Bulletin of the National Research Centre

RESEARCH Open Access

Intron retention in *Cathelicidin-4* in river buffalo

Ahlam A. Abou Mossallam*, Soheir M. El Nahas, Esraa A. Balabel, El Hassan T. Shahwan and Noha M. Osman

Abstract

Background: The function of cathelicidins-4 *(CATH4)* is not limited to microbial killing, but extends to other aspects of immunity and tissue repair. The presence of different *CATH4* variants including intron retention affects the immunity system. Intron retention, in buffalo, is not fully studied. In this study, we investigated *CATH4* mRNA in river buffalo and their variants, which can be used in the future for selecting buffalo resistant to diseases.

Results and conclusion: Analysis of *CATH4* mRNA in river buffalo (Egyptian breed) revealed the presence of a novel variant (1073 bp) which includes unspliced part of intron 3 (469 bp) in addition to previously reported unspliced complete intron 1 (103) and intron 2 (137 bp). Identification of intron retention was conducted by comparing the amplified unspliced cDNA and DNA sequences. Analysis of the 3 retained intronic regions revealed the presence of the 4 splice signals, needed for splicing which include the 5′ (GT) and 3′ (AG) intron splice sites, the branch point, and the polypyrimidine tract. However, in the intron-retained sequence, the polypyrimidine tract was weak. It contained 6 and 4 non-continuous uridine stretch in introns 1 and 2, respectively, (intron 3 was partial) which may have caused introns retention. In addition, analysis of the unspliced sequence showed three unique exonic SNPS located close to the splice sites (1 to 22 nucleotides) and five SNPs in retained intronic regions located near the splice sites (18 to 246 nucleotides away from exon/intron boundaries) which may be related to the retention of the three introns.

Keywords: Egyptian buffalo, *Cathelicidine-4*, *CATH4* splice sites, cDNA, SNPs

Background

Cathelicidin-4 (CATH4), popularly known as indolicidin, has broad and rapid microbicidal effect that may be critically important to clear tissues from pathogens and to prevent the onset of infection (Dorschner et al. 2001). The function of cathelicidins is not limited to microbial killing, but extends to other aspects of immunity and tissue repair (Gallo et al. 2002). CATH4 contains 4 exons and 3 introns: the first 3 exons comprise signal peptide and cathelin prodomain (N-terminal) while the fourth exon encodes the cleavage site and variable C-terminal antimicrobial peptide (Zanetti et al. 2000; Zaiou and Gallo 2002).

Water buffalo (*Bubalis bubalis*) population includes river (*Bubalus bubalis bubalis*) and swamp buffalo (*Bubalus bubalis carabanesis*), 77% of which are river buffalo (FAO, 2013). Buffalo are major source of meat, milk, and its by-products. Buffalo surpass the cattle in its ability to adapt to the hot, humid areas of muddy and swampy lands (Marai and Habeeb 2010). Buffalo *CATH4* was cloned by Das et al. (2006). The complete *CATH4* coding region was found to be 92.9% similar to *Bos taurus* nucleotide sequence. Several *cathelicidin* genes were identified in cattle and buffaloes. In cattle, SNPs, insertions, and deletions have been reported in different breeds of *Bos taurus* and *Bos indicus* (Gillenwaters et al. 2009). In a study by Brahma et al. (2015), amplicons of *cathelicidin* genes of 5 breeds of cattle and buffalo were investigated. Buffalo *CATH4* genes showed higher single-nucleotide variations compared to cattle genes.

Constitutive splicing of intronic sequences from RNA is the dominant form of gene expression. However, alternative splicing leading to intron retention (IR) has been reported in many bovine genes (Chacko and Ranganathan 2009). Examples of IR have been found in bovine *growth hormone* (Dirksen et al. 1995), *CD46* (Wang et al. 2014), and *NCF4* (Ju et al. 2015). A higher relative frequency of IR has been associated with genes with overall shorter intron lengths (~ 100–200 nt), higher expression levels, weaker splice sites, and particular densities of *Cis*-regulatory elements (Sakabe and de Souza 2007). Recent evidences suggest that single-

* Correspondence: ahlammasry@yahoo.com
Cell Biology Department, Genetic Engineering and Biotechnology Division, National Research Centre, Dokki, Cairo 11622, Egypt

Table 1 Sequences of primer pairs used

Primer number	*CATH4*	5′-sequence-3′	Target length	Annealing temp.	Accession no.
1	F	CATGCAGACCCAGAGGGCCA	502 bp	62 °C	gb|EF050453.1|
	R	ATCAGACACTTAGGACTCTTCCCCG			
2	F	GGGCCATCCTCTCGTTGG	607 bp	54 °C	XM_006065186.1
	R	GGAGCATGTGGTTGAATC			

nucleotide polymorphisms (SNPs) are the main factor that contribute to the generation of alternative splice variants, which can cause degenerative axonopathy (Drögemüller et al. 2011) and a congenital mechanobullous skin disorder (Menoud et al. 2012) in cattle.

In this study, we investigated the different splice variants of mRNA *CATH4* in river buffalo (Egyptian breed) which can be used for selecting disease-resistant breeds of buffalo.

Materials and methods

1. Collecting buffalo samples

Blood samples collected on ethylene diamine tetra acetic acid (EDTA) from healthy river buffalo (Egyptian breed) were kindly provided by the veterinarian of the buffalo farm "United Farms Group Company."

2. DNA extraction

Genomic DNA was extracted from whole blood using salting out method according to Miller et al. (1988). The DNA concentrations were measured using Nanodrop 1000 (Thermoscientific) and were adjusted to 50 ng/μL for polymerase chain reaction (PCR).

3. mRNA extraction and cDNA synthesis

Total RNA from blood was extracted using Easy–RED™ iNtRON Biotechnology, Inc. according to the manufacturer's instructions. RNA was considered to be free of DNA and proteins with a 260/280 optical density ratio of ~ 2.0. cDNA synthesis was performed using Revert Aid First Strand Synthesis Kit according to the manufacturer's instructions. To ensure that the RNA was not contaminated with genomic DNA, a PCR reaction was performed using RNA in absence of reverse transcriptase, as a negative control.

4. Primers design

Two primer pairs were designed to investigate *CATH4* in Egyptian buffalo, using Primer3 software (Untergasser and Cutcutache, 2012). Table 1 presents the primer pair sequences and the accession number from which they were designed.

5. PCR amplification

PCR amplification for *CATH4* was performed in 50 μl reaction volume which includes 20 μl of water (nuclease-free), 25 μl of PCR Master Mix (2X), 1 μl of forward primer (10 μM), 1 μl of reverse primers (10 μM), and 4 μl of 50 ng of DNA or cDNA template. The reaction mixture was run in a Q-Cycler, HVD LifeSciences. The thermal cycling program was initial denaturation at 95 °C for 3 min followed by 40 cycles of the following: denaturation at 95 °C for 30s, annealing for 30s (54 °C or 62 °C), extension at 75 °C for 1 min, and then final extension at 75 °C for 10 min. Detection of PCR products were performed by agarose gel electrophoresis according to the method described by Ausubel et al. (1990). The gels were inspected by Gel documentation system (In Genius, Syngenebioimaging). PCR products were purified using MEGAquick-spin™ Total Fragment DNA Purification Kit (iNtRON biotechnology) according to the kit's instructions.

6. Sequence analysis and SNPs identification

Purified PCR products were sequenced by Macrogen (Korea) using reverse and forward primers. The specificity of the nucleotide sequences were verified by BLAST analysis (Basic Local Alignment Search Tool) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al. 1990). Sequences

ATGCAGACCCAGAGGGCCATCCTCKYGTTGGGGCGGTGGTCACYGTGGCTTCTGCTGCTGGGGCTTG
TGGTGTCCTCGACCAGYGCCCAGGACCTCAGCTACAGGGARGCCGTGCTTCGTGCTGTGGATCAGCT
CAATGAGCGGTCYTCAGAAGCTAATCTCTACCGCCTCCTGGAGCTAGAMCCRCCTCCCAAGGAT**GAT**
GCAGATCTGGGCACTCGAAAGCCTGTGAGCTTCACGGTGAAGGAGACTGTGTGCCCCAGGACGAC
TCAGCAGCCYRCGGAGCRGTGTGACTTCAAGGAGGAAGGGCGGGTGAARCAGTGTGTGGGGACA
GTCACCCTGGACCCGTCCAATGACCAGTTTGACCTAAACTGTAATGAG**CTCCAGAGTGTCAGGATAC**
GCTTTCCATGGCCATGGCCATGGCCATGGTGGCGCAGATTCCGAGGTTGA*TGGACAAGAACTGTTG*
GATCCCRAGCCCGGGGAAGAGTCCWAAGTGTCTGATGTTTGTTCAGATTCGGGCTTCTGGACA

Fig. 1 Buffalo *CATH4* full-spliced cDNA sequence. Exons 1 and 3 are underlined, and 2 and 4 are in bold. 3′ Untranslated region is in italic. SNPs: M (A/C), R (G/A), K (T/G), Y(C/T), and W (T/A)

cctctcgttggggcggtggtcactgtggcttctgctgctgggcctagtggtaccctcggccagcgcccaggacctcag
ctacagggaggccgtgcttcgtgctgtggatcagctcaatgagcggtcctcagaagctaatctctaccgcctcctgg
wgctagaccc**r**cctc**y**caaggat***GT****GAGTTGGGGAGGGGACTG****Y****CTAGGTGAGGGGCAGGG*
AGACAGATCAGAGAAGGAAAAATGAGCCTGAACCCAGTTTCCCCCGCACTTTAATCCGT
*CACC****AG***gatgcagatctgggcactcgaaagcctgtgagcttcacggtgaaggagactgtgtgccccaggacgac
tcagcagcc**yr**cggagc**r**gtgtgacttcaaggagaaaggg***GT****GAGGCTGGGGGTTGGGGGTCAATG*
*TTTCC****Y****AGGGAGCTGAACAGGGAGCTTCTGGGA****W****GGTTTCCTGTCTCTGGGGTGAGGC*
*TGGGAGGTTATGGCCAAGGGGATTCCAGTTTGACCTTGAGCCTCTCCTTCC****AG***cgggtgaa
rcagtgtgtggggacagtcaccctggacccgtccaatgaccagtttgacctaaactgtaatg**m**g***GT****GAGTGGT*
CCCTTCTGGACTGGGGGGTTTCTAGGGAAGATAGTGTGTGGAACATCCTTTGTACCAAT
*GACCCGCTGTCCCATCCAGGGCAGA****S****AGAGGCCCTCCTACCCTGGCCCCTCCTTCCCTAA*
*GCCCCAGGTCTCCAGCCCTGGGGCTGCC****T****CCCTTAGAGCAGTGGTCCTCTACTGGGGTTC*
CCATCTGGGAACTGACATGAGATGGATTCTCAGGACCCACTCGGACTTCCTGAATCTGA
*C****W****CTGGAGTGGGGCCCAGCCATTTGGATTTTCCCAAGACCTCCAGGGAATTGTGACTG**
AGCTCAGCTTACGACACACTGACTGCAGTCGTGGGTTTTCAGCCTCTGCTCTCATCCAGC
TTTGCTGGAATGTGCTTGTGACCCTGGGAAGCCCCATGCCATCTGTGGCCTCAGTTTTCC
TGTCCGTCTGGGGTAGGGAATCAACCCCCCCTTTCCCAAAAAGG

Fig. 2 Buffalo *CATH4* unspliced cDNA sequence. Starting position is c.21. Introns are in italic with the 5′ and 3′ splice site (GT-AG). Exons 1, 2 and 3 are in small letters. SNPs: M(A/C), R(G/A), K(T/G), Y(C/T), S(G/C), and W(T/A)

a

Score	Expect	Identities	Gaps	Strand	Frame
273 bits(302)	2e-77()	165/179(92%)	0/179(0%)	Plus/Plus	

```
Spliced    21 CCTCKYGTTGGGGCGGTGGTCACYGTGGCTTCTGCTGCTGGGGCTTGTGGTGTCCTCGAC  80
              ||||  |||||||||||||||| |||||||||||||||||| || |||||  ||||| |
Unspliced   1 CCTCTCGTTGGGGCGGTGGTCACTGTGGCTTCTGCTGCTGGGCCTAGTGGTACCCTCGGC  60

Spliced    81 CAGYGCCCAGGACCTCAGCTACAGGGARGCCGTGCTTCGTGCTGTGGATCAGCTCAATGA  140
              ||| ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Unspliced  61 CAGCGCCCAGGACCTCAGCTACAGGGAGGCCGTGCTTCGTGCTGTGGATCAGCTCAATGA  120

Spliced   141 GCGGTCYTCAGAAGCTAATCTCTACCGCCTCCTGGAGCTAGAMCCRCCTCCCAAGGATG  199
              |||||| |||||||||||||||||||||||||| |||||| ||||||| ||||||||
Unspliced 121 GCGGTCCTCAGAAGCTAATCTCTACCGCCTCCTGGWGCTAGACCCRCCTCYCAAGGATG  179
```

b

Score	Expect	Identities	Gaps	Strand	Frame
183 bits(202)	3e-50()	107/108(99%)	0/108(0%)	Plus/Plus	

```
Spliced   199  GATGCAGATCTGGGCACTCGAAAGCCTGTGAGCTTCACGGTGAAGGAGACTGTGTGCCCC  258
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Unspliced 282  GATGCAGATCTGGGCACTCGAAAGCCTGTGAGCTTCACGGTGAAGGAGACTGTGTGCCCC  341

Spliced   259  AGGACGACTCAGCAGCCYRCGGAGCRGTGTGACTTCAAGGAGGAAGGG  306
               |||||||||||||||||||||||||||||||||||||||||||||| |||||
Unspliced 342  AGGACGACTCAGCAGCCYRCGGAGCRGTGTGACTTCAAGGAGAAAGGG  389
```

c

Score	Expect	Identities	Gaps	Strand	Frame
127 bits(140)	2e-33()	72/73(99%)	0/73(0%)	Plus/Plus	

```
Spliced   307 GCGGGTGAARCAGTGTGTGGGGACAGTCACCCTGGACCCGTCCAATGACCAGTTTGACCT  365
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Unspliced 526 GCGGGTGAARCAGTGTGTGGGGACAGTCACCCTGGACCCGTCCAATGACCAGTTTGACCT  585

Spliced   366 AAACTGTAATGAG  378
              |||||||||| |
Unspliced 586 AAACTGTAATGMG  598
```

Fig. 3 Alignment between the spliced and unspliced *CATH4* amplicons showing three separate matches which correspond to exons 1 (**a**), 2 (**b**), and 3 (**c**)

were analyzed by multiple alignments using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Polymorphic sites were determined by visual examination of sequence's charts.

7. Determining the four potential splice sites

To predict the potential splice sites 5′ and 3′ (GT/AG), the Splice Port tool (http://spliceport.cbcb.umd.edu/) was used. Prediction of the branch points (BPs) and

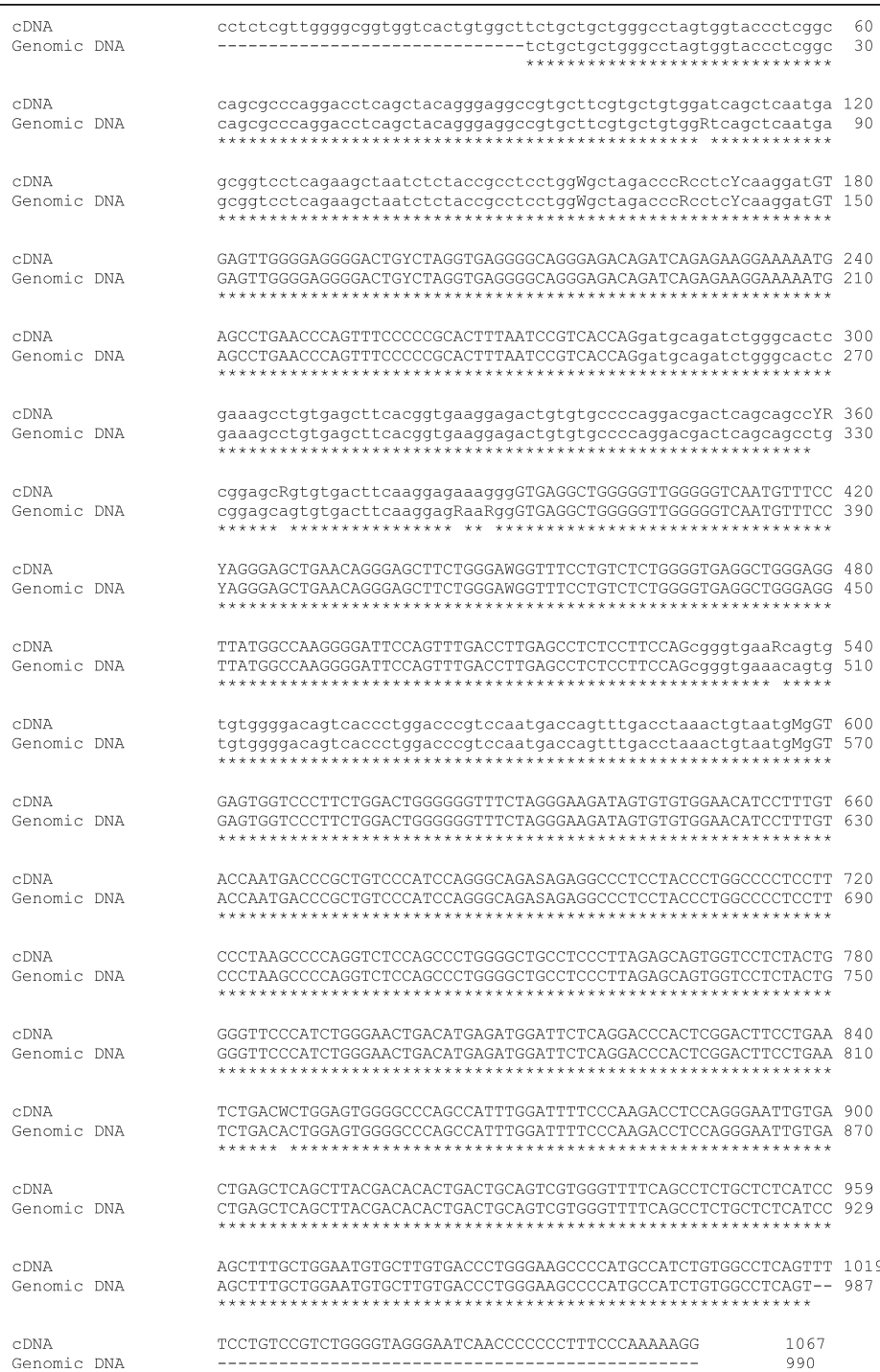

Fig. 4 Multiple sequence alignment of buffalo *CATH4* genomic DNA and 1067 bp cDNA. Introns are in capital; exons 1, 2, and 3 are in small letters. SNPs: M(A/C), R(G/A), K(T/G), Y(C/T), S(G/C), and W(T/A)

Table 2 The potential predicted 5′ and 3′ splice sites (GT-AG) in Buffalo *CATH4* unspliced cDNA sequence

Donor/acceptor splice site		Short sequence	FGA score
Donor (c.198+g.1)	5′ Intron 1	aggatGTgagtt	1.0048
Acceptor (c.199–g.1)	3′ Intron 1	tcaccAGgatgc	0.5185
(Donor (c.306+g.1)	5′ Intron 2	aagggGTgaggc	0.8146
Acceptor (c.307–g.1)	3′ Intron 2	cttccAGcgggt	0.7447
Donor (c.378+g.1)	5′ Intron 3	atgagGTgagtg	1.5098

Exonic region is in small letter, and intronic region in capital letter. FGA score indicates the SplicePort computational score

polypyrimidine tract (ppt) were carried out using SVM-BP finder tool (http://regulatorygenomics.upf.edu/Software/SVM_BP/) (Corvelo et al. 2010).

Results

PCR was conducted on cDNA of buffalo. Only in three samples that the first primer pairs resulted in 526 bp amplicon which corresponded to *CATH4* full-spliced coding region with complete 4 exons and part of 3′ untranslated region (Fig. 1). The second primer pair was used with the cDNA samples that were not amplified by the first primer. It amplified segments of 1067 bp (Fig. 2). In the alignment between the two amplified amplicons, 526 bp (full-spliced segment) and 1067 bp segments showed three separate matches (Fig. 3) covering exons1, 2, and 3. This left segments of 102 bp between exons 1 and 2, 136 bp between exons 2 and 3, and 469 bp after exon 3.

In order to determine the nature of the sequences between the exons and beyond, a PCR reaction was conducted using the second primer pair on genomic DNA buffalo sample. The resulted DNA amplicon was aligned with the cDNA (1067 bp segment) using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Fig. 4).

The alignment showed that the 1067 bp cDNA (accession# *MK507762.1*) is an intron-retained amplicon with full retention of intron 1 (102 bp) and intron 2 (136 bp) and a large segment of 5′ of intron 3 (469 bp). In order to find out reasons for unsplicing of the introns, the cDNA intron-retained sequence was analyzed. The analysis includes identifying the locations of any occurring SNPs and the sequences of cis-acting 4 splice signs, needed for splicing. The latter includes the 5′ (GT) and 3′ (AG) intron splice sites, the branch points (BPs), and polypyrimidine tract (ppt).

Prediction of the potential 5′ and 3′ splice sites (GT-AG) in *CATH4* mRNA intron-retained form was determined using the Splice Port tool (http://spliceport.cbcb.umd.edu/), and feature generation algorithm (FGA) scores were calculated (Table 2). The analysis showed that the donor splice sites in the three introns are the top splice site candidates which indicate that the three introns should have been spliced.

The potential branch points (BPs) and polypyrimidine tract (ppt) in introns 1 and 2 (intron 3 was only 5′ partial) of the intron-retained sequence were investigated using the SVM-BP finder (http://regulatorygenomics.upf.edu/Software/SVM_BP/).

The candid positions were the ones with svm_score of 1.3308 and 0.9529 located at distances of 35 nt and 15 nt from the 3′ splice site of intron 1 and intron 2, respectively, (Table 3).

Figure 5 shows all candidate branch points and polypyrimidine tract in intron 1 and intron 2 nucleotide sequences of the intron-retained cDNA. In intron 1 and intron 2, the uridine tracts had only 6 and 4 non-continuous uridines (t), respectively.

Table 3 The potential predicted branch point (BP) and polypyrimidine tract (ppt) in buffalo *CATH4* unspliced cDNA sequence

seq_id	agez	ss_dist	bp_seq	bp_scr	y_cont	ppt_off	ppt_len	ppt_scr	svm_scr
Intron 1	28	75	aggtgaggg	− 1.5456	0.41	42	20	30	− 2.5503
Intron 1	28	56	agatcagag	− 3.8131	0.49	23	20	30	− 2.1565
Intron 1	28	41	aaatgagcc	− 0.2929	0.67	8	20	30	0.2686
Intron 1	28	35	gcctgaacc	1.3833	0.70	2	20	30	1.3308
Intron 2	13	115	gggtcaatg	− 1.7699	0.50	40	12	27	− 2.5034
Intron 2	13	96	agctgaaca	0.3411	0.51	21	12	27	− 0.4227
Intron 2	13	57	gggtgaggc	− 0.4101	0.52	44	9	22	− 2.2737
Intron 2	13	44	aggttatgg	− 2.4447	0.59	31	9	22	− 2.1868
Intron 2	13	21	gtttgacct	1.4336	0.81	8	9	22	0.9218
Intron 2	13	15	ccttgagcc	0.4239	0.90	2	9	22	0.9529

seq_id - Sequence Identifier, *agez* AG dinucleotide exclusion zone length, *ss_dist* distance to 3′ splice site, *bp_seq* branch point sequence (nonamer; from − 5 to +3 relative to the BP adenine), *bp_scr* BP sequence score using a variable order Markov model, *y_cont* pyrimidine content between the BP adenine and the 3′ splice site, *ppt_off* polypyrimidine tract offset relative to the BP adenine, *ppt_len* polypyrimidine tract length, *ppt_scr* polypyrimidine tract score, *svm_scr* final BP score using the SVM classifier

```
>intron1
gtgagttggggaggggactgYctaggtgAggggcagggagacagatcAgagaagga
aaaatgAgcctgAac***ccagtttcccccgcactttaatccgtcaccag***

>intron2
gtgaggctgggggttgggggtcAatgtttccYagggagctgAacagggagcttctg
ggaWggtttcctgtctctgggggtgAggctgggaggttAtggccaaggggattccag
tttgaccttgAgc***ctctccttcc***cag
```

Fig. 5 SVM-BP finder outputs all candidate branch points (BPs) of Egyptian buffalo *CATH4* intron-retained cDNA. Where BP adenine (A) are in capital letters, BP sequence (nonamer; from −5 to +3 relative to the BP adenine) are underlined, and polypyrimidine tract length (ppt_len) are in italic

SNPs positions relative to splicing sites

Polymorphic sites and their distances from the GT splice site were determined (Table 4) in nucleotide sequences of the spliced cDNA (8 SNPs) and the intron-retained cDNA (13 SNPs; 8 in exonic regions and 5 in intron-retained segments).

Discussion

Intron retention, a form of alternative splicing that affects the mechanism of gene expression control in mammals, is not fully yet studied. It enhances gene regulatory complexity in vertebrates (Schmitz et al. 2017) and plays an essential conserved role in normal physiology and in diverse diseases (Wong et al. 2016). It was in 1997 that Coolidge et al. reported the role of the cis-regulatory elements in intron retention. Galante et al. (2004) and Sakabe and de Souza (2007) reported that retained introns are on the average shorter, more C/G rich, and associated with weaker splice sites than constitutive introns.

In the present work, we have detected retention of the three introns in *CATH4* of Egyptian buffalo. Accurate intron splicing occurs in the presence of strong cis-acting 4 splice signs which include the 5′

Table 4 Single-nucleotide polymorphisms in spliced and Intron-retained sequences of Egyptian buffalo *CATH4* mRNA and their distances from GT splice site

Regions	Positions of single-nucleotide polymorphisms		Distance from GT splice site
	Full-spliced cDNA	Intron-retained cDNA	
Exon 1(c.1–198)	c.25 T/G(K)	C	173
	c.26 T/C(Y)	T	172
	c.44 T/C(Y)	T	154
	c.84 T/C(Y)	C	114
	c.108G/A(R)	G	90
	c.147C/T(Y)	C	51
	c.176A	c. 176 T/A(W)	22
	c.183A/C(M)	C	15
	c.186G/A (R)	c. 186 A/G(R)	12
	c.191C	c. 191 T/C(Y)	7
Retained intron 1	–	c.198+21 C/T(Y)	18
Exon 2 (c.199–306)	c276C/T(Y)	c.276 T/C(Y)	30
	c277A/G(R)	c.277 G/A(R)	29
	c284G/A(R)	c. 284A/G(R)4	22
Retained intron 2	–	c.306+32 T/C(Y)	29
	–	c.306+60 A/T(W)	57
Exon 3 (307-378)	c.315G/ A (R)	c.315 G/A(R)	63
	c.377A	c.377 C/A(M)	1
Retained intron 3	–	c.378+94 G/C(S)	91
	–	c.378+249 T/A(W)	246

(GT) and 3′ (AG) intron splice sites, the branch points (BPs), and polypyrimidine tract (ppt) sequences which are essential for accurate splicing (Black 2003). However, these cis-sequences provide only one half of the information required for recognition by the splicing machinery (Lim and Burge 2001). The number of nucleotides between the branchpoint and the nearest 3′ acceptor site, ranging from 18 to 40 bp, was found to affect splice site selection (Taggart et al. 2012; Clancy 2008). In the present study, the branch sites in the intron-retained cDNA were located at distances of 35 nt and 15 nt from the 3′ splice site for introns 1 and 2, respectively, which are within the distances suggested by Clancy (2008). No information was available for intron 3 since it was only 5′ partially retained.

For intron-splicing, pyrimidine tracts play a role. Strong pyrimidine tracts contain 11 continuous uridines, whereas decreasing the continuous uridine stretch to five or six residues requires that the tract be located immediately adjacent to the AG for optimal competitive efficiency (Coolidge et al. 1997). In the present investigation, intron retention may have been caused by the presence of weak pyrimidine tracts which contained only 6 and 4 non-continuous uridine stretch for intron 1 and intron 2, respectively, and are not immediately next to AG. No results were available for intron 3 since only the 5′ segment of intron 3 was present in the sequence. It is worth mentioning that the fully retained introns in buffalo CATH4 gene were short. Retaining introns in genes with short introns have been reported (Sakabe and de Souza 2007).

Recently, SNPs have been considered to play a role in alternative splicing leading to intron retention (Wang et al. 2014; Ju et al. 2015). Exonic SNPs have direct effects on the properties of proteins, while SNPs within introns and untranslated region can affect the expression and splicing of mRNA (Wang et al. 2013, 2014). In a study by Estivill (2015), it was reported that 20,000 SNPs were located close to splice sites. However, there were cases where SNPs were > 30 nucleotides away from the splice sites that disrupted splicing in 10,000 exons with evidence of alternative splicing and that splicing mutations located deeper in intronic regions (within 300 nucleotides from splice sites) were associated with disease. In the present study, 8 SNPs were detected in buffalo intron-retained cDNA, located at distances ranging from 7 to 246 nt away from the GT splice site which may be related to the retention of the three introns.

The presence of different *CATH4* variants has been reported in other breeds of Indian buffaloes (Brahma et al. 2015). Individual buffalo from Mehsana and Murrah breeds was reported to carry 4-6 variants of *CATH4* gene and that the gene could be present in multiple copies. Differences in *CATH4* copy number has been reported to be breed-specific in indicine (Nelore) relative to the taurine cattle (Bickhart et al. 2012).

Conclusion

CATH4 in river buffalo (Egyptian breed) mRNA is present in spliced variant and intron-retained variants (three introns were retained) despite the presence of the conserved cis-sequences (the 4 splice signals) essential for accurate splicing. Retention of introns 1, 2, and 3 may have occurred as a result of short introns, weak polypyrimidine tracts containing 6 and 4 non-continuous uridine stretch, and/or SNPs located close to AG splice site.

Abbreviations

BPs: Branch points; CATH4: Cathelicidin-4; CD46: Cluster of differentiation46; IR: Intron retention; NCF4: Bovine neutrophil cytosolic factor 4; nt: nucleotide; PCR: Polymerase chain reaction; ppt: Polypyrimidine tract; SNPs: Single-nucleotide polymorphisms

Acknowledgements

Not applicable

Authors' contributions

AAAM and SMEN, designed the experiment, analyzed the data, and wrote the manuscript. EAB and ETS conducted the practical section of the work. NMO carried out the statistical analysis. All authors read and approved the final manuscript.

Funding

No specific fund was supplied for this work.

Availability of data and materials

We declare that all data analyzed during this study are included in this published article.

Ethics approval and consent to participate

Blood samples used in this study was collected and provided by the buffalo farm experienced veterinary.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 13 March 2019 Accepted: 3 July 2019
Published online: 23 July 2019

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215:403–410

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1990) Current protocols in molecular biology (editors). Greene Publishing and Wiley-Interscience, New York

Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK et al (2012) Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res. 22:778–790

Black DL (2003) Mechanisms of alternative pre-mRNA splicing. Annu Rev Biochem. 27:291–336

Brahma B, Patra MC, Karri S, Chopra M, Mishra P et al (2015) Diversity, antimicrobial action and structure-activity relationship of buffalo cathelicidins. PLOS ONE. 10(12):e0144741

Chacko E, Ranganathan S (2009) Genome-wide analysis of alternative splicing in cow: implications in bovine as a model for human diseases. BMC Genomics. 10(Suppl3):S11

Clancy S (2008) RNA splicing: introns, exons and spliceosome. Nat Educ. 1(1):31
Coolidge CJ, Seely RJ, Patton JG (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. Nucleic Acids Res. 25(4):888–896
Corvelo A, Hallegger M, Smith CW, Eyras E (2010) Genome-wide association between branch point properties and alternative splicing. PLoS Comput Biol. 6(11):e1001016
Das H, Sharma B, Kumar A (2006) Cloning and characterization of novel cathelicidin cDNA sequence of *Bubalus bubalis* homologous to *Bos taurus* cathelicidin-4. DNA Seq. 17(6):407–414
Dirksen WP, Sun Q, Rottman FM (1995) Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. J Biol Chem. 270:5346–5352
Dorschner RA, Pestonjamasp VK, Tamakuwala S, Ohtake T, Rudisill J, Nizet V et al (2001) Cutaneous injury induces the release of cathelicidin anti-microbial peptides active against group A Streptococcus. J Investig Dermatol. 117:91–97
Drögemüller C, Reichart U, Seuberlich T, Oevermann A, Baumgartner M, Kühni Boghenbor K et al (2011) An unusual splice defect in the mitofusin 2 gene (MFN2) is associated with degenerative axonopathy in Tyrolean Grey cattle. PLoS One 6:e18931
Estivill X (2015) Genetic variation and alternative splicing. Nat Biotechnol. 33(4): 357–359
FAO (2013) FAO Statistical year book - world food and agriculture
Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ (2004) Detection and evaluation of intron retention events in the human transcriptome. RNA. 10: 757–765
Gallo RL, Murakami M, Ohtake T, Zaiou M (2002) Biology and clinical relevance of naturally occurring antimicrobial peptides. J Allergy Clin Immunol. 110:823–831
Gillenwaters EN, Seabury CM, Elliott JS, Womack JE (2009) Sequence analysis and polymorphism discovery in 4 members of the bovine cathelicidin gene family. J Hered. 100(2):241–245
Ju Z, Wang C, Wang X, Yang C, Sun Y, Jiang Q et al (2015) Role of an SNP in alternative splicing of bovine NCF4 and mastitis susceptibility. PLoS ONE 10(11):e0143705
Lim LP, Burge CB (2001) A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci USA. 98:11193–11198
Marai IFM, Habeeb AAM (2010) Buffaloes' reproductive and productive traits as affected by heat stress. Tropical Subtropical Agroecosystems. 12:193–217
Menoud A, Welle M, Tetens J, Lichtner P, Drögemüller C (2012) A COL7A1 mutation causes dystrophic epidermolysis bullosa in Rotes Höhenvieh cattle. PLoS One. 7:e38823
Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16:1215
Sakabe NJ, de Souza SJ (2007) Sequence features responsible for intron retention in human. BMC Genomics. 8:59
Schmitz U, Pinello N, Jia F, Alasmari S, Ritchie W, Keightley MC, Shini S, Lieschke GJ, Wong JJ, Rasko JEJ (2017) Intron retention enhances gene regulatory complexity in vertebrates. Genome Biol. 18:216
Taggart AJ, DeSimone AM, Shih JS, Filloux ME, Fairbrother WG (2012) Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat Struct Mol Biol. 19(7):719–721
Untergasser A, Cutcutache I, koressaar T, ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer 3 new capabilities and interfaces. Nucleic Acids Res. 40:e115
Wang X, Li T, Zhao HB, Khatib H (2013) A mutation in the 3′ untranslated region diminishes micro RNA binding and alters expression of the OLR1 gene. J Dairy Sci. 96:6525–6528
Wang X, Zhong J, Gao Y, Ju Z, Huang J (2014) A SNP in intron 8 of CD46 causes a novel transcript associated with mastitis in Holsteins. BMC Genomic. 15:630
Wong JJ, Au AY, Ritchie W, Rasko JE (2016) Intron retention in mRNA: no longer nonsense: known and putative roles of intron retention in normal and disease biology. Bioessays. 38(1):41–49
Zaiou M, Gallo RL (2002) Cathelicidins, essential gene-encoded mammalian antibiotics. J Mol Med. 80:549–561
Zanetti S, Deriu A, Volterra L, Falchi MP, Molicotti P, Fadda G, Sechi LA (2000) Virulence factors in Vibrio alginolyticus strains isolated from aquatic environments. Annali di Igiene. 12(6):487–491

Publisher's Note