ACADEMIA JOURNAL OF BIOLOGY 2023, 45(3): 23–33
DOI: 10.15625/2615-9023/18295

ANALYSIS OF GENETIC POLYMORPHISM IN GENES ENCODING CATHELICIDINS FROM VIETNAM INDIGENOUS YELLOW CATTLE BREEDS

Thi Tuoi Do[1], Hong Ngoc Ta[1], Thi Thuy Anh Tran[1], Doan Lan Pham[2], Thi Phuc Do[1], Thi Hong Van Nguyen[1,*]

[1]VNU University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Ha Noi, Vietnam
[2]National Institute of Animal Science, 9 Tan Phong, Ha Noi, Vietnam

Received 24 April 2023; accepted 18 September 2023

ABSTRACT

Cathelicidins include antimicrobial peptides (AMPs) and are involved in the innate immune defense against infections in mammals. Polymorphisms in the DNA sequence of cathelicidin genes could be relevant to inherited variations of host innate immunity, adaptation, and pathogen resistance. This study aims to investigate the sequence polymorphism of cathelicidin genes including *CATHL2* and *CATHL4* from local indigenous yellow cattle breeds of Vietnam. Genomic DNA samples were extracted from 52 individuals collected from different cattle populations in Vietnam including Ha Giang, Thanh Hoa, Nghe An and Phu Yen. The *CATHL2* and *CATHL4* genes were amplified by PCR, following sequencing for identification of the single nucleotide polymorphisms (SNPs) and/or insertion-deletion (indel) DNA sequence variations. The copy number variations (CNVs) of the *CATHL4* gene were determined by cloning, single strand conformation polymorphism (SSCP) and real-time PCR. The comparative analysis results showed that there are 13 SNPs detected in all sequences of *CATHL2* gene including 9 SNPs in intron 2 and intron 3, and 4 missense substitutions in exon 1, 2, and 3. These SNPs do not affect the predicted tertiary structure of Bac5 encoded by this gene. For *CATHL4*, the results revealed 15 SNPs, 3 indels and 1 repeat variation as motif (TG)n(G)n/(AC)n(C)n. Among these variations, the deletion 46delC in exon 1 of *CATHL4* was detected in a sample of the Ha Giang breed causing a truncated polypeptide as predicted. The average copy number of *CATHL4* of Vietnam indigenous cattle breeds using *GADPH* as a reference gene showed in range from 2.53 to 2.89 copies.

Keywords: *CATHL2*, *CATHL4*, cathelicidins, genetic polymorphism, Vietnam indigenous cattle.

Citation: Thi Tuoi Do, Hong Ngoc Ta, Thi Thuy Anh Tran, Doan Lan Pham, Thi Phuc Do, Thi Hong Van Nguyen, 2023. Analysis of genetic polymorphism in genes encoding cathelicidins from Vietnam indigenous yellow cattle breeds. *Academia Journal of Biology*, 45(3): 23–33. https://doi.org/10.15625/ 2615-9023/18295

[*]*Corresponding author email*: nguyenthihongvan@hus.edu.vn

INTRODUCTION

Antimicrobial peptides (AMPs), also known as host defense peptides (HDPs), are molecules belonging to innate immunity. These AMPs play a role in defense and can react quickly to pathogens including Gram-negative and Gram-positive bacteria, fungi, and viruses. They have a multi-functional biological activity which can be chemotactic and immune-regulatory (Lazarev et al., 2010), therefore have the ability to act as endogenous antibiotics and kill the microbes by disrupting cell membranes. Cathelicidins (CATHLs) are categorized in a family of host-defense peptides and identified in many species, including hagfish, amphibians, fish, birds, snakes, and mammals (Zanetti et al., 1993; Zanetti et al., 2005). Cathelicidins are encoded by *CATHL* genes sharing a conserved N-terminal cathelin domain and present in the pro-form of their peptides. These genes display a conserved structure with four introns and three exons (Dorin et al., 2015). Several reports showed the diversity in cathelicidin gene sequences which relate to the antimicrobial activity of the encoded AMPs in the hosts (Lazarev et al., 2010; Zanetti et al., 2005; Gillenwaters et al., 2009). The recent studies provided novel genetic variants, and these enable future case-control studies and functional assays designed to elucidate whether CATHL variation may potentially underlie inherited differences in bovine innate immunity (Gillenwaters et al., 2009). In mammal genomes, the number of cathelicidin genes is various. In human and mice genomes, there is a single cathelicidin gene which codes for an α-helical mature peptide. However multiple cathelicidin genes have been identified in cattle and sheep genomes (Whelehan et al., 2014). In human and bovine, cathelicidins are encoded by very polymorphic genes but shared conserved functions (Mookherjee et al., 2006). The diversity and biological activity of cathelicidins of buffalo, a species known for its disease resistance, also were investigated (Brahm et al., 2015). Bac5 encoded by *CATHL2* is the first cathelicidin which was discovered from bovine neutrophils (Zanetti et al., 1993). A series of new homologs of cathelicidin 4 (CATHL4), which were structurally diverse in their antimicrobial domain, was identified in buffalo. AMPs of newly identified buffalo CATHL4s (buCATHL4s) displayed potent antimicrobial activity against selected Gram positive and Gram-negative bacteria (Brahm et al., 2015). The bovine lineage was one of the first in which cathelicidins were discovered during studies of the antimicrobial activity of bovine neutrophil lysates (Romeo et al., 1988). An assay of milk from cows with mastitis showed that BMAP-27 and BMAP-28 peptides, encoded by *CATHL6* and *CATHL5* respectively, display antimicrobial activity against *Escherichia coli*. To date, at least seven distinct protein-coding cathelicidin genes have been identified in the bovine genome and are found in a single cluster on chromosome 22 (Tomasinsig et al., 2010). There were three additional cathelicidin genes, which are cathelicidin 8, cathelicidin 9 and cathelicidin 10 (Elsik et al., 2009). A comprehensive annotation of the cathelicidin locus in *Bos taurus* using homology-based search methods of the BLAST family of programs and the more sensitive Hidden Markov Models (HMM) approach was conducted. The result of bioinformatic analysis of the bovine genome (BosTau7) revealed seven protein-coding cathelicidin genes, *CATHL1-7*, including two identical copies of the *CATHL4*, as well as three additional putative cathelicidin genes, all clustered on the long arm of chromosome 22 (Whelehan et al., 2014). The study of comparative sequence analysis of *CATHLs 2, 5, 6*, and *7* for 10 domestic bovine breeds belonging to *Bos taurus* and *Bos indicus* showed 60 SNPs, in which 7 SNPs were non-synonymous and 5 of those were indel mutations (Gillenwaters et al., 2009).

Vietnam indigenous bovine breeds descended from the process of crossbreeding of *Bos indicus* and *Bos taurus*, and domestication for a long time. Vietnam yellow cattle are classified by locality, such as

Thanh Hoa, Nghe An, Lang Son, Phu Yen, Ba Ria-Vung Tau breeds or by identifying characteristics as the yellow, H'mong and U dau riu breeds (Le et al., 1999). Vietnam indigenous cattle breeds have good resistance to pathogenic microorganisms and strong adaptation to adverse environments of local climate and extreme ecological conditions. These characteristics may indicate the potential genetic diversity and include the genes relevant to innate immunity such as cathelicidins. However, up to date, there are no publications on cathelicidins regarding nucleotide and amino acid sequences and the genetic polymorphisms of Vietnam indigenous cattle breeds. This study aims to investigate the polymorphism of genes encoding cathelicidins and provide novel genetic variants to elucidate whether CATHL variations may potentially underlie inherited differences in Vietnamese indigenous bovine breeds. In this study, we focused on polymorphism analysis of *CATHL2* and *CATHL4* genes, which are genes encoding respectively Bac5 and indolicidin showing diversity and potent antimicrobial activities.

MATERIALS AND METHODS

Sample collection and DNA extraction

The ear tissue samples of 52 yellow cattle individuals from 4 local breeds of Vietnam (Phu Yen, Ha Giang, Thanh Hoa, and Nghe An) were collected by the scientists of Vietnam National Institute of Animal Science. These samples were collected following ethical permission for research use. All tissue samples were kept in liquid nitrogen and then stored at -80 ºC until DNA extraction. DNA was isolated using the Bioneer kit (Korean) following the manufacturer's protocol. The purity and concentration of DNA were checked by spectrophotometer A260/A280 ratio and gel electrophoresis on 1% agarose gel.

PCR for *CATHL2* and *CATHL4* gene amplification

Specific primers were designed to amplify the studied genes based on cathelicidin gene sequences published in the Genbank database (Accession No NC.037349.1 (*CATHL4*), EU380692.1 (*CATHL2*), AJ786261.1 (*GADPH* as the reference gene for Real-time PCR) and checked by using BLAST, NCBI (Table 1). These primers were distributed by IDT and Thermo Fisher Scientific, USA.

PCR was carried out in 20 µL volume for each negative control without a DNA template. The reactions contained 4 µL of 5×Phusion HF Buffer, 200 µM of each dNTP, 0.5 µM of each primer, and 1U Phusion DNA Polymerase (Thermo Fisher Scientific, USA) and 100 ng genomic DNA as template. The thermocycler was done as follows: 95 ºC for 10 min., 25 cycles including denaturing at 95 ºC for 45 sec., annealing at 63–67 °C for 40 sec., elongation at 72 ºC for 1 min., additional elongation for 10 min. PCR product was visualized by electrophoresis conducting with 1% agarose gel in TAE 1X buffer, 100 V voltage for 30 min., then purified using QIAquick Gel Extraction Kit (Qiagen, USA) or MeGAquick-Spin PlusTotal Fragment kit (Intron, Korea).

CATHL4 gene cloning

Amplicons of the *CATHL4* gene from all samples of 4 indigenous local cattle breeds were ligated into pJET1.2/blunt vector (Thermo Fisher Scientific, USA) for cloning. Recombinant vectors were transformed into competent *E. coli* DH5α and screened for colonies containing recombinant vectors in LB medium added with 100 µg/mL ampicillin. Colony PCR was conducted using pJET 1.2F and pJET 1.2R vector primers (Table 1). Plasmids were extracted with a plasmid purification kit of Qiagen (NSW, Australia).

SSCP analysis for screening *CATHL4* polymorphism

For SSCP analysis, 50 colonies containing *CATHL4* recombinant vector from each cattle sample were selected. The plasmids were isolated and then used for PCR in SSCP analysis. To amplify *CATHL4* exon 4 (approximately 170 bp in length) from these plasmid templates, PCR was conducted using C4SS-F and C4SS-R primers (Table 1). The components of the reaction included 1X HF

Buffer, 0.5 µM for each primer, 1 ng *CATHL4* plasmid DNA template and de-inonized H_2O up to the volume of 10 µl. The reaction was employed following the thermocycler: 98 ºC for 3 min., 30 cycles of denaturing at 98 ºC for 20 sec., annealing at 63 ºC for 30 sec. and elongation at 72 ºC for 10 min, additional elongation for 10 min. The PCR products were denatured with 60% formamide, 1X dye, incubated at 95 ºC for 10 min. and the DNA single strands were stabilized by keeping them at -20 ºC for 10 min. The electrophoresis of DNA single strands was carried out in 12% polyacrylamide (Acrylamide: bisacrylamide (30:1)), 100 V for 4 hours. SSCP visualization was conducted by staining the gel with silver staining protocol (Bassam et al., 1993). The plasmids displaying different patterns of SSCP were identified at the full length of the *CATHL4* nucleotide sequence.

Table 1. Nucleotide sequences of primers used for polymorphism analysis of *CATHL2* and *CATHL*4 genes

Primers	Target amplification	Nucleotide sequence (5'-3')	Amplicon length (kb)	Reference
CAT-D2F	*CATHL2*	GAGGACTGGGGACCATGGAGACC	1.52	Biswajit et al. (2015)
CAT-D2R		TCCCAAGAGGTCTTCCCTGGGCT		
CAT-D4F	*CATHL4*	AGACTGGGGACCATGCAGACCCAG	1.35	Biswajit et al. (2015)
CAT-D4R		TCACTGTCCAGAAGCCCGAATCTG		
C4SS-F	SSCP *CATHL4* Exon 4	GCTTGTGGCCTCCTTTTTCATAGCTCCA	0.17	This study
C4SS-R		TCACTGTCCAGAAGCCCGAATCTG		
CAT-4F	qPCR *CATHL4* Exon 4	CTGTCAGATCCTGAGCCTCGGGAA	0.07	This study
CAT-4R		TCACTGTCCAGAAGCCCGAATCTG		
pJET1.2F	*Cloned CATHL4*	CGACTCACTATAGGGAGAGAGGC	1.49	IDT
pJET1.2R		AAGAACATCGATTTTCCATGGCAG		
GAPDH-F	*GAPDH*	TCTTCACTACCATGGAGAAGGCTGG	0.13	
GAPDH-R		TGCAGCAAGGAGCAAGTTAATTGCA		

DNA sequencing and polymorphism analysis of *CATH2* and *CATHL4* genes

Purified PCR products (PCR products of genomic DNA template for *CATHL2* and colony PCR for *CATHL4*) were sent for sequencing at Macrogen Company (Korea). Both investigated *CATHL* gene sequences were aligned and compared with the corresponding *CATHL* gene sequences of *Bos taurus* codes including EU380692.1 for *CATHL2* and NC037349.1 for *CATHL4*. Bioinformatic software was used for comparative analysis including BioEdit version 7.0.5.3, and Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI, USA, http://www.genecodes.com/). Ambiguous polymorphisms were validated and confirmed by repeated PCR and sequencing. Prediction of amino acid sequences was identified using the online SMART tool (http://smart.embl-heidelberg.de/). Analysis of the effect of amino acid changes on the spatial structure of proteins was performed using Phyre2 software (Kelly et al., 2015).

Determination of the absolute copy number of *CATHL4* by real-time PCR

For identification of the absolute copy number of *CATHL4* in the 14 samples of 4 breeds, *CATHL4* exon 4 was amplified by real-time PCR in a 10 µL reaction mixture containing 5 µL of 2X SYBR Green Master Mix, 0.3 µL of each 10 µM/L primer CAT-4F and CAT-4R, 1 µL of 50 ng DNA template and 3.4 µL of H_2O. The copy number was calculated by comparison to *GAPDH* as a

reference gene. Amplified *CATHL4* exon 4 or *GADPH* gene was cloned into pJET1.2, transformed into *E. coli* DH5α. Recombinant plasmids were isolated, diluted, adjusted concentration using the Nano-Drop ™ 1000 Spectrophotometer system (Thermo Fisher). A tenfold diluted series of plasmids was used to construct the calibration curve (10^7-10^4 copies/µL). Plasmid concentrations were calculated and converted to number of DNA copies using the published equation (Whelan et al., 2003). The samples were analyzed to determine the cycle threshold (Ct) values. Real-time PCR was performed using a 7500 Real Time PCR System (Applied Biosystem, USA) in 3 replicates and standard deviation < 0.5. Normalized copy number was obtained by calculating a ratio of *CATHL4* copy number to *GADPH* copy number.

RESULTS

Polymorphism analysis of the *CATHL2* gene

Genomic DNA were extracted from 52 tissue samples of cattle collected in 4 populations including Ha Giang (11 samples), Thanh Hoa (13 samples), Nghe An (14 samples), and Phu Yen (14 samples), then used for PCR to amplify the *CATHL2*. All samples gave specific amplicons of 1.46 kb in length as estimated for the *CATHL2* gene (Fig. 1). The 1.46 kb sequence covers 236 bp upstream region of exon 1 and extends downstream of exon 4 more than 113 bp.

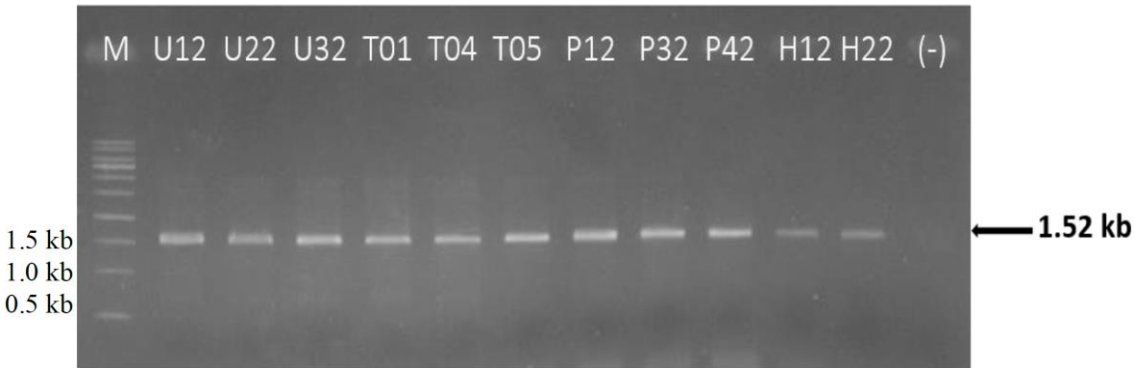

Figure 1. Electrophoresis of PCR products for amplification of *CATHL2*: 1% agarose, 1X TAE buffer, M: 1kb DNA Ladder (PhileKorea). U: Nghe An breed, T: Thanh Hoa breed, P: Phu Yen breed, H: Ha Giang breed, (-): negative control

All amplicons were purified and sequenced successfully. The achieved sequences were compared with RefSeq EU380692.1 (*Bos taurus* breed Holstien) from Genbank using BioEdit and Sequencher 4.7 software. The comparison of all sequences showed a homology of above 99%. From alignment results, there are 13 SNPs detected in the *CATHL2* gene from investigated samples (Table 2), including 9 SNPs found in intron 2 and intron 3, 4 missense substitutions found in the coding region of exon 1, 2, and 3. Among 13 substitutions, there are 9 transitions and 4 transversions. There is no indel polymorphism detected in this gene.

Among 13 SNPs in the *CATHL2* gene, 5 SNPs including 432G>A, 1355T>C, 1382C>G, 1393G>A, and 1397G>A are coincident with SNPs detected in domestic cattle breeds previously reported (Gillenwaters et al., 2009; Romeo et al., 1988). Using Bovine Genome Variation Database (BGVD, http://animal.omics.pro/code/index.php/BosVar/searchBySNP, Ningbo, 2020) to analyse, we identified that 8 SNPs including 411G>A, 615G>C, 672G>C, 677C>G, 771T>C, 842A>G, 924A>G, 964A>G are new variants. Especially, sample T07 from Thanh Hoa breed has the double SNPs at 411G>A and 615G>C, which cause amino acid substitutions at p.59

Glu>Lys and p.89 Ser>Thr. CATHL2 protein 3D structure was predicted via Phyre2 online software. The result showed that these substitutions do not affect the structure and function of CATHL2. The 3D structure prediction showed a fold structure domain of Cystatin-like, belonging to superfamily Cystatin/monellin, family Cathelicidin motif with confidence 100%. Using the SMART online software, all missense substitutions showed the conserve domains include signal peptide domain from amino acids 1–29, PFAM Cystatin domain (amino acids 24–116), PFAM Cathelicidin domain (amino acids 30–130). The confidently predicted domains of the SNP 432G>A (p.66 Asp>Asn) showed Cystatin-like domains (amino acids 18–125) which are similar to the amino acid sequence deduced from reference EU380692.1. However, the confidently predicted domains of SNPs 411 G>A (p.59 Glu>Lys), 842A>G (p.117 Asn>Asp) are PFAM Cystatin and PFAM Cathelicidin. The SMART analysis of the T07 sample with double amino acid substitutions 411G>A and 615G>C showed the confidently predicted domains including signal peptide and PFAM Cathelicidin.

Table 2. SNPs identified in *CATHL2* sequences of Vietnam indigenous yellow cattles in comparative analysis to *Bos taurus* with accession number EU380692.1

No.	Exon or Intron	Genomic position[a], substitution[b]	Amino acid change[c]	Local breeds (number of sample)[d]
1	Exon 1	411 G>A	p.59 Glu>Lys	T (3)
2	Exon 1	432G>A	p.66 Asp>Asn	H (1)
3	Exon 2	615G>C	p.89 Ser>Thr	T (2), U (1)
4	Intron 2	672G>C	None	H (3), P (1)
5	Intron 2	677C>G	None	T (2)
6	Intron 2	771T>C	None	T (3), U (1)
7	Exon 3	842A>G	p.117 Asn>Asp	P (4)
8	Intron 3	924A>G	None	T (2), U (1)
9	Intron 3	964A>G	None	T (3), U (1)
10	Intron 3	1355T>C	None	All breeds
11	Intron 3	1382C>G	None	All breeds
12	Intron 3	1393G>A	None	All breeds
13	Intron 3	1397G>A	None	All breeds

Note: a: Genomic position identified based on CATHL2 GenBank Accession No. EU380692.1 (*Bos taurus*); b: Reference > query alleles are shown; c: Amino acids encoded by Reference and query alleles were compared; d: The breed abbreviations: T: Thanh Hoa, U: Nghe An, P: Phu Yen, H: Ha Giang The number in parentheses indicates the number of samples.

Genetic polymorphism analysis of the *CATHL4* gene

PCR amplification and sequencing

CATHL4 from all samples of 4 local breeds was amplified by PCR with primer pairs CATD4-F and CATD4-R (Table 1). The PCR products were run on 1% agarose gel and showed amplified fragments of about 1.37 kb in length as expected (Fig. 2).

All PCR amplicons of *CATHL4* were sequenced by Sanger sequencing. Analysis of sequences revealed the polymorphisms of CNV of repetitive sequence (C)n(CA)n/(TG)n(G)n, which showed the noisy signals for all sequences.

To identify the *CATHL4* nucleotide sequence and polymorphism as SNPs and CNVs, we cloned the *CATHL4* gene into

pJET1.2/blunt vector. The recombinant vectors were transformed into the competent *E. coli* DH5α and colony PCR using primers pJET1.2F and pJET1.2R was conducted to amplify the *CATHL4* gene and select the expected colonies.

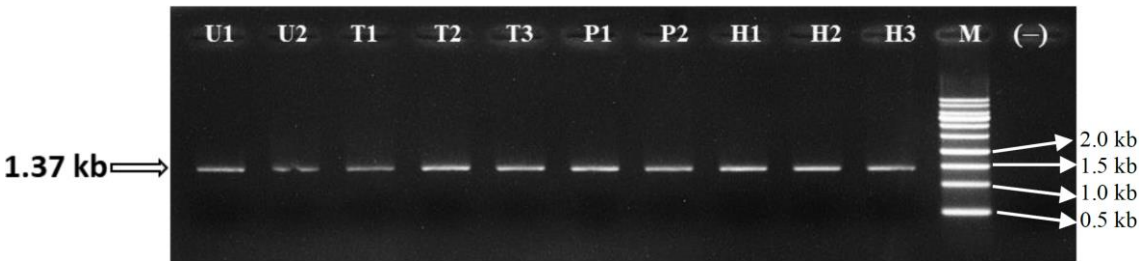

Figure 2. Electrophoresis of PCR products for amplification of *CATHL4*, M: 1kb DNA Ladder (PhileKorea). U1-2: Nghe An breed, T1-3: Thanh Hoa, P1-2: Phu Yen, H1-3: Ha Giang, (-): negative control without DNA template

Table 3. List of SNPs identified in *CATHL4* sequences of Vietnam indigenous yellow cattle breeds in comparative analysis to *Bos taurus* with accession number NC 037349.1

No.	Exon or Intron	Genomic position[a], substitution[b]	Amino acid[c]	Local breeds (clone)[d]
1	Exon 1	46delC	Truncated protein	H (1)
2	Exon 1	123C>T	p.12 Trp>Arg	P (1)
3	Intron 1	296G>A	None	P (1)
4	Intron 1	300T>C	None	P (1), U (3)
5	Intron 1	306G>A	None	H (3), P (1)
6	Intron 1	487T>C	None	H (2), P (3) T (1)
7	Intron 2	522 delA	None	P (1), T (2)
8	Intron 2	534T>C	None	H (1), P (3), T (2)
9	Intron 2	597A>G	None	H (1), P (1), T (1)
10	Exon 3	620A>C	p.123 Asn>Thr	H (1)
11	Exon 3	621C>G	p.123 Asn>Lys	H (2), P (1), T (1)
12	Intron 3	927C>T	None	All samples
13	Intron 3	957C>A	None	All samples
14	Intron 3	1093-1116Ins (24 bp)	None	H (4), P (3), T (2)
15	Intron 3	1075G>T	None	P (2)
16	Intron 3	1148Ins ((TG)n(G)n/(AC)n(C)n)	None	All breeds (13)
17	Intron 3	1196T>C	None	H (1), P (2), T (2)
18	Exon 4	1245C>G	p.137 Pro>Arg	H (1), P (3), T(2)
19	3'UTR	1299C>G	None	H (2), P (4), T (2)

Note: a: Genomic position identified based on *CATHL4* GenBank Accession No. NC 037349.1 (breed Hereford *Bos taurus*); b: Reference > query alleles are shown; c: Amino acids encoded by reference and query alleles were compared; d: The breed abbreviations: T: Thanh Hoa, U: Nghe An, P: Phu Yen, H: Ha Giang. The number in parentheses indicates the number of samples.

For preliminary screening of the polymorphism in the *CATHL4* clones, we used PCR-SSCP analysis of *CATHL4* exon 4 and detected the different SSCP patterns of electrophoresis on the gel. 14 cattle samples showing the different patterns of SSCP were

selected for DNA sequencing of full length *CATHL4*. Besides, the SSCP patterns showed the potential copy number variation. If there is one locus of the gene, there are possibly two allelic variations and if there are more than two variants, it suggests that the gene has more than one locus. In our study, the SSCP showed 4 different electrophoretic patterns of the *CATHL4* gene in one sample of Ha Giang breed. Based on the preliminary results of SSCP analysis and DNA sequencing, *CATHL4* in the yellow cattle breeds of this study is supposed to have more than one copy of the gene. The comparative polymorphism analysis using the results of *CATHL4* DNA sequencing was conducted and aligned with NC 037349.1 and NM_174827.2 as references. The results revealed the sequence of NC 037349.1 contains a delG variation at the first nucleotide of exon 3 compared to all aligned sequences including all *CATHL4* sequences of this study and NM_174827.2. In *CATHL4* sequences of this study, 19 polymorphic sites were identified, including 15 SNPs, 3 indels and 1 variation of complex repeat sequence model (TG)n(G)n/(AC)n(C)n that was detected in intron 3 of all samples in this study (Table 3).

Among 15 SNPs, 10 SNPs are in introns 1, 2 and 3, 1 SNP is in 3'UTR. Four nonsynonymous SNPs in exons 1, 3, and 4 are 123C>T, SNP 620A>C, 621C>G and 1245C>G causing amino acid substitutions p.12Trp>Arg, p.123Asn>Th, p.123Asn>Lys, and p.137Pro>Arg respectively. Phyre2 software was used to analyse and identify the effect of these substitutions on the 3D structure of CATHL4 cathelicidin. The result showed that the CATHL4 amino acid sequences have the typical structure that was identified as a Cystatin-like fold, Cystatin/monellin superfamily and cathelicidin motif family. Among the 3 indels detected, there is only one deletion (46delC) in exon 1 of *CATHL4* of the Ha Giang breed sample. This deletion is a frameshift mutation in the coding sequence and the predicted polypeptide chain is truncated because of the presence of a premature stop codon. Phyre2 analysis showed an extremely changed 3D structure. The other indels are in intron 2 and intron 3, thus do not affect to amino acid sequence.

Analysis of the absolute copy number of CATHL4

The CNV is one of the sources of genetic diversity in the cathelicidin gene family. In this study, the CNV of *CATHL4* was identified via real-time PCR. The 72 bp long fragment of *CATHL4* exon 4 from 16 individuals belonging to 4 populations was amplified by real-time PCR using CAT-4F and CAT-4R primers (Table 1). The *GADPH* gene was used as a reference with the amplified fragment of 130 bp. The absolute copy number of *CATHL4* was determined based on the calculation of the linear equations of the corresponding calibration curve, converted in log10 values, and normalized with the *GADPH* gene (Table 4). The result indicated that the copy number of *CATHL4* is polymorphic within the breed, and the average copy number of 4 local breeds ranges from 2.53 to 2.89 copies.

Table 4. Absolute copy number of *CATHL4* gene in four Vietnam indigenous yellow cattle breeds

Local breed	Number of samples	Log copy number (Mean ± SEM)	Copy number (Mean ± SEM)	Normalized copy number (Mean ± SEM)	Maximum copy number (Mean ± SEM)
Phu Yen	3	5.07 ± 0.03	1.18E+05 ± 6.54E+03	2.86 ± 0.006	5.30
Ha Giang	5	4.76 ± 0.03	7.12E+04 ± 6.24E+03	2.56 ± 0.011	6.69
Thanh Hoa	5	5.11 ± 0.08	1.39E+05 ± 2.46E+04	2.53 ± 0.002	3.61
Nghe An	3	4.97 ± 0.02	9.29E+04 ± 3.18E+03	2.84 ± 0.002	3.60

DISCUSSION

Characterization of SNPs, indel polymorphisms and CNVs within the bovine CATHL gene family has been focused on in different publications (Gillenwaters et al., 2009; Bickhart et al., 2012). The novel polymorphisms are central in these studies to understand the effect of nonsynonymous *CATHL* variation on bovine innate immunity. Although the cathelicidins of human and bovine origin are highly variable in sequence but share conserved functions (Mookherjee et al., 2006). The data on the frequency and distribution of CATHL genetic polymorphisms in indigenous bovines of Vietnam is quite limited.

In our study, among 13 SNPs in *CATHL2* gene sequences, there is no change in the 3D structure of the variant protein compared to the wild type. SNPs 432G>A the amino acid replacements 66Asp>Asn was also the SNP has been reported in the previous study (Gillenwaters et al., 2009). Three nonsynonymous substitutions which are 411G>A in exon 1, 615G>C in exon 2, and 842A>G in exon 3 are newly detected SNPs causing amino acid changes at the positions p.59 Glu>Lys, p.89 Ser>Thr, and p.117 Asn>Asp respectively. These amino acid changes belong to the catheline like domain (CLD), but these substitutions do not affect the predicted tertiary structure of the CATHL2 protein. The results confirmed that the coding sequence of the *CATHL2* is highly conserved within the species, especially in the sequence encoding AMD region (exon 4). However, via SMART analysis, these synonymous substitutions may cause the difference in confidently predicted domains of CATHL2 protein.

In *CATHL4*, most SNP loci and indels identified in this study are in the noncoding sequence of this gene, only four missense substitutions were detected that do not affected the conformation of encoded cathelicidin. Notably, a 24 bp insertion was found in intron 3 of the *CATHL4* gene detected in three samples and a complex repeat sequence combined with a polymorphic multiple G repeat region as (TG)n(G)n/(AC)n(C)n in all samples. This result coincided with a previous study by Gillenwaters et al. (2009). CpGPlot for sequence analysis of the CATHL4 gene in all samples also showed that no putative CpG islands were detected in this gene.

The previous studies on cattle and buffalo have shown that *CATHL4* displays the high polymorphism not only in nucleotide sequence but also in copy number in genome (Gillenwaters et al., 2009; Biswajit et al., 2015). In the previous study (Whelehan et al., 2014), the BosTau7 breed identified the existence of two copies of *CATHL4* in the genome. In comparison with buffalo which has from 5 to 8 copies of *CATHL4* in its genome (Tomasinsig et al., 2009), Vietnam yellow cattle breeds have fewer copies. Among four local breeds in this study, Ha Giang breeds had the most diversity of *CATHL4* copy number variations, copy number (ratio of *CATHL4/GAPDH* copy number) varied from 1.22 to 6.69.

Hiller et al. (2019) identified SNPs mutations in the *CATHL2* gene and determined their potential association with dairy performance traits in Polish Black-and-White Holstein-Friesian (phf) cows. Using PCR-RFLP, their study genotyped *CATHL2* and calculated the frequencies of *CATHL2*/*Dde*I alleles and *CATHL2*/*Hha*I polymorphisms. From that genotyping, their study confirmed that CC (*CATHL2*/*Dde*I) and CG (*CATHL2*/*Hha*I) genotypes produced higher milk yield than the other cattle genotypes. In our study, the initial data on genetic polymorphisms of *CATHL2* and *CATHL4* genes was reported. However, the case-control studies and functional assays designed to elucidate genetic variants of *CATHL* genes have not been performed in this study. Besides that, research on the relationship between the genetic polymorphism of *CATHL* genes and the other genes associated with the tolerance, performances, and other typical traits of Vietnamese Yellow cattle is necessary. This may contribute to understanding the influence

of the *CATHL* genotypes on selected performance traits and inherited variation in cattle innate immunity.

CONCLUSION

Our study has identified the genetic polymorphism of *CATHL2* and *CATHL4* of indigenous yellow cattle in Vietnam including Ha Giang, Thanh Hoa, Nghe An and Phu Yen breeds. The SNPs, indels, and copy number variation detected in this study confirmed the genetic polymorphisms of these cathelicidin genes from investigated breeds. These preliminary data showed new variants of DNA sequences of cathelicidin genes in Vietnam indigenous yellow cattle. The study offers the necessity of study on other members of the cathelicidin gene family and the expression of these genes to give new insights into the role of cathelicidins in the immune system and tolerance of Vietnam indigenous yellow cattle breeds.

Acknowledgements: The authors would like to thank the Faculty of Biology, VNU University of Science, Vietnam National University, Hanoi, and Key Laboratory of Enzyme and Protein Technology for the administrative support and equipment in laboratories. This research was funded by Vietnam National University, Hanoi, grant number KLEPT.18-02.

REFERENCES

Bassam B. J., Caetano-Anolles G., 1993. Silver staining of DNA in polyacrylamide gels. *Appl Biochem Biotechnol*, 42: 181–188.

Bickhart D. M., Hou Y., Schroeder S. G., Alkan C., Cardone M. F., Matukumalli L. K., 2012. Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Research*, 22: 778–790.

Biswajit B., Mahesh C. P., Satyanagalakshmi K., Meenu C., Purusottam M., Bidhan C. D., Sushil K., Sourav M., Kiran T., Krishna M. P., Tirtha K. D., Sachinandan D., 2015. Diversity, antimicrobial action and structure activity relationship of buffalo cathelicidins. *PLoS ONE* 10(12): e014474.

Dorin J. R., McHugh B. J., Cox S. L., Davidson D. J., 2015. Mammalian antimicrobial peptides: defensins and cathelicidins. *Molecular Medicinal Microbiology*, Second Edition, Volume 1: 539–565.

Elsik C. G., Tellam R. L., Worley K. C., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324: 522–528.

Gillenwaters E. N., Seabury C. M., Elliott J. S., Womack J. E., 2009. Sequence analysis and polymorphism discovery in 4 members of the bovine cathelicidin gene family. *Journal of Heredity*. 100(2): 241–245.

Kelly A. L., Mezulis S., Yates C. M., Wass N. M., Sternberg M. J. E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocol*, 10: 845–858.

Kumar P., Henikoff S., Pauline C. Ng., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocol*, 4(7): 1073−1081.

Lazarev V. N., Govorun V. M., 2010. Antimicrobial Peptides and Their Use in Medicine, *Apply Biochemistry and Microbiology*, 46(9): 803–814.

Lawrence A. K., Stefans M., Christopher M. Y., Mark N. W. and Michael J. E. S, 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10: 845–858.

Mookherjee N., Brown, K. L., Bowdish D. M., Doria S., Falsafi R., Hokamp K., Roche F. M., Mu R., Doho G. H., Pistolic J., Powers J. P., Bryan J, Brinkman F. S., Hancock R. E., 2006. Modulation of the TLR-mediated inflammatory response by the endogenous human host defense peptide LL-37. *Journal of Immunology*, 176: 2455–2464.

Ningbo C., Weiwei F., Jianbang Z., Jiafei S., Qiuming C., Zhuqing Z., Hong C., Tad S. S., Chuzhao L., Yu J., 2020. BGVD: An Integrated Database for Bovine Sequencing Variations and Selective Signatures. *Genomics Proteomics Bioinformatics*, 18: 186–193.

Tomasinsig L., De Conti G., Skerlavaj B., Piccinini R., Mazzilli M., D'Este F., Tossi A., Zanetti M., 2010. Broad-spectrum activity against bacterial mastitis pathogens and activation of mammary epithelial cells support a protective role of neutrophil cathelicidins in bovine mastitis. *Infection and Immunity*, 78: 1781–1788.

Tomasinsig L., Zanetti M., 2005. The cathelicidins- structure, function and evolution. *Current Protein and Peptide Science*, 6: 23–34.

Whelehan J. A., Russel N. B., Whelan M. A., 2003. A method for the absolute quantification of cDNA using real-time PCR. *Journal of Immunology Methods*, 278(1-2): 261–269.

Whelehan C. J., Barry-Reidy A., Meade K. G., Eckersall P. D., Aspinas C., Fernando N., Andrew T. L., Cliona O'F., 2014. Characterisation and expression profile of the bovine cathelicidin gene repertoire in mammary tissue. *BMC Genomics*, 15: 128.

Zanetti M., 2005. The role of cathelicidins in the innate host defenses of mammals. *Current Issues of Molecular Biology*, 7: 179–196.

Zanetti M., Del S. G., Storici P., Schneider C., Romeo D., 1993. The cDNA of the neutrophil antibiotic Bac5 predicts a prosequence homologous to a cysteine proteinase inhibitor that is common to other neutrophil antibiotics. *The Journal of Biological Chemistry*, 268: 522–526.