing through vas deferens. The differential gene expression levels among the tissues were examined for normality of distribution, transformed where appropriate using log transformation, and analysed by ANOVA and Tukey post-hoc test as implemented in GraphPad Prism 5.0 (for Windows, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com)) and a  $P$ -value  $<0.05$  was considered to be statistically significant. The Kruskal-Wallis non-parametric one-way ANOVA was used to measure the heterogeneity quantified transcripts in bulls & cows. For visualization of clustering of data based on

relative expression, a heat map was plotted by the Clustvis rendering tool [134] using the means for all the replicates for all tissues & genes, to depict the expression pattern of BuBDs across MRT and other body tissues.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12862-019-1535-8>.

**Additional file 1.** Orthologs of BDs. CDS & protein sequence data for 30 BDs in all the species. The data for gene loci, coordinates, orientation etc. is available on request.

**Additional file 2.** Assemblies & novel defensins. CDS & protein sequence data for those BuBDs whose information i.e. the mRNA, CDSs aren't yet listed in the Genbank alongwith the latest RefSeq assemblies from where the sequences were mined.

**Additional file 3.** Results from the MSA (Figure S1) & various tools implemented on the predict-protein server (Figure S2-S7). Also a pictorial representation of the mean expression profiles of these BDs in buffalo tissues (Figure S8).

**Additional file 4.** qRT-PCR optimization. The protocol followed for standardization of the qRT-PCR for the six class-A BuBDs in accordance with MIQE guidelines.

## Abbreviations

AMPs: Antimicrobial peptides; E1: Exon 1; E2: Exon 2; BD/ DEFb: Beta-defensin; BuBD: Buffalo beta-defensin; BUSTED: Branch-Site Unrestricted Statistical Test for Episodic Diversification; CA-BD: Class-A beta-defensin; CCR: Chemokine receptor; CR: Conception rate; EBD: Enteric beta-defensin; eEF2: Eukaryotic elongation factor; FEL: Fixed Effects Likelihood; FRT: Female reproductive tract; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase; GARD: Genetic Algorithm for Recombination Detection; aBSREL: Adaptive Branch-Site Random Effects Likelihood; MEME: Mixed Effects Model of Evolution; GSM: Gene-specific motif; GSM: Gene-specific motif; CMP: Cervical mucus penetration; HDPs: Host defence peptides; HyPhy: Hypothesis testing using Phylogenies; LAP: Lingual antimicrobial peptide; MEGA: Molecular Evolutionary Genetic Analyses; MLM: Maximum likelihood method; MRT: Male reproductive tract; MSA: Multiple sequence alignment; RT: Rete testis; SNP: Single nucleotide polymorphism; SPP: Seminal plasma protein; SSR: Sperm surface remodelling; SVM: Support vector machine; TAP: Tracheal antimicrobial peptide; VD: Vas deferens

## Acknowledgements

The critical inputs of Dr. Suneel Onteru, Sr. Scientist, Animal Biochemistry, NDRI in the final version of the manuscript are gratefully acknowledged.

## Authors' contributions

Study was designed by VB, RK & TKD. Sample collection & expression analysis was done by VB, AM, KD & AS performed phylogenetic analyses, Manuscript was written by VB. Figures were designed by AK. All authors read and approved the final manuscript.

## Funding

This work was supported by the Bill & Melinda Gates Foundation [grant number OPP1154401].

## Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

## Ethics approval and consent to participate

Not Applicable

## Consent for publication

Not Applicable

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Animal Genomics Lab, National Dairy Research Institute, Karnal 132001, India. <sup>2</sup>Theriogenology Lab, SRS of NDRI, Bengaluru 560030, India.

Received: 7 April 2019 Accepted: 22 October 2019

## References

- 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. *J Biol Chem*. 1993;268(9):6611–48.
- Ryan LR, Diamond G. Modulation of Human  $\beta$ -defensin-1 production by viruses. *Viruses*. 2017;9(6):153.
- Raschig J, Mailänder-Sánchez D, Berscheid A, Berger J, Strömstedt AA, Courth LF. Ubiquitously expressed human beta-defensin 1 (hBD1) forms bacteria-entrapping nets in a redox-dependent mode of action. *PLoS Pathol*. 2017;13(3):e1006261.
- Yoo YJ, Kwon I, Oh SR, Perinpanayagam H, Lim SM, Ahn KB, Lee Y, Han SH, Chang SW, Baek SH, Zhu Q, Kum KY. Antifungal effects of synthetic Human Beta-defensin-3-C15 peptide on *Candida albicans*-infected Root Dentin. *J Endod*. 2017;43(11):1857–61.
- Meade KG, Cormican P, Narciandi F, Lloyd A, O'Farrelly C. Bovine  $\beta$ -defensin gene family: opportunities to improve animal health? *Physiol Genomics*. 2014;46(1):17–28.
- Wu Z, Hoover DM, Yang D, Boulègue C, Santamaria F, Oppenheim JJ, Lubkowski J, Lu W. Engineering disulfide bridges to dissect antimicrobial and chemotactic activities of human beta-defensin. *Proc Natl Acad Sci*. 2003;100(15):8880–5.
- Valore EV, Park CH, Quayle AJ, Wiles KR, McCray Jr PB, Ganz T. Human  $\beta$ -defensin-1: an antimicrobial peptide of urogenital tissues. *J Clin Invest*. 1998; 101:1633–42.
- Ryan LK, Dai J, Yin Z, Megjugorac N, Uhlhorn V, Yim S. Modulation of human  $\beta$ -defensin-1 (hBD-1) in plasmacytoid dendritic cells (PDC), monocytes, and epithelial cells by influenza virus, Herpes simplex virus, and Sendai virus and its possible role in innate immunity. *J Leukoc Biol*. 2011;90:343–56.
- McGlasson SL, Semple F, MacPherson H, Gray M, Davidson DJ, Dorin JR. Human  $\beta$ -defensin 3 increases the TLR9-dependent response to bacterial DNA. *Eur J Immunol*. 2017;47(4):658–64.
- Shi N, Jin F, Zhang X, Clinton SK, Pan Z, Chen T. Overexpression of human  $\beta$ -defensin 2 promotes growth and invasion during esophageal carcinogenesis. *Oncotarget*. 2014;5:11333–44.
- Roupé M, Nybo M, Sjöbring U, Alberius P, Schmidtchen A, Sørensen OE. Injury is a major inducer of epidermal innate immune responses during wound healing. *J Invest Dermatol*. 2010;130(4):1167–77.
- Van Kilsdonk JWJ, Jansen PAM, Van den Bogaard EH, Bos C, Bergers M, Zeeuwen PLJM, Schalkwijk J. The Effects of human beta-defensins on skin cells in vitro. *J Dermatol*. 2017;233(2-3):155–63.
- Röhr J, Huber B, Koehl GE, Geissler EK, Hehlgans T. Mouse  $\beta$ -defensin 14 (Defb14) promotes tumor growth by inducing angiogenesis in a CCR6-dependent manner. *J Immunol*. 2012;188(10):4931–9.
- 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. *Biol Reprod*. 2016;95(6):122.
- Légaré C, Akintayo A, Blondin P, Calvo E, Sullivan R. Impact of male fertility status on the transcriptome of the bovine epididymis. *Mol Hum Reprod*. 2017;23(6):355–69.
- Yang D, Liu ZH, Tewary P, Chen Q, de la Rosa G, Oppenheim J. *Curr Pharm Des*. 2007;13(30):3131–9.
- Funderburg NT, Jadlowsky JK, Lederman MM, Feng Z, Weinberg A, Sieg SF. The Toll-like receptor 1/2 agonists Pam (3) CSK(4) and human beta-defensin-3 differentially induce interleukin-10 and nuclear factor-kappaB signalling patterns in human monocytes. *Immunology*. 2011;134:151–60.
- Semple F, Webb S, Li HN, Patel HB, Perretti M, Jackson IJ, Gray M, Davidson DJ, Dorin JR. Human beta-defensin 3 has immunosuppressive activity in vitro and in vivo. *Eur J Immunol*. 2010;40:1073–8.

19. Fusco A, Savio V, Cammarota M, Alfano A, Schiraldi C, Donnarumma G. Beta-Defensin-2 and Beta-Defensin-3 reduce intestinal damage caused by *Salmonella typhimurium* modulating the expression of cytokines and enhancing the probiotic activity of *Enterococcus faecium*. *J Immunol Res*. 2017. <https://doi.org/10.1155/2017/6976935>.
20. Feng J, Luo J, Mack MR, Yang P, Zhang F, Wang G, Gong X, Cai T, Mei Z, Kim BS, Yin S, Hu H. The antimicrobial peptide hBD2 promotes itch through Toll-like receptor 4 signaling in mice. *J Allergy Clin Immunol*. 2017;140(3):885–888.e6.
21. Otri AM, Mohammed J, Al-Aqaba MA, Fares U, Peng C, Hopkinson A, et al. Variable expression of human beta-defensins 3 and 9 at the human ocular surface in infectious keratitis. *Invest Ophthalmol Vis Sci*. 2012;53:757–61.
22. Semlali A, Witold C, Alanazi M, Rouabhia M. Whole cigarette smoke increased the expression of TLRs, HBDs, and proinflammatory cytokines by human gingival epithelial cells through different signaling pathways. *PLoS One*. 2012;7(12):e52614.
23. Lütjhe P, Hirschberg AL, Brauner A. Estrogenic action on innate defense mechanisms in the urinary tract. *Maturitas*. 2014;77(1):32–6.
24. Pujianto DA, Loanda E, Sari P, Midoen YH, Soeharso P. Sperm-associated antigen 11A is expressed exclusively in the principal cells of the mouse caput epididymis in an androgen-dependent manner. *Reprod Biol Endocrinol*. 2013;11:59.
25. Rengaraj D, Truong AD, Lillehoj HS, Han JY, Hong YH. Expression and regulation of avian beta-defensin 8 protein in immune tissues and cell lines of chickens. *Asian-Australas J Anim Sci*. 2018;31(9):1516–24.
26. Sørensen OE, Thapa DR, Rosenthal A, Liu L, Roberts AA, Ganz T. Differential regulation of  $\beta$ -defensin expression in human skin by microbial stimuli. *J Immunol*. 2005;174(8):4870–9.
27. Semple F, Dorin JR. Beta-defensins: multifunctional modulators of infection, inflammation and more? *J Innate Immun*. 2012;4:337–48.
28. Patil AA, Cai Y, Sang Y, Blecha F, Zhang G. Cross-species analysis of the mammalian beta-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. *Physiol Genomics*. 2005;23:5–17.
29. Bagnicka E, Strzałkowska N, Szreder T, Prusak B, Józwick A, Kościuczuk E, Krzyżewski J, Zwierzchowski L. A/C polymorphism in the  $\beta$ -4 defensin gene and its association with phenotypic and breeding values of milk production traits in Polish-Friesian cows. *Anim Sci Papers and Rep*. 2008;26(4):239–50.
30. Narciandi F, Lloyd AT, Chapwanya A, O'Farrelly C, Meade KG. Reproductive tissue-specific expression profiling and genetic variation across a 19 gene bovine  $\beta$ -defensin cluster. *Immunogenetics*. 2011;63:641–51.
31. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res*. 2012;22(4):778–90.
32. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, et al. Analysis of copy number variations among diverse cattle breeds. *Genome Res*. 2010;20(5):693–703.
33. Li P, Chan HS, He B, So SC, Chung YW, Shang Q, Zhang YD, Zhang YL. An Antimicrobial peptide gene found in the male reproductive system of rats. *Science*. 2001;291(5509):1783–1785.
34. Belleannee C, Thimon V, Sullivan R. Region-specific gene expression in the epididymis. *Cell Tissue Res*. 2012;349(3):717–31.
35. Browne JA, Yang R, Leir SH, Eggner SE, Harris A. Expression profiles of human epididymis epithelial cells reveal the functional diversity of caput, corpus and cauda regions. *Mol Hum Reprod*. 2016;22(2):69–82.
36. Ribeiro CM, Silva EJR, Hinton BT. Avellar MCW  $\beta$ -defensins and the epididymis: contrasting influences of prenatal, postnatal, and adult scenarios. *Asian J Androl*. 2016;18(2):323–8.
37. Rich TD, Turman EJ. Reproductive tract anatomy and physiology of the cow. In: *Beef cattle Handbook publication BCH-2200 Stillwater: Oklahoma State University*. 2014. <http://www.iowabeefcenter.org/bch/CowReproductiveAnatomy>. Accessed 15 Jan 2019.
38. Tollner TL, Bevins CL, Cherr GN. Multifunctional glycoprotein DEF126—a curious story of defensin-clad spermatozoa. *Nat Rev Urol*. 2012;9(7):365–75.
39. Dorin JR, Barratt CL. Importance of  $\beta$ -defensins in sperm function. *Mol Hum Reprod*. 2014;20(9):821–6.
40. Whiston R, Finlay EK, McCabe MS, Cormican P, Flynn P, Cromie A, Hansen PJ, Lyons A, Fair S, Lonergan P, O'Farrelly C, Meade KG. A dual targeted  $\beta$ -defensin and exome sequencing approach to identify, validate and functionally characterise genes associated with bull fertility. *Sci Rep*. 2017;7(1):12287.
41. Colledge WH. Defending sperm function. *PLoS Genet*. 2013;9(10):e1003889.
42. Tollner TL, Yudin AI, Treece CA, Overstreet JW, Cherr GN. Macaque sperm coating protein DEF126 facilitates sperm penetration of cervical mucus. *Hum Reprod*. 2008;23(11):2523–34.
43. 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. *Biol Reprod*. 2005;73(6):1243–52.
44. Lyons A, Narciandi F, Donnellan E, Romero-Aguirregomezcorra J, O'Farrelly C, Lonergan P, Meade KG, Fair S. Recombinant  $\beta$ -defensin 126 promotes bull sperm binding to bovine oviductal epithelia. *Reprod Fert Dev*. 2018;30(11):1472–81.
45. Yudin AI, Tollner TL, Li MW, Treece CA, Overstreet JW, Cherr GN. ESP13.2, a member of the  $\beta$ -defensin family, is a macaque sperm surface-coating protein involved in the capacitation process. *Biol Reprod*. 2003;69(4):1118–28.
46. 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. *Mol Reprod Dev*. 2004;69(3):325–37.
47. Dyck SM, Polu SA, Julliard V, Babiuk LA, Hurk SD. The synthetic peptides bovine enteric  $\beta$ -defensin (EBD), bovine neutrophil  $\beta$ -defensin (BNBD) 9 and BNBD 3 are chemotactic for immature bovine dendritic cells. *Veterinary Immunol Immunopathol*. 2011;143(1–2):87–107.
48. Cormican P, Meade KG, Cahalane S, Narciandi F, Chapwanya A, Lloyd AT, O'Farrelly C. Evolution expression and effectiveness in a cluster of novel bovine beta-defensins. *Immunogenetics*. 2008;60(3–4):147–56.
49. Yudin AI, Tollner TL, Treece CA, et al. Beta-defensin 22 is a major component of the mouse sperm glycocalyx. *Reproduction*. 2008;136(6):753–65.
50. Meade KG. Farrelly  $\beta$ -Defensins: Farming the microbiome for homeostasis and health. *Front Immunol*. 2019. <https://doi.org/10.3389/fimmu.2018.03072>.
51. Dejuçq N, Jegou B. Viruses in the mammalian male genital tract and their effects on the reproductive system. *Microbiol Mol Biol Rev*. 2001;65:208–31.
52. Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libbertson B, et al. Antagonistic coevolution accelerates molecular evolution. *Nature*. 2010;464:275–8.
53. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A*. 1998;95(12):6578–83.
54. Machado LR, Ottoloni B. An evolutionary history of defensins: a role for copy number variation in maximizing host innate and adaptive immune responses. *Front Immunol*. 2015;6:115.
55. Luenser K, Ludwig A. Variability and evolution of bovine  $\beta$ -defensin genes. *Genes and Immunity*. 2005;6:115–22.
56. 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  $\beta$ -Defensin peptides in vertebrates. *Evolutionary Bioinformatics*. 2015;11:105–14.
57. Cobo ER, Chadee K. Antimicrobial Human  $\beta$ -Defensins in the colon and their role in infectious and non-infectious diseases. *Pathogens*. 2013;2(1):177–92.
58. Bauer F, Schweimer K, Klüver E, Conejo-Garcia JR, Forssmann WG, Rösch P, Adermann K, Sticht H. Structure determination of human and murine  $\beta$ -defensins reveals structural conservation in the absence of significant sequence similarity. *Protein Sci*. 2001;10(12):2470–9.
59. Klüver E, Schulz-Maronde S, Scheid S, Meyer B, Forssmann WG, Adermann K. Structure-activity relation of human beta-defensin 3: influence of disulfide bonds and cysteine substitution on antimicrobial activity and cytotoxicity. *Biochemistry*. 2005;44(28):9804–16.
60. Cornwall GA. Role of posttranslational protein modifications in epididymal sperm maturation and extracellular quality control. *Adv Exp Med Biol*. 2014;759:159–80.
61. Tollner TL, Yudin AI, Tarantal AF, Treece CA, Overstreet J, Cherr GN. Beta-defensin 126 on the surface of macaque sperm mediates attachment of sperm to oviductal epithelia. *Biol Reprod*. 2008;78:400–12.
62. Tollner TL, Venners SA, Hollox EJ, Yudin AI, Liu X, Tang G, Xing H, Kays RJ, Lau T, Overstreet JW, Xu X, Bevins CL, Cherr GN. A common mutation in DEF126 causes impaired sperm function and subfertility. *Sc Transl Med*. 2011;3(92):92ra65.
63. Teclé E, Gagneux P. Sugar-coated sperm: Unraveling the functions of the mammalian sperm glycocalyx. *Mol Reprod Dev*. 2015;82(9):635–50.
64. Malm J, Sorensen O, Persson T, Frohm-Nilsson M, Johansson B, Bjartell A, Lilja H, Stahle-Backdahl M, Borregaard N, Egesten A. The human cationic antimicrobial protein (hCAP-18) is expressed in the epithelium of human

- epididymis, is present in seminal plasma at high concentrations, and is attached to spermatozoa. *Infect Immun.* 2000;68:4297–302.
65. Rodriguez-Jimenez FJ, Krause A, Schulz S, Forssmann WG, Conejo-Garcia JR, Schreeb R, Motzkus D. Distribution of new human beta-defensin genes clustered on chromosome 20 in functionally different segments of epididymis. *Genomics.* 2003;81:175–83.
  66. Dorin JR, McHugh BJ, Cox SL, Davidson DJ. Mammalian antimicrobial peptides; defensins and cathelicidins. In: Yi-Wei Tang, Max Sussman, Joseph Schwartzman, editors. *Molecular Medical Microbiology (Second Edition)*. Academic Press; 2015. p. 539–565.
  67. Zhao Y, Diao H, Ni Z, Hu S, Yu H, Zhang Y. The epididymis-specific antimicrobial peptide  $\beta$ -defensin 15 is required for sperm motility and male fertility in the rat (*Rattus norvegicus*). *Cell Mol Life Sci.* 2010;68:697–708.
  68. Zhou CX, Zhang YL, Xiao L, Zheng M, Leung KM, Chan MY, Lo PST, Sang LL, Wong HY, Ho LS, Chung YW, Chan HC. An epididymis-specific  $\beta$ -defensin is important for the initiation of sperm maturation. *Nat Cell Biol.* 2004;6:458–64.
  69. Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, Gutenkunst RN, White TJ, Green ED, Bustamante CD. Targets of balancing selection in the human genome. *Mol Biol Evol.* 2009;26:2755–64.
  70. Barreiro LB, Quintana-Murcil. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 2010;11:17–30.
  71. Zhu S, Gao B. Evolutionary origin of  $\beta$ -defensins. *Dev Comp Immunol.* 2013;39(1–2):79–84.
  72. Zhang ZY, Zhang HM, Li DS, Xiong TY, Fang SG. Characterization of the  $\beta$ -defensin genes in giant panda. *Sci Rep.* 2018;8:12308.
  73. Narciandi F, Lloyd A, Meade KG, O'Farrelly C. A novel subclass of bovine b-defensins links reproduction and immunology. *Reprod Fert Dev.* 2014;26(6):769–77.
  74. Mukherjee S, Ganguli D, Majumder PP. Global footprints of purifying selection on Toll-like receptor genes primarily associated with response to bacterial infections in humans. *Genome Biol Evol.* 2014;6:551–8.
  75. Beutler B. Innate immunity: an overview. *Mol Immunol.* 2004;40:845–59.
  76. Majumder PP. Pathogen pressure and molecular evolutionary genetics of innate immunity genes in humans. In: *Nature at work: ongoing saga of evolution*. New Delhi: Springer India; 2010. p. 249–65.
  77. Ferrer-Admetlla A, Bosch E, Sikora M, Marqués-Bonet T, Ramírez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J. Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol.* 2008;181:1315–22.
  78. Chapman JR, Hellgren O, Helin AS, Kraus RH, Cromie RL, Waldenström J. The evolution of innate immune genes: purifying and balancing selection on  $\beta$ -defensins in waterfowl. *Mol Biol Evol.* 2016;33(12):3075–308.
  79. Hollox EJ, Armour JA. Directional and balancing selection in human beta-defensins. *BMC Evol Biol.* 2008;8:113.
  80. Semple CA, Maxwell A, Gautier P, Kilanowski FM, Eastwood H, Barran PE, et al. The complexity of selection at the major primate  $\beta$ -defensin locus. *BMC Evol Biol.* 2005;5:32.
  81. Cheng Y, Prickett MD, Gutowska W, Kuo R, Belov K, Burt DW. Evolution of the avian  $\beta$ -defensin and cathelicidin genes. *BMC Evol Biol.* 2015;15:188.
  82. Hurles M. Gene duplication: the genomic trade in spare parts. *Plos Biol.* 2004;2:e206.
  83. Hollox EJ, Abujaber R. Evolution and diversity of defensins in vertebrates. In: Pontarotti P, editor. *Evolutionary biology: self/nonself evolution, species and complex traits evolution, methods and concepts*. Springer International Publishing; 2017. p.27–50.
  84. Tennessen JA, Blouin MS. Selection for antimicrobial peptide diversity in frogs leads to gene duplication and low allelic variation. *J Mol Evol.* 2007;65:605–15.
  85. Lazzaro BP. Natural selection on the *Drosophila* antimicrobial immune system. *Curr Opin Microbiol.* 2008;11:284–9.
  86. Viljakainen L, Pamilo P. Selection on an antimicrobial peptide defensin in ants. *J Mol Evol.* 2008;67:643–52.
  87. Ghatge M, Sharma A, Maity S, Kakkar VV, Vangala RK. Danger-recognizing proteins,  $\beta$ -defensin-128 and histatin-3, as potential biomarkers of recurrent coronary events. *Int J Mol Med.* 2017;40:531–8.
  88. Choi MY, Le MT, Nguyen DT, Choi H, Kim W, Kim JH, Chun J, Hyeon J, Seo K, Park C. Genome-level identification, gene expression, and comparative analysis of porcine  $\beta$ -defensin. *BMC Genetics.* 2012;13:98.
  89. Guo M, Wu F, Hao G, Qin Q, Li R, Li N, Wei L, Chai T. *Bacillus subtilis* improves immunity and disease resistance in rabbits. *Front Immunol.* 2017;8:354.
  90. Johnson GP, Lloyd AT, O'Farrelly C, Meade KG, Fair S. Comparative genomic identification and expression profiling of a novel b-defensin gene cluster in the equine reproductive tract. *Rep Fer Dev.* 2016;28:1499–508.
  91. Narciandi F, Fernandez-Fuertes B, Khairulzaman I, Jahns H, King D, Finlay EK, Mok KH, Fair S, Lonergan P, O'Farrelly C, Meade KG. Sperm-coating  $\beta$ -defensin 126 is a dissociation-resistant dimer produced by epididymal epithelium in the bovine reproductive tract. *Biol Reprod.* 2016;95(6):121.
  92. Jelinsky SA, Turner TT, Bang HJ, Finger JN, Solar MK, Wilson E, Brown EL, Kopf GS, Johnston DS. The rat epididymal transcriptome: comparison of segmental gene expression in the rat and mouse epididymides. *Biol Reprod.* 2007;76:561–70.
  93. Basic local alignment search tool.1990. [http:// www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST). Accessed 5th Dec 2018.
  94. GenBank. 2005 [http:// www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank). Accessed 18th Sep 2018.
  95. ExPASy: the proteomics server for in-depth protein knowledge and analysis. 2003. [http:// www.expasy.org](http://www.expasy.org). Accessed 20th Dec 2018.
  96. The PredictProtein server. <http://www.predictprotein.org> Accessed 24th Dec 2018.
  97. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variant. *BMC Genomics.* 2015;16(Suppl 8).
  98. Schlessinger A, Punta M, Yachdav G, Kajan L, Rost B. Improved Disorder Prediction by Combination of Orthogonal Approaches. *Plos one.* 2009;4(2):e4433.
  99. Hamp T, Kassner R, Seemayer S, Vicedo E, Schaefer C, Achten D, Auer F, Boehm A, Braun T, Hecht M, Heron M, Honigschmid P, Hopf TA, Kaufmann S, Kiening M, Denis K, Landerer C, Mahlich Y, Roos M, Rost B. Homology-based inference sets the bar high for protein function prediction. *BMC Bioinformatics (Suppl. Automated Function Prediction Meeting 2011)*. 2013; 14: Suppl. S3–S7
  100. Ofra Y, Rost B. ISIS: interaction sites identified from the sequence. *Bioinformatics.* 2007;23:e13–6.
  101. Hönigschmid P, Frishman D. Accurate prediction of helix interactions and residue contacts in membrane proteins. *J Struct Biol.* 2016;194(1):112–123.
  102. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, Brunak S, Wandall HH, Lavery SB, Clausen H. Precision mapping of the human O-GalNAc glycoproteome through simple cell technology. *EMBO J.* May 15,2013;32(10):1478–1488.
  103. Gupta R. Prediction of N-glycosylation sites in human proteins. 2004. [http:// www.cbs.dtu.dk/services/NetNGlyc/](http://www.cbs.dtu.dk/services/NetNGlyc/). Accessed 25 Dec 2018.
  104. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol.* January 2001;305(3):567–80.
  105. Waghu FH, Barai RS, Gurusu P, Idicula-Thomas S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic acids research.* 2015;44(D1):D1094–D1097. CAMPR3: A database on sequences, structures and signatures of antimicrobial peptides. 2015. [http:// www.camp3.bicnirrh.res.in/predict/](http://www.camp3.bicnirrh.res.in/predict/). Accessed 24th Dec 2018.
  106. PDB. Announcing the worldwide Protein Data Bank Nature Structural Biology. 2003. <https://www.rcsb.org/>. Accessed 19th Feb 2019.
  107. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296–303.
  108. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605–12.
  109. Chenn VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. All-atom structure validation for macromolecular crystallography. *Acta Cryst.* 2010;66:12–21.
  110. Katoh M, Kuma M. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66.
  111. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2- A multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25:118991.

112. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignments of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 2010;38:7–13.
113. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547–9.
114. Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. 1st ed. New York: Oxford University Press; 2000.
115. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.* 1985;39:783–91.
116. Kosakovsky Pond SL, et al. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol.* 2006;23:1891–901.
117. Kishino H, Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominioidea. *J Mol Evol.* 1989;29(2):170–9.
118. Datamonkey. <https://www.datamonkey.org>. Accessed December 2018.
119. Smith MD, et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol.* 2015;32:1342–53.
120. Wertheim JO, et al. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 2015;32:20–832.
121. Pond KSL, Frost SDW. Not So Different After All: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005;22:1208–22.
122. Murrell B, et al. Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics.* 2012;8(7):e1002764.
123. Murrell B, et al. Gene-wide identification of episodic selection. *Mol Biol Evol.* 2015;32:1365–71.
124. Kumar S, Gadagkar SR. Disparity Index: A simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics.* 2001;158:1321–7.
125. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol and Evol.* 1986;3:418–26.
126. Zhang J, Kumar S, Nei M. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Mol Biol and Evol.* 1997; 14:1335–8.
127. Tajima F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123:585–95.
128. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics.* 2012;13:134.
129. Kibbe WA. An online oligonucleotide properties calculator. *Nucleic Acids Res.* 2007;35(Suppl 2):W43–6.
130. UCSC Genome Browser, Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res.* 2002;6:996–1006.
131. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Briefings in Bioinformatics.* 2013;14(2):144–61.
132. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The MIQE Guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009;55(4):611–22.
133. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔCt</sup> Method. *Methods.* 2001;25:402–8.
134. Metsalu T, Vilo J. Clustvis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nuc Acids Res.* 2015;43(1):566–70.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

