

In Silico Characterization of Functional Divergence of Two Cathelicidin Variants in Indian Sheep

Kamaljeet K. Dhaliwal¹, Jaspreet S. Arora¹, Chandra S. Mukhopadhyay¹ and Prem P. Dubey²

¹School of Animal Biotechnology, Post Graduate Institute of Veterinary Education and Research, ²Department of Animal Genetics and Breeding, College of Veterinary Sciences, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India.

ABSTRACT: The present work focuses on the *in silico* characterization of functional divergence of two ovine cathelicidin coding sequence (cds) variants (ie, Cath1 and Cath2) of Indian sheep. Overlapping partial cds of both the cathelicidin variants were cloned in pJet1.2/blunt vector and sequenced. Evolutionary analysis of the Cath2 and Cath1 indicated that the mammalian cathelicidins clustered separately from avian fowlicidins. The avian fowlicidins, which are very different from mammalian cathelicidins (Caths), clearly displayed signatures of purifying selection. The pairwise sequence alignments of translated amino acid sequences of these two sheep cathelicidins showed gaps in the antimicrobial domain of Cath1 variant; however, the amino terminal cathelin regions of both the Caths were conserved. Amino acid sequence analysis of full-length cathelicidins available at public database revealed that Cath1, Cath2, and Cath7 of different ruminant species (including our Cath1 and Cath2 variants) formed individual clads, suggesting that these types have evolved to target specific types of microbes. *In silico* analysis of Cath1 and Cath2 peptide sequences indicated that the C-terminal antimicrobial peptide domain of Cath2 is more immunogenic than that of the ovine Cath1 due to its higher positive antigenic index, making Cath1 a promising antigen for production of monoclonal antibodies.

KEYWORDS: antimicrobial peptides, ovine Cathelicidins, phylogeny, positive selection, sequence analysis

CITATION: Dhaliwal et al. *In Silico* Characterization of Functional Divergence of Two Cathelicidin Variants in Indian Sheep. *Evolutionary Bioinformatics* 2015;11 189–196 doi: 10.4137/EBO.S29779.

TYPE: Original Research

RECEIVED: May 21, 2015. **RESUBMITTED:** July 05, 2015. **ACCEPTED FOR PUBLICATION:** July 12, 2015.

ACADEMIC EDITOR: Jike Cui, Associate Editor

PEER REVIEW: Five peer reviewers contributed to the peer review report. Reviewers' reports totaled 1,510 words, excluding any confidential comments to the academic editor.

FUNDING: The funds for this project were provided by the RKVY Scheme, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: drarora2003@gmail.com

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

Introduction

The cathelicidin peptide family of antimicrobial peptides (AMPs) is structurally characterized by the N-terminal cathelin domain,¹ which is very similar to the cathepsin-L inhibitor cathelin (henceforth called cathelicidin),² and a variable C-terminal antimicrobial domain. Most cathelicidins are differentially expressed in a variety of tissues and exhibit broad spectral antimicrobial activity against several fungi, protozoa, Gram-negative and Gram-positive bacteria, and enveloped viruses.³ Cathelicidins have been identified in mammalian^{4–6} and nonmammalian animal species.^{7–10} A wide array of cathelicidin variants with diverse range of antimicrobial functions¹¹ have been discovered in domesticated species. In contrast, humans, mice, and rats have only one known multifunctional cathelicidin gene.¹²

In general, the cathelicidin genes are approximately of 2 kb size, and have four exons interleaved by three introns.¹³ The first three exons encode the signal peptide and cathelin domains, and the last exon codes for the antimicrobial domain with elastase processing site.¹² Huttner and colleagues² reported that eight ovine cathelicidin genes, located on the ovine chromosome 19, encode for seven different cathelicidins. Out of these, two are α -helical peptides: SMAP29 (sheep

myeloid antimicrobial peptide, 29 amino acids in length) and SMAP34 (also called OaMAP28 and Oa-MAP34); four are arginine/proline-rich extended structures (OaBac5, OaBac6, OaBac7.5, and OaBac11); and remaining two encodes identical dodecapeptides (OaDodeA and OaDodeB), which are known as cathelicidin-1. SMAP29 exhibits antimicrobial activity against a number of pathogens. Brogden and coworkers¹⁴ showed that mouse, rabbit, and sheep cathelicidins were effective against nine ovine pathogens. Ovine cathelicidins have been studied in exotic sheep; however, no report is available on Indian indigenous sheep. Besides, the evolution of the cathelicidin variants in Indian sheep has not been investigated. The present study has been designed to analyze sequences of Indian ovine cathelicidins (Cath1 and Cath2) to unravel the evolutionary perspective with the rest of the cathelicidins.

Materials and Methods

Collection of samples. The experiment was conducted on a local sheep breed of Punjab, India, which is being maintained by farmers for wool and mutton. Peripheral blood (5 mL) was aseptically collected in an anticoagulant from the jugular vein of healthy adult sheep using a sterile 24-gauge



needle in a 15 mL conical tube containing 0.5 M EDTA. Ethical approval was obtained from the Institutional Animal Ethics Committee at Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana.

Total RNA extraction and cDNA synthesis. Total RNA was extracted from whole blood by erythrolysis using chilled $1 \times$ RBC lysis buffer (ammonium chloride 0.83 g, ammonium bicarbonate 0.1 g, EDTA 0.5 M 20 μ L, and DEPC-treated water (0.1%) up to 100 mL) followed by centrifugation at 3,000 rpm for 10 minutes at 4 °C to precipitate a white pellet. The isolated white pellet was resuspended in 1 mL of Trizol reagent (Invitrogen) to isolate the total RNA. The quantity and the purity of the extracted total RNA was checked using NanoDrop (Thermo Scientific). The RNA samples with concentration $>1,000$ ng/ μ L and ratio of absorbance (260/280) between 1.95 and 2.15 were used for cDNA synthesis.

cDNA synthesis and cloning. The first-strand cDNA was synthesized using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific), as per the manufacturer's protocol. The primers (Table 1) were designed using the ovine (exotic breed) coding sequences (cds) of *Ovis aries* 5 kDa bactinecin precursor (BAC5) (NCBI Acc. No. NM_001009787.1) and *O. aries* procyclic dodecapeptide (CATHL1B) (NCBI Acc. No. NM_001009772.1) using ABI Primer Express software and custom synthesized from Integrated DNA Technologies (IDT, USA). BLASTn sequence analysis of the PCR amplicons revealed homology with Cath2 and Cath1 of other ruminant species, and therefore these primer pairs were named Cath2 and Cath1, respectively.

The purified PCR products were ligated to pJet1.2/blunt-cloning vector and transformed into *E. coli* TOP10 strain, as per standard protocol.¹⁵ Restriction endonuclease digestion using *Bgl*II confirmed the insert in the recombinant vector. The isolated plasmids were sequenced (University of Delhi, New Delhi, India).

Sequence analysis. The pipeline of the present experiment is depicted in Supplementary Figure 1 (Pipeline of the experimental methods.png).

Processing and homology search of coding sequence data. The cloned sequences of Cath1 and Cath2 were trimmed and edited using BioEdit, Version 7.2.3.¹⁶ The full-length coding sequences (cds) were submitted to the DNA Data Bank of Japan. BLASTn¹⁷ search of the obtained cds yielded 83 homologous and heterologous full-length coding sequences (E-value $<10^{-5}$) of cathelicidin variants of divergent animal species. Those sequences were downloaded in the FASTA format

from the NCBI database (<http://blast.ncbi.nlm.nih.gov/>). The *in silico* translated (using the ExPasy Translation tool) complete amino acid sequences of 83 selected cathelicidins were retrieved in FASTA format.

Pairwise and multiple sequence alignment. The DNASTar (Lasergene, DNASTar, Inc.) software and the online tools MAFFT (<http://www.ebi.ac.uk/Tools/msa/mafft/>) and Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) were used for pairwise sequence alignment of the ovine cathelicidins (Cath2 and Cath1) as well as multiple sequence alignment of all the 85 complete cathelicidin peptide sequences (Alignment data available in Supplementary File S1. 85 AA Aln.FAS.MDSX).

Construction of phylogenetic tree. The best evolutionary model was selected using MEGA6¹⁸ software, based on the lowest Bayesian information criterion (BIC) scores. The Akaike information criterion (AICc)-corrected values were determined for each of the models. The best model for analyzing the amino acid data was the Jones–Taylor–Thornton (JTT) matrix-based model.¹⁹ MEGA6 software was used for construction of phylogenetic tree and estimation of evolutionary divergence, and Fisher's exact test and codon-based test were used for determining the selection pressure on the cathelicidin variants. The evolutionary tree was constructed using the maximum likelihood method, considering the JTT substitution model and five discrete Gamma categories for rates of substitution among sites. The reliability of the branching of the tree was checked by 1,000 bootstrap resampling (phylogenetic tree file: Supplementary File S2. AA ML Non-Compressed Tree1, opens in MEGA6).

Evolutionary divergence. The evolutionary divergence between the peptides was estimated using the JTT substitution model.¹⁹ The heatmap and the sample dendrogram for evolutionary divergence of the ruminant cathelicidin variants were generated using the WGCNA package of R (version 3.0.2) (codes available in Supplementary File S4. R Codes Ev. Analysis Data). The heatmap was constructed using the evolutionary divergence matrix (data available in Supplementary File S3. Ev. Div. Rum Matrix).

Disparity index and Fisher's exact test. The amino acid sequences were tested for homogeneity of the substitution pattern (disparity index) using 100 Monte Carlo replications and complete deletions of gaps.²⁰ The probability (P) of rejecting the null hypothesis of strict neutrality in favor of the alternative hypothesis of positive selection (Fisher's exact test) for each sequence pair²¹ was calculated. The numbers of synonymous and nonsynonymous differences between sequences were estimated

Table 1. Primer pairs used for PCR amplification of Cath1 and Cath2 coding sequence of Indian sheep.

OLIGO	PRIMER SEQUENCES (5'–3')	GC CONTENT (%)	Tm (°C)	AMPLICON SIZE
Cath2	F-ATGGAGACCCAGGGAGCCAG	65	58	602
	R-ATGCTGACTCAGGCCAAATGAGAT	46	56	
Cath1	F-ATGGAGACCCAGAGGCCAGCCT	65	62	547
	R-CAGATCCAGTAGCTTGAGGCCGAGA	56	61	



using the Nei–Gojobori method (assumed transition/transversion bias = 2).²²

Determining the sites of positive or negative selection.

Nineteen complete cds representing the Cath variants of ruminant species (ie, taurine (NCBI Acc. nos BC133480, BC120477, L02650), bubaline (XM006065184, AJ812216, EF050453, XM006065185), caprine (XM005696006, NM001285577), ovine (AB973433, NM001009787), yak (XM005910806)), along with some non-ruminant Cath sequences as outgroup [equine (AJ224927, NM001081930), canine (AY392089), feline (NM001204778), donkey (FJ803910), primates (BC055089, AF181954)] were selected. The positive, negative, or neutral selection pressure being experienced by each of the codons was determined by different statistical methods, namely, single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), internal branch FEL (IFEL), and finally, random effects likelihood (REL)²³ using the Datamonkey (www.datamonkey.org/) online server.²⁴ At the beginning, a model selection was run to test all the 203 time-reversible models. A hierarchical testing (<http://www.datamonkey.org/help/models.php>) amalgamated with nested likelihood ratio test (LRT) with AIC-corrected selection was done by the server to obtain a single “best-fitting” rate matrix. Significance level was set separately for each of the analyses (SLAC, FEL, IFEL, and REL). Branch-site REL test (BSR)²⁵ for episodic diversifying selection (EDS) was done to endure any variation in evolutionary rate along branches vis-à-vis sites (<http://www.datamonkey.org/help/index.php>). EDS identifies the branches that has experienced a transient but not easily detectable selection during evolution.²⁶ The impact of positive selection on the subsets of sites was determined by this analysis, without any a priori proof about the lineages of interest.

Comparison of predicted protein structures. The amino acid sequences of both cathelicidins (Cath2 and Cath1) were deduced from the cds using ExPasy Translation tool (<http://web.expasy.org/translate/>) and then were subjected to secondary structure prediction using Pspred v3.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>)²⁷ and tertiary structure prediction by an ab initio approach using the online available tool RaptorX (<http://raptorx.uchicago.edu/>).²⁸ The tertiary protein structure obtained from RaptorX protein was subjected to structure validation using the WhatIf online tool (<http://swift.cmbi.ru.nl/whatif/>), and Ramachandran’s plot analysis was done for validation of the predicted protein structure using the RAMPAGE online tool (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>).²⁹ The DNASTar (Lasergene) software was used for determining the biochemical features of both ovine Cathelicidins (Cath1 and Cath2). The antigenicity of the peptides (Cath1 and Cath2) was predicted by the Pro-tein tool of DNASTar, Lasergene, software suit.

Results and Discussion

Cloning and sequencing of cathelicidin. The positive clones were confirmed for the presence of the insert by

restriction enzyme digestion, and two isolated plasmids of Cath2 and Cath1 were sequenced. The sequences of both cathelicidin cds were submitted to DNA Data Bank of Japan [DDBJ getentry Acc. No. of Cath2: AB973433.1 (<http://getentry.ddbj.nig.ac.jp/getentry/na/AB973433.1/>) and Cath1: AB974690.1 (<http://getentry.ddbj.nig.ac.jp/getentry/na/AB974690.1/>)].

Pairwise sequence alignment of Cath1 and Cath2 peptides. The pairwise alignments of Cath2 and Cath1 (Fig. 1) showed considerable variations in the AMP region; however, the signal peptide and cathelin domain regions were relatively conserved. These two variants of cathelicidin peptides may have thus evolved to confer protection against different microbes.

Multiple sequence alignment of ruminant cathelicidin peptides. The multiple sequence alignment of the cathelicidin types in ruminant species revealed that the lengths of the signal peptide and the cathelicidin domains are nearly same; however, considerable variation was found in the mature peptide (harboring the AMP domain) region. It can also be inferred that the variations in the AMP region signify different specificity of the cathelicidin variants for wider array of microorganisms.

Phylogenetic analysis. The tree depicts (Fig. 2) that the mammalian cathelicidin variants cluster distinctly from that of the avian cluster. The marine animals cluster along with other mammals, as the marine animals included in this study are also mammals (ie, different types of whales and dolphins). Cath1, Cath2, and Cath7 of different ruminant species form individual clads (irrespective of species specificity), suggesting that the cathelicidin types have evolved to target specific cathelicidin types of microbes. The tree also demonstrates that the different cathelicidin variants Cath4, Cath5, and Cath6 belonging to same species have clustered in different clads due to the difference in sequences among variants within same species vis-à-vis homology among the same Cath variant belonging to different ruminant species. The primates have formed a separate clad from the non-primate, hoofed species, amongst the mammals. This could indicate that cathelicidin AMPs have undergone natural selection in accordance with the habitat/ecological niche and species-specific needs to defend against microbes. All the porcine Caths are clustered together in a single clad, which suggests that the cathelicidin genes of porcine origin have evolved through independent rounds of duplications and subsequent diversification after the separation of Suidae and Bovidae families. The cathelicidin types or variants belonging to the same species presented high bootstrapping values (>70%), indicating higher consistency of the given data and taxonomical bipartitioning.³⁰

Estimation of evolutionary divergence. The departure of the ovine Cath1 and Cath2 cds was compared with the cds of cathelicidin variants belonging to different species, with respect to certain parameters. The evolutionary divergence estimates³¹ among ruminant species such as *Ovis aries* (sheep), *Bos taurus* (Cattle), *Capra hircus* (Goat), *Bubalus bubalis* (buffalo), *Bos mutus*

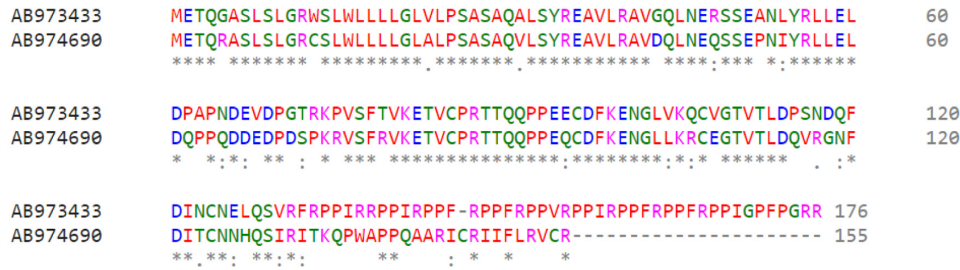


Figure 1. Pairwise sequence alignment of the ovine cathelicidin peptide sequences (Cath2: A973433 and Cath1: AB974690) *in silico* translated from the coding sequences obtained from the present study.

(Yak), and *Pantholops hodgsonii* (Tibetan wild antelope) ranged from 0.000 to 0.365. The maximum evolutionary divergence was observed between the Caths of *Phasianus colchicus* (common pheasant) and that of the ruminant species (dark red cells in heatmap, Fig. 3), while the least divergent homologues were the cathelicidins belonging to same families but different species (dark blue). The sample dendrogram along with heatmap of the evolutionary divergence values (Fig. 4) among

the cathelicidin peptides of the ruminant species enables clear visualization of the evolutionary divergence. Here, the rainbow color along the left and top of the sample dendrogram helps to associate the names of the gene/transcript variants of the ruminant species with the corresponding node of the sample dendrogram, row-wise versus column-wise. The ovine Cath1 and Cath2 (named in the figure as NewCath1 Oae and NewCath2 Oae, respectively) have clustered in two different major

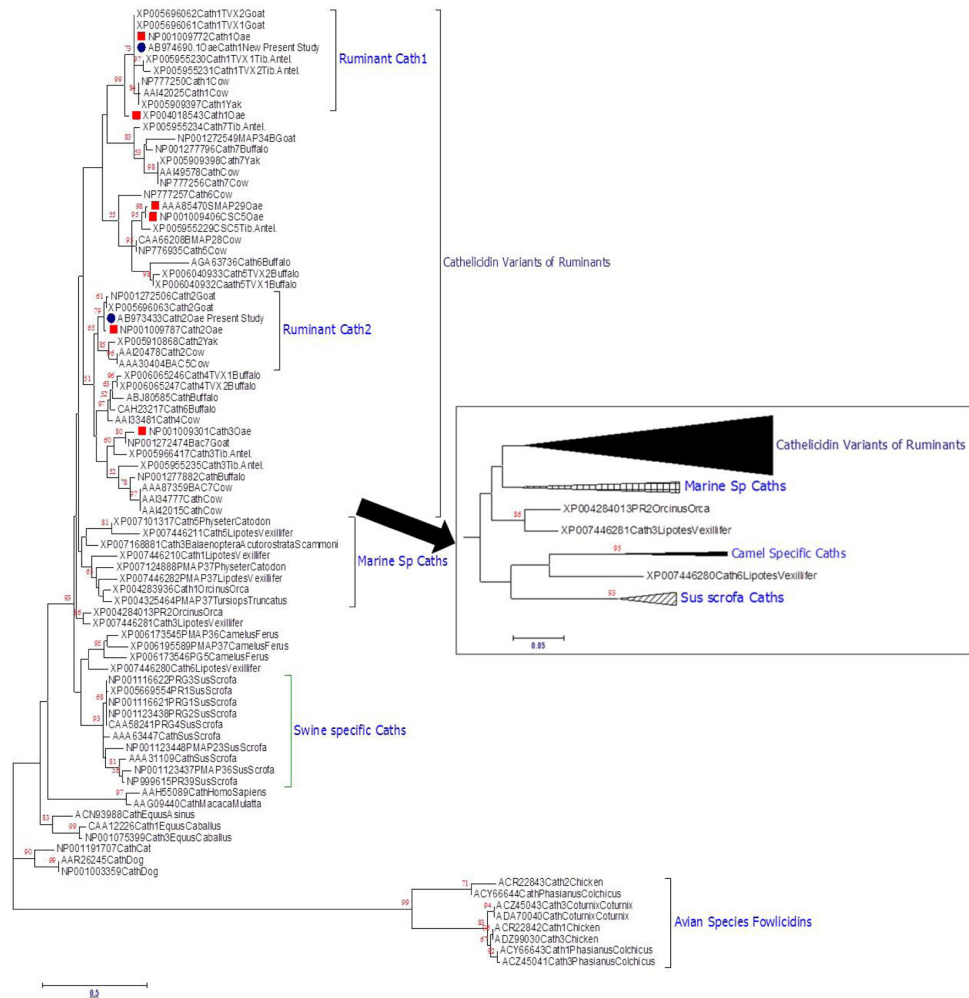


Figure 2. Phylogenetic tree constructed from the cathelicidin peptide sequences, using maximum likelihood with 1,000 bootstrap resampling. The subtree within the box depicts the evolutionary relatedness of Caths among land and marine mammals.

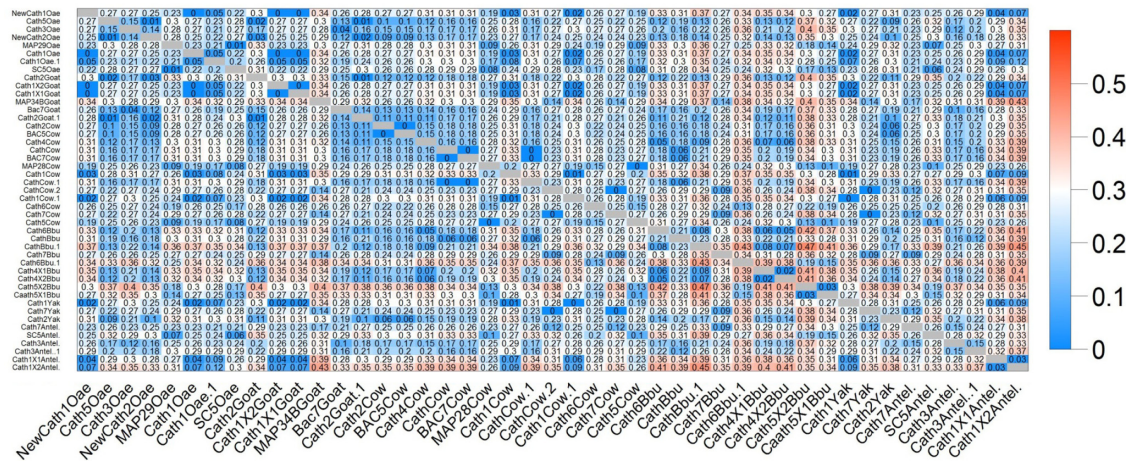


Figure 3. Evolutionary divergence heatmap. Relative distance among the transcript variants of the ruminant species.

clads of the dendrogram (Fig. 4). The color variation in this heatmap (Fig. 4) is only qualitative and does not represent the actual values of evolutionary divergence.

Homogeneity of substitution patterns between sequences.

The discrepancy in evolutionary patterns for the divergent cathelicidin cds was estimated by disparity index,

with the assumption of homogeneity (same pattern of base substitution) of the evolutionary process.²⁰ The null hypothesis of same prototype of substitution pattern among the cathelicidin cds was rejected for many of the sequence pairs, either closely related ruminant species (ie, sheep and goat) or distantly related (ie, some rodents and aves) from the sheep

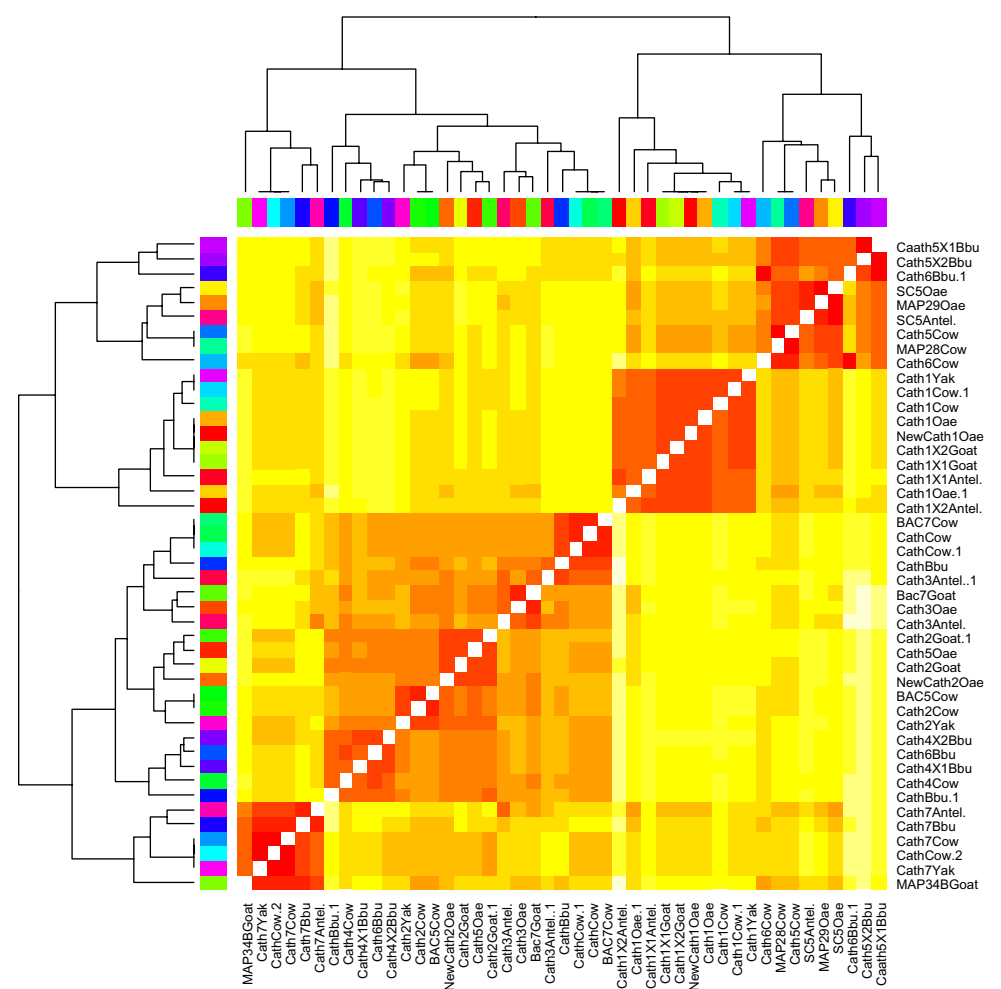


Figure 4. Sample dendrogram and evolutionary divergence heatmap of the ruminant specific cathelicidin peptide sequences.

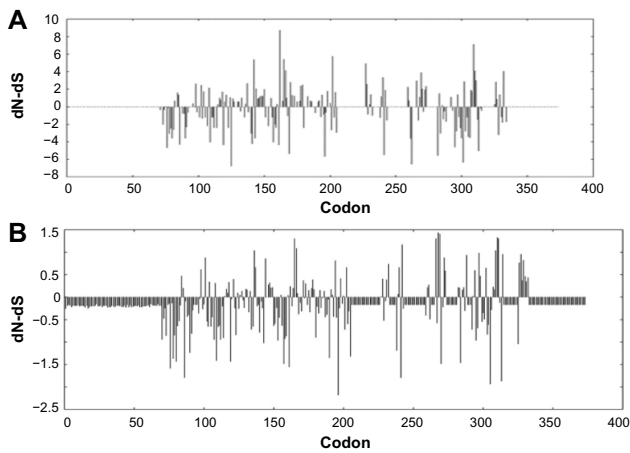


Figure 5. Plot for the value of test statistic (dN-dS) against the codon positions obtained from the analyses, ie, single likelihood ancestor counting [SLAC: (A)] and random effects likelihood [REL: (B)].

cathelicidins (based on the sequence divergence estimation). Different cathelicidin types in cattle and non-ruminant species (ie, swine and equines species) have shown significant disparity during evolution, indicating that the substitution pattern varied among species with respect to that of sheep. This implies that the branching pattern obtained in the phylogenetic tree could be misleading, as the assumption of homogeneity of substitution pattern has not been met. To address this question, the coding sequence data of 19 cathelicidin sequences (belonging to different cathelicidin types of ruminants and one representative cathelicidin sequence from some non-ruminant species) were subjected to

“branch-site REL analysis” (BSR)²⁵ in Datamonkey online server (<http://www.datamonkey.org/>).²³

Estimation of selection pressure site by site. The test statistic “dN-dS” determines the type of selection pressure (due to positive, negative, or neutral selection) over individual codon. SLAC, which is the fastest and most conservative method (out of the six methods available in Datamonkey server) to test selection at a single codon by comparing site-by-site selection,²⁴ detected three positively selected codon sites (162, 165, and 202). In Figure 5, we have comparatively plotted the test statistic (dN-dS) against the codon positions obtained from the analyses using SLAC and REL. The REL method, which is the most powerful and the least conservative out of SLAC, FEL, and REL, detected no positively selected sites and 20 negatively selected codons as compared to SLAC. REL is a stringent test that declares types of either positive or negative selection of codons. The fixed effects likelihood (FEL) and internal branch FEL (IFEL) methods identified 8 and 10 positively selected sites, respectively. It is evident from the Figure 5 that positive and negative selection pressures have acted over different domains of cathelicidin peptides. Singh et al.³² have recently reported evolutionary analyses of similar kind for bubaline Dicer1 cds, where the REL method was found to identify more number of sites experiencing selection positive or negative pressure. The branch-site REL (BSR) analysis analyzes the episodic diversifying selection of divergent homologous coding sequences. In this analysis, in total five branches (or nodes) could be detected ($P < 0.05$) to undergo episodic diversifying selection. This test estimates the ratio (ω) between nonsynonymous (β) to synonymous (α)

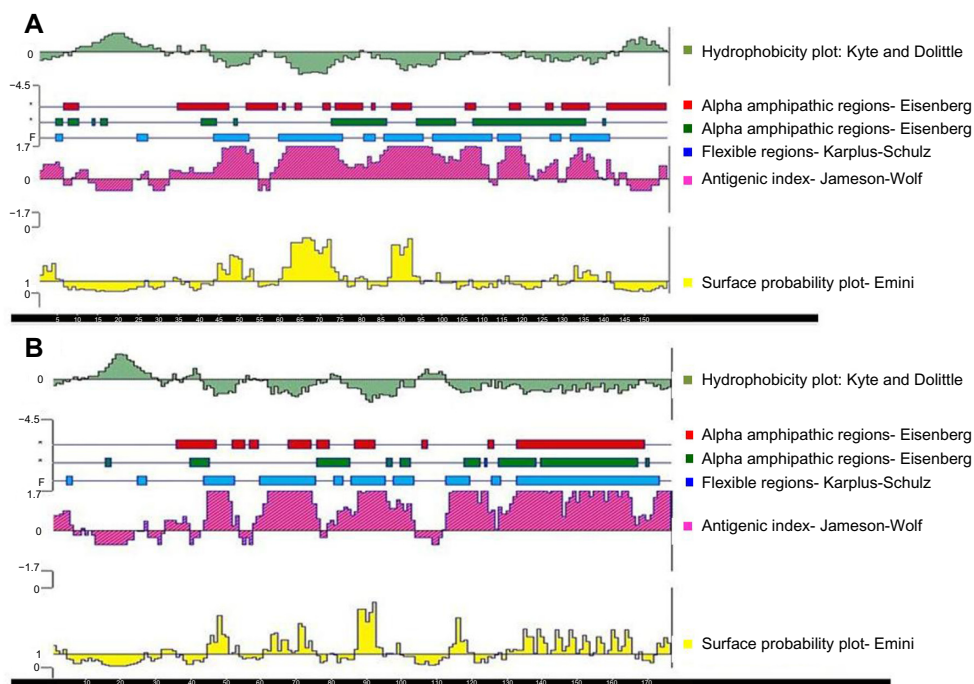


Figure 6. Graphical representation of the salient features (hydrophobicity, antigenic index, surface probability, etc) of the ovine Cath1 (A) and Cath2 (B) peptides, using Protean tool of DNASTar (Lasergene) software.



substitution rates along a lineage of interest, as a measure of selective pressure.

The results indicate that almost all the positively selected sites are positioned in the highly variable AMP region of the cathelicidin. This also indicates that the various cathelicidin variants/types that persist in the species under study have some specific role in conferring immunity and thereby fitness of the organism. Similarly, in sheep, a very limited number of cathelicidin types have been experimentally reported.

Comparative protein structure analysis. The secondary structure of ovine Cath2 showed four β -pleated sheets, two α helices, and several loop structures some of which formed interconnecting links, while the Cath1 protein showed five β -pleated sheets, three α helices, and several loop structures (Supplementary Fig. 2). Cath1 and Cath2 differed in terms of hydrophobicity and antigenicity. Cath1 has more positive hydrophobicity at the N-terminal as compared to Cath2, while the antigenic index of Cath2 has positive antigenic index at the N-terminal region as compared to Cath1 (Fig. 6). This indicates that Cath2 could be more immunogenic than Cath1. The 3D protein structure showed the α -helix, β -strand, and loop structures (Supplementary Fig. 3). The 3D structures of the two variants of ovine Caths differed at the AMP domain. Ramachandran's plot of Cath1 and 2 predicted structures revealed that 96.1% and 93.1% of the residues, respectively, were in favored region (Supplementary Fig. 4). It can be concluded from the present structure analyses that the cathelicidin variants differ in sequence (at the N-terminus), in secondary structure, and perhaps in function. Evolution of the Cath variants has been specific to the type of microbial defense.

Conclusion

Evolutionary analysis of different types and transcript variants of cathelicidin from divergent species showed that the variants of cathelicidin among different animal species have evolved due to positive selection pressure. Avian cathelicidins/fowlicidins have evolved independently (purifying selection) and are very divergent from that of mammalian cathelicidins. Selection pressure varies with codons versus the domains of cathelicidin peptides. Structural differences between the Cath1 and 2 peptides of Indian sheep were observed. Further studies on protein–ligand docking are warranted to understand the evolutionary aspects of the cathelicidin variants against specific microbe-associated molecular patterns (MAMP).

Acknowledgments

The facilities provided by the School of Animal Biotechnology, GADVASU, Ludhiana, are sincerely acknowledged. The authors thank Dr Nikhlesh Singh, Assistant Professor, Department of Physiology, University of Tennessee, Knoxville, Tennessee, USA, Prof. G. S. Brah, Adjunct Professor and Ex-Director, and Dr. R. K. Chaudhary, Assistant Scientist, School of Animal Biotechnology, GADVASU, for critically reviewing the manuscript for English editing.

Author Contributions

Conceived and designed the experiments: JSA. Analyzed the data: CSM. Did bench work and wrote the first draft of the manuscript: KKD. Contributed to the writing of the manuscript: PPD. Agree with manuscript results and conclusions: KKD, JSA, CSM, PPD. Jointly developed the structure and arguments for the paper: KKD, JSA, CSM, PPD. Made critical revisions and approved final version: JSA, CSM. All authors reviewed and approved of the final manuscript.

Supplementary Materials

Supplementary Figure 1. Pipeline of the experimental methods demonstrating the tools and software used in the present study.

Supplementary Figure 2. Secondary structure prediction of ovine Cath2 (left) and Cath1 (right) using Psipred v3.0.

Supplementary Figure 3. Predicted tertiary structure of ovine Cath2 and Cath1 peptides, using RaptorX online server.

Supplementary Figure 4. Ramachandran's Plot of predicted tertiary structures of cathelicidins (Cath1 and Cath2) of Indian sheep, using RAMPAGE.

Supplementary File 1. Alignment data for pairwise sequence alignment of the ovine cathelicidins (Cath2 and Cath1) as well as multiple sequence alignment of all the 85 complete cathelicidin peptide sequences.

Supplementary File 2. Phylogenetic tree file, AA ML Non-Compressed Tree1, opens in MEGA6.

Supplementary File 3. Evolutionary divergence matrix data.

Supplementary File 4. R Codes for evolutionary analysis data.

REFERENCES

1. Bals R, Wilson JM. Cathelicidins – a family of multifunctional antimicrobial peptides. *Cell Mol Life Sci.* 2003;60:711–20.
2. Huttner KM, Lambeth MR, Burkin HR, Burkin DJ, Broad TE. Localisation and genomic organisation of sheep antimicrobial peptide genes. *Gene.* 1998;206:85–91.
3. Chromek M, Slamova Z, Bergman P, et al. The antimicrobial peptide Cathelicidin protects the urinary tract against invasive bacterial infection. *Nat Med.* 2006;12:636–41.
4. Termen S, Tollin M, Olsson B, Svenberg T, Agerberth B, Gudmundsson GH. Phylogeny, processing and expression of the rat Cathelicidin rCRAMP: a model for innate antimicrobial peptides. *Cell Mol Life Sci.* 2003;60:536–49.
5. Treffers C, Chen L, Anderson RC, Yu PL. Isolation and characterisation of antimicrobial peptides from deer neutrophils. *Int J Antimicrob Agents.* 2005;26:165–9.
6. Sang Y, Teresa Ortega M, Rune K, et al. Canine Cathelicidin (K9CATH): gene cloning, expression, and biochemical activity of a novel pro-myeloid antimicrobial peptide. *Dev Comp Immunol.* 2007;31:1278–96.
7. Uzzell T, Stolzenberg ED, Shinnar AE, Zasloff M. Hagfish intestinal antimicrobial peptides are ancient Cathelicidins. *Peptides.* 2003;24:1655–67.
8. Lynn DJ, Higgs R, Gaines S, et al. Bioinformatic discovery and initial characterization of nine novel antimicrobial peptide genes in the chicken. *Immunogenetics.* 2004;56:170–7.
9. Chang CI, Pleguezuelos O, Zhang YA, Zou J, Secombes CJ. Identification of a novel Cathelicidin gene in the rainbow trout, *Oncorhynchus mykiss*. *Infect Immun.* 2005;73:5053–64.
10. Chang CI, Zhang YA, Zou J, Nie P, Secombes CJ. Two Cathelicidin genes are present in both rainbow trout (*Oncorhynchus mykiss*) and atlantic salmon (*Salmo salar*). *Antimicrob Agents Chemother.* 2006;50:185–95.
11. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol.* 2005;3:238–50.



12. Tomasinsig L, Zanetti M. The Cathelicidins-structure, function and evolution. *Curr Protein Pept Sci.* 2005;6:23–34.
13. Scocchi M, Bontempo D, Boscolo S, Tomasinsig L, Giulotto E, Zanetti M. Novel Cathelicidins in horse leukocytes. *FEBS Lett.* 1999;457:459–64.
14. Brogden KA, Kalfa VC, Ackermann MR, Palmquist DE, McCray PB Jr, Tack BF. The ovine Cathelicidin SMAP29 kills ovine respiratory pathogens *in vitro* and in an ovine model of pulmonary infection. *Antimicrob Agents Chemother.* 2001;45:331–4.
15. Green MR, Sambrook J. *Molecular Cloning: A Laboratory Manual.* 4th ed. New York: Cold Spring Harbor Laboratory Press; 2012.
16. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser.* 1999;41:95–8.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725–9.
19. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 1992;8:275–82.
20. 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.
21. Zhang J, Kumar S, Nei M. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Mol Biol Evol.* 1997;14:1335–8.
22. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986;3:418–26.
23. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 2005;22:1208–22.
24. Kosakovsky Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics.* 2005;21:2531–3.
25. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SD, Delpont W, Scheffler K. A random effects branch site model for detecting episodic diversifying selection. *Mol Biol Evol.* 2011;28:3033–43.
26. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012;8(7):e1002764. doi: 10.1371/journal.pgen.1002764.
27. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999;292:195–202.
28. Källberg M, Wang H, Wang S, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc.* 2012;7:1511–22.
29. Lovell SC, Davis IW, Arendall WB III, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins.* 2002;50:437–50.
30. Hedges SB. The number of replications needed for accurate estimation of bootstrap P value in phylogenetic studies. *Mol Biol Evol.* 1992;9:366–9.
31. Tajima F. Simple methods for testing molecular clock hypothesis. *Genetics.* 1993;135:599–607.
32. Singh J, Mukhopadhyay CS, Arora JS, Kaur S. Biocomputational characterization and evolutionary analysis of bubaline dicer1 enzyme. *Asian Australas J Anim Sci.* 2015;28:876–87.