Molecular Evolutionary Analysis of β-Defensin Peptides in Vertebrates

Jianbo Tu[1,*], Diyan Li[1,*], Qingqing Li[2,3,*], Long Zhang[1,*], Qing Zhu[1], Uma Gaur[1], Xiaolan Fan[1], Huailiang Xu[1], Yongfang Yao[1], Xiaoling Zhao[1] and Mingyao Yang[1]

[1]Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu, P.R. China. [2]School of Life Sciences, Key Laboratory of Biomass Energy and Environmental Biotechnology, Yunnan Province, Yunnan Normal University, Kunming, P.R. China. [3]Kunming Xianghao Technology Co, Ltd, Kunming, Yunnan, P.R. China. *These authors contributed equally to this work.

ABSTRACT: Vertebrate β-defensins comprise an important family of antimicrobial peptides that protect organisms from a diverse spectrum of bacteria, viruses, fungi, and protozoan parasites. Previous studies have shown a marked variation in the number of β-defensins among species, but the underlying reason is unclear. To address this question, we performed comprehensive computational searches to study the intact β-defensin genes from 29 vertebrates. Phylogenetic analysis of the β-defensin genes in vertebrates identified frequent changes in the number of β-defensin genes and multiple species-specific gene gains and losses that have been occurring throughout the evolution of vertebrates. The number of intact β-defensin genes varied from 1 in the western clawed frog to 20 in cattle, with numerous expansions and contractions of the gene family throughout vertebrates, especially among tetrapods. The β-defensin gene number in a species is relevant to the ever-changing microbial challenges from the environment that they inhabit. Selection pressure analysis shows there exist three amino acid sites under significant positive selection. Protein structural characteristics analysis suggests that structural diversity determines the diverse functions of β-defensins. Our study provides a new perspective on the relationships among vertebrate β-defensin gene repertoires and different survival circumstances, which helps explain how β-defensins have evolved.

KEYWORDS: β-defensins, evolution, vertebrates, gene duplication, positive selection, structural characteristic

CITATION: Tu et al. Molecular Evolutionary Analysis of β-Defensin Peptides in Vertebrates. *Evolutionary Bioinformatics* 2015:11 105–114 doi: 10.4137/EBO.S25580.

RECEIVED: March 12, 2015. **RESUBMITTED:** April 14, 2015. **ACCEPTED FOR PUBLICATION:** April 20, 2015.

ACADEMIC EDITOR: Jike Cui, Associate Editor

TYPE: Original Research

FUNDING: This work was supported by Natural Science Foundation of China (31471998 and 31402063), The Innovative Research Team from Sichuan Bureau of Education (14TD0002), Key Project of Yunnan Provincial Department of Education (2011Z108), and the Open Cooperation Found of Key Laboratory of Yunnan for Biomass Energy. 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: diyanli@sicau.edu.cn, yangmingyao@sicau.edu.cn

Introduction

Gene duplication has been an important mechanism for shaping immune defenses against the high diversity of pathogens faced by plants, invertebrates, and vertebrates.[1,2] This can be seen from the gene clusters of Toll-like receptors,[3] major histocompatibility complex class I and II,[4] and the antimicrobial peptides (AMPs), ie, defensins[5] and cathelicidins.[6] Cysteine-rich AMPs are abundant in animal and plant tissues involved in host defense. In insects, most AMPs are synthesized in the fat body, an organ analogous to the liver of vertebrates. Cysteine-rich AMPs are summarized in Supplementary Table 1. The number of disulfide linkages varies from one to four. Beta-defensins contain three pairs of disulfide-linked cysteines, and plant AMPs often contain eight cysteines, which form four disulfide linkages. This could correspond to a hypothesis that the main function of the disulfides may be to protect the backbone from proteolysis rather than maintaining the microbicidal activity of the AMP molecules.[7]

Defensins are a group of small cationic peptides, which are a first line of host defense against pathogenic infections.[8] They have a broad spectrum of antimicrobial activities against bacteria, virus, fungi, and protozoan parasites.[9–12] On the basis of the cysteine pairing to form intramolecular disulfide bonds, vertebrate defensins can be classified into three subfamilies, α, β, and θ.[13,14] All three types of defensins have six conserved cysteines. Specifically, the six cysteines of α-defensins are disulfide-linked C1–C6, C2–C4, and C3–C5, but in β-defensins they are connected C1–C5, C2–C4, and C3–C6.[15] Theta-defensin, a cyclic peptide also containing three pairs of disulfide bonds, is believed to arise from peptide splicing of two-hemi θ-defensins.[11] Alpha-defensins are specific to mammals and are mainly produced by leukocytes of myeloid origin and Paneth cells of the small intestine.[11,16] Beta-defensins have been found in most vertebrate species including fish,[8] amphibians,[17,18] lizards,[19] birds,[20] and mammals[21,22] with a much wider range of tissue expression pattern. Theta-defensins were first isolated from the leukocytes of rhesus macaques[23] and are the only backbone-cyclic peptides known in animals up to date,[24] which are believed to have arisen from α-defensins.[25] Comparisons of α- and β-defensins in vertebrates have shown more evidences favoring a closer relationship between vertebrate β-defensins and insect defensins.[21,26]

In recent years, β-defensins have been discovered in various vertebrate species, including teleost fish,[8] Chinese brown

frog,[18] salamander,[17] chicken,[20] zebra finch,[20] duck,[27] lizard,[19] cattle,[28] mouse, and human.[29] The β-defensins are defined by a six-cysteine motif [usual spacing, C-X_6-C-X_4-C-X_9-C-X_6-C-C[30,31]] and abundant basic amino acid residues, and their coding sequences consist of two to four exons. The majority of β-defensins are two-exon genes with the signal sequence being encoded by the first and part of the second exon while the mature peptide is encoded by the second exon.[32] In humans and mice, more than 30 β-defensins were identified in five different regions that are scattered over four chromosomes.[29] In birds, some 13 β-defensin homologs have been discovered recently by in silico analysis and are clustered in a 86-kb region in chromosome 3.[13] Reptiles are also known to possess β-defensin–like peptides, with two toxin peptides being identified in rattlesnake, which contain the same six cysteines as β-defensins.[33] In amphibians, only two β-defensins have been identified.[17,18] These studies strongly suggest that the β-defensins have expanded throughout vertebrate evolution and may have arisen from a common ancestral gene through gene or genome duplication events.[8]

To date, no β-defensin family members have been described in dolphin, panda, manatee, and platypus. How the β-defensin family evolved (birth-and-death processes) during the vertebrate evolution is also unknown. Taking advantage of several vertebrate genomes including fish, amphibians, reptiles, birds, and mammals, we obtained multiple intact β-defensin–like peptides in 29 vertebrates and provided a comprehensive view of birth-and-death processes involving β-defensin genes during the evolution of vertebrates in this study.

Materials and Methods

Identification of novel β-defensin genes. To identify potential sequences in vertebrates with whole genome sequences available at the Ensembl (http://www.ensembl.org/index.html), University of California–San Cruz (UCSC) genome browser (http://genome.ucsc.edu/) and NCBI (http://www.ncbi.nlm.nih.gov/), all 86 intact β-defensin genes previously reported (including 24 human, 10 mouse, 27 cattle, 15 chicken, 1 salamander, 1 Chinese brown frog, 3 lizard and 5 fish sequences) were retrieved from GenBank and used as query sequences to conduct a TBLASTN search of each genome sequence and cDNA sequence using the *E*-value 1e-10. All potential hits were then examined for the presence of the characteristic β-defensin motif or highly conserved signal/prosegment sequence. For every β-defensin sequence identified, additional iterative BLAST searches were performed as described above until no more β-defensin sequences can be revealed.

The topology of the species tree was downloaded from the UCSC Genome Browser (http://hgdownload-test.cse.ucsc.edu/goldenPath/mm10/multiz60way/mm10.60way.commonNames.nh, last accessed May 4, 2014).

Prediction of full-length coding sequences of β-defensins. Most mammalian β-defensins are encoded in two separate exons set apart by a short intron of less than 2 kb, with one exon encoding signal/prosegment sequence and the other exon encoding the mature peptide containing the six-cysteine β-defensin motif.[34] Unlike most mammalian β-defensin genes, which primarily consist of two exons and one intron, the chicken β-defensin genes were found to be composed of four short exons separated by three introns with variable lengths ranging from 117 bp to 3,322 bp.[14] We used a combination of GenomeScan[35] or GENSCAN[36] to identify the full-length coding sequence of β-defensin genes.

Molecular phylogenetic analyses of identified β-defensins in vertebrates. The deduced β-defensin sequences were aligned by ClustalX[37] with appropriate manual adjustments. The neighbor-joining tree of β-defensin amino acid sequences from 29 vertebrates was constructed. The p-distance method[38] was used. The reliability of the estimated trees was evaluated by the bootstrap method[39] with 1,000 replications. The zebrafish preprohepcidin1 (GenBank: AY363452.1) and preprohepcidin2 genes (GenBank: AY363453.1) were used as out-groups because preprohepcidin genes are relatively closer to β-defensin genes among zebrafish AMPs.[40] Phylogenetic trees were constructed in MEGA 5.[41] Multiple sequence alignment was visualized with GeneDoc software.[42]

Selection pressure analyses. Maximum likelihood methods were used to study the selective pressure acting on β-defensin genes, and all tests were conducted using the CODEML program in PAML 4.7 package.[43] Natural selection was examined using the site-specific models of heterogeneous selection pressure among sites. Potential positive selection was tested based on the ratio (ω) of nonsynonymous[28] to synonymous (d_S) substitutions rates ($\omega = d_N/d_S$).[7] Generally, if $\omega = 1$, the amino acid substitutions were assumed to be largely neutral; $\omega > 1$ is evidence of positive selection; $\omega < 1$ was consistent with purifying selection. Five models were used to test for positively selected sites: M1, M2, M7, M8, and M8a. Parameters for the models M0 (one ratio), M1a (neutral), M2a (selection), M7 (beta), M8 [beta and ω (equivalent to Ka/Ks)], and M8a (beta and $\omega = 1$) were calculated. The M0 model assumes a uniform selective pressure among sites. The M1a model assumes a variable selective pressure but no positive selection. The M2a model assumes a variable selective pressure with positive selection. The M7 model assumes a beta-distributed variable selective pressure. The M8 model assumes a beta-distributed variable selective pressure plus positive selection. The M8a assumes a beta-distributed variable selective pressure without positive selection. Four likelihood ratio tests were carried out: M0 vs. M3, M1a vs. M2a, M7 vs. M8, and M8a vs. M8.[43]

Peptide sequence logo was based on the alignment of all intact β-defensin peptide sequences identified in this study. The alignments were slightly modified to adjust the gap positions by visual inspection. Peptide sequence logo of β-defensins was generated from the WEBLOGO (http://weblogo.berkeley.edu/logo.cgi).

Structural analyses. To infer structural characteristics of vertebrate β-defensin peptides, we predicted the folding pattern of several sequences using the Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)[44] and SWISS-MODEL (http://swissmodel.expasy.org)[45] web servers. Electrostatic potential was calculated using the PBEQ-Solver web server (http://www.charmm-gui.org/?doc=input/pbeq-solver).[46] All figures were generated with PyMOL software (DeLano Scientific; http://pymol.org).

Results

Number of intact β-defensin genes in 29 vertebrate species. We used the published data from the NCBI, Ensembl, and UCSC database and applied the known full-length β-defensins from the human, mouse, cattle, chicken, amphibians (salamander and Chinese brown frog), lizard and fish as queries to identify β-defensin genes from the genome sequences of 29 vertebrates. There are a total of 490 β-defensin genes identified in this study (Fig. 1). The nucleotide sequences of all intact β-defensins are provided in Supplementary File 1. The results suggested a remarkable diversity existing in the number of β-defensins in vertebrates. The number of putative β-defensin genes varies from 1 in frog to 42 in cattle. Cattle β-defensin genes present a big expansion compared with other mammals, while marine mammals have fewer β-defensin genes (dolphin and manatee have 8 and 17 β-defensins, respectively). To sum up, β-defensins show a distinct tendency for gene expansion along the evolutionary path from reptiles to mammals.[34]

Phylogenetic relationships of vertebrate β-defensin genes. To clarify and analyze the evolutionary relationships and evolutionary dynamics of the vertebrate β-defensin gene family, we used MEGA 5[41] to construct a neighbor-joining tree (Fig. 2 and Supplementary Fig. 1) by calculating the proportion difference (p-distance) of aligned amino acid sites of the 490 β-defensin sequences in vertebrates.

Because of the relatively smaller number of aligned gap-free sites, the bootstrap values in the tree are generally low. Several lineages show a cluster of genes from the same species or group of closely related species (marked with one color, Fig. 2), whereas other lineages show genes from distantly related species (marked with many colors, Fig. 2).

As shown in Figure 2, most laurasiatherians β-defensins form some species-specific lineages, such as cattle and dolphin. A majority of platypus β-defensin genes form one platypus-specific lineage. Most lizard β-defensin genes form

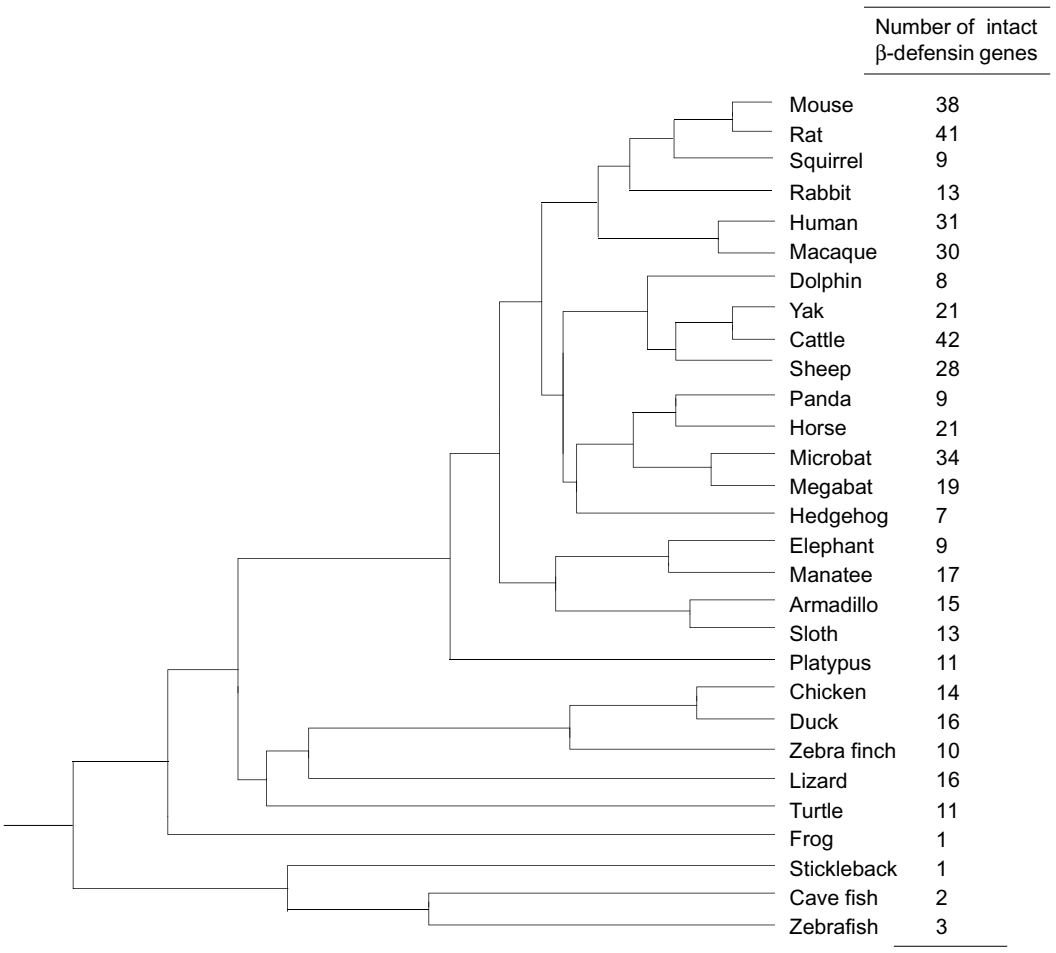

Figure 1. The β-defensin genes of 29 vertebrates determined in this study. Intact β-defensin genes identified in this study.

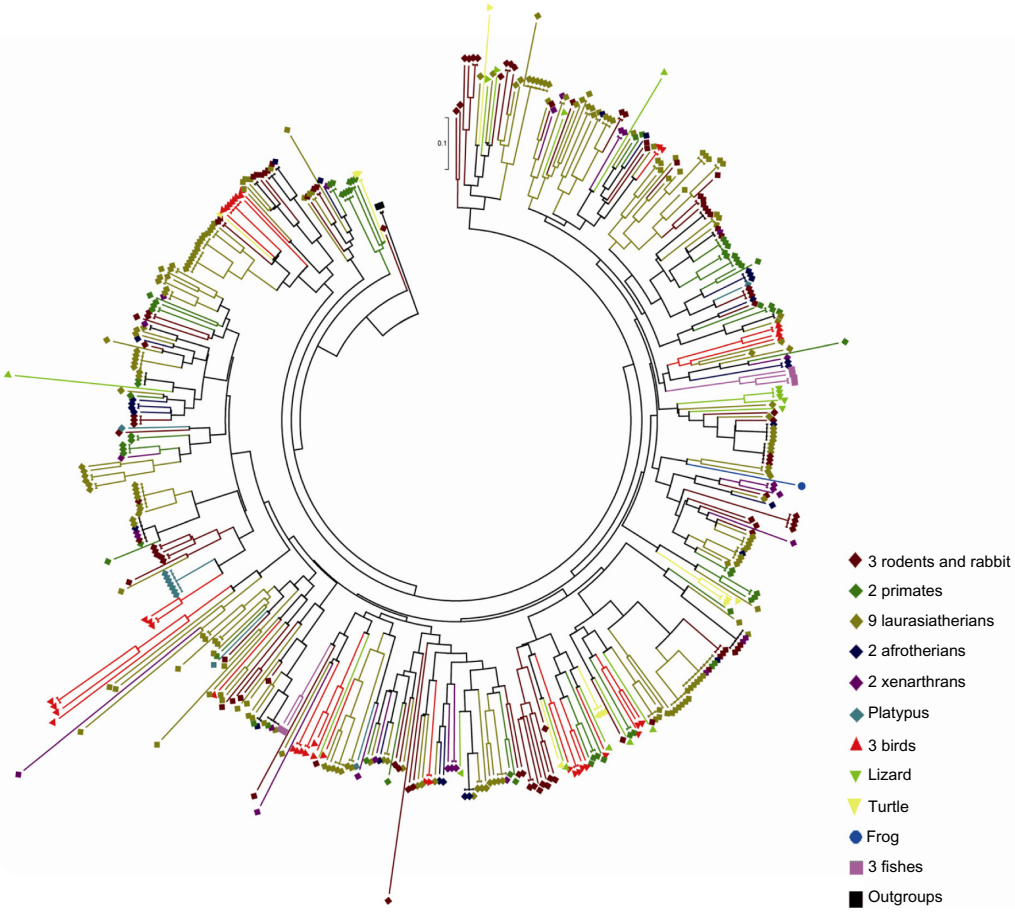

Figure 2. Phylogenetic tree for intact β-defensin genes of 29 vertebrates. The tree was constructed by calculating the proportion difference (p-distance) of aligned amino acid sites of the β-defensin sequences according to the neighbor-joining method and was rooted with zebrafish preprohepcidin1 (GenBank:AY363452.1) and preprohepcidin2 (GenBank:AY363453.1) genes. The reliability of each branch was tested using 1,000 bootstrap replications. Branch lengths are drawn to scale, which is measured by the number of amino acid substitutions per site. Refer to Supplementary Figure 1 for the detailed tree with species and gene names.

three main separated lizard-specific lineages. The primates and birds β-defensin genes cluster together to form several lineages. These results suggested that multiple species-specific gene duplications occurred in these vertebrates during evolution, indicating that most β-defensin genes duplicated after the divergence of these species. In addition to the marked gene duplication between lizard and mammals, we also found several cases of gene loss during the evolution. There is a marked contraction in the lineage leading to amphibian (frog); it is unclear whether the frog lost β-defensins after its divergence from fish. The marked varied number of β-defensins in vertebrates suggested that varying β-defensin gene numbers give animals a mechanism to adapt to markedly changeable conditions.

Analysis of selective pressures in vertebrate β-defensin genes. Earlier results[21] showed that mean d_N was significantly greater than mean d_S in the mature peptide, but there was no significant difference between mean d_S and mean d_N in the signal peptide or prosegment. Therefore, to identify which amino acid positions may be under positive selection in our research, we estimated selection for d_N/d_S for each codon by using the PAML 4.7 package.[43] In the site models M0 vs. M3, the model M3 (discrete) was significantly better than the M0 (one-ratio) model, with $2\Delta\ln L = 4{,}657.482288$ and showed statistical significance ($P < 0.001$). Therefore, the model M3 indicated extreme selection variation among sites and that β-defensins were under weak purifying selection. We employed site models M1a vs. M2a, M7 vs. M8, and M8a vs. M8 to test for positive selection of specific codons. The models M2a and M8, which are also significantly favored over the other models, indicated positive selection. Positively selected sites with posterior probabilities greater than 0.95 are listed in Table 1. The results from the site model (M1a vs. M2a, M7 vs. M8, and M8a vs. M8) indicated that β-defensin genes have three positively selected sites (Table 1) based on the BEB analysis.

To examine the distribution of positively selected sites, the deduced positively selected sites in Table 1 were mapped to the sequence logo of β-defensins (Fig. 3). As shown in Figure 3, positively selected sites are located primarily in the mature peptide. Two of three positively selected sites are located in the mature peptide region. These two positively selected sites are all in the N-terminal of the mature peptide region.

Table 1. Log-likelihood values and positively selected sites for β-defensin genes under site models.

MODEL	P	lnL	*dN/dS*	ESTIMATES OF PARAMETERS	POSITIVELY SELECTED SITES	2Δlnl	DF	*P* VALUE
M0(one-ratio)	979	−107188.361644	0.4873	$\omega_0 = 0.48733$		M0 vs.M3 (4657.482288)	6	$P < 0.001$
M3(discrete)	985	−104859.620500	0.6417	$p_0 = 0.02249$, $p_1 = 0.97751$, $p_2 = 0.00000$, $\omega_0 = 0.06784$, $\omega_1 = 0.65495$,$\omega_2 = 126.3523$				
M1a(neutral)	980	−105025.500488	0.9780	$p_0 = 0.02432$ ($p_1 = 0.97568$), $\omega_0 = 0.09364$, $\omega_1 = 1.00000$		M1a vs. M2a (588.80411)	2	$P < 0.001$
M2a(selection)	982	−104731.098433	1.6119	$p_0 = 0.01980$, $p_1 = 0.46015$ ($p_2 = 0.52005$),$\omega_0 = 0.08698$, $\omega_1 = 1.00000$, $\omega_2 = 2.21137$	16 Q 0.999** 24 D 1.000** 31 N 0.997**			
M7 (beta)	980	−104335.949980	0.4471	$p = 0.44808$, $q = 0.55317$		M7 vs. M8 (677.119058)	2	$P < 0.001$
M8(beta & ω)	982	−103997.390451	1.0770	$p_0 = 0.17689$ ($p_1 = 0.82311$), $p = 0.59523$, $q = 1.37117$, $\omega = 1.24501$	16 Q 0.966* 24 D 1.000**			
M8a(beta & ω=1)	981	−104038.878243	0.8691	$p_0 = 0.16854$ ($p_1 = 0.83146$), $p = 0.57759$, $q = 1.89914$, $\omega = 1.00000$		M8a vs. M8 (82.975584)	1	$P < 0.001$

Note: The site models are implemented using the control variable NS sites in CODEML and *p* is the number of free parameters in the ω distribution. "**" indicates positively selected sites with significance at the 99% level. "*" indicates positively selected sites with significance at the 95% level. The likelihood ratio tests are analyzed by comparing the following pairs of models: M0–M3, M1a–M2a, M7–M8, and M8a–M8.

Implicitly, these sites that have been subjected to positive selection are important in functional diversity of β-defensins. Therein, one positively selected site is located within a region forming an α-helix. Since α-helical regions are often embedded within the membranes, they may be involved in anchoring the β-defensins to the bacterial cell wall.[47] Thus, this positively selected site within the α-helix may play important roles in the immune specificity of β-defensins. We furthermore found an exceptional site subjected to positive selection in the first β-strand. Beta-strands form the structural core of the β-defensins. It is generally suggested that the triple β-strands are characteristic of β-defensins, so these sites are mainly unaffected by positive selection. Thus, the positively selected site identified in this region may represent alterations in the oligomerization of β-defensins.[48]

It is believed that sites within the prosegment region have often suffered negative selection. Exceptionally, we also found a positively selected site located in the prosegment region. The positively selected site in this region was also subject to positive selection, suggesting an uncharted important function of prosegment peptide to date.

Sequence characteristics of cattle intact β-defensins. In this study, we have revealed that 42 intact β-defensin genes exist in cattle genome. In view of cattle β-defensin genes having the most diverse repertoire in vertebrate β-defensins identified so far, we further dissected characteristics of β-defensins in cattle. We constructed phylogenetic tree for the 42 intact β-defensin genes using the NJ method (Fig. 4). This phylogenetic tree showed four distinct phylogenetic gene clusters (Cluster I–IV) existing in cattle β-defensins corresponding to the fact that all cattle β-defensin genes were contained in four chromosome clusters.[49] We speculated that Cluster IV may represent the most ancient cluster from the phylogenetic tree.

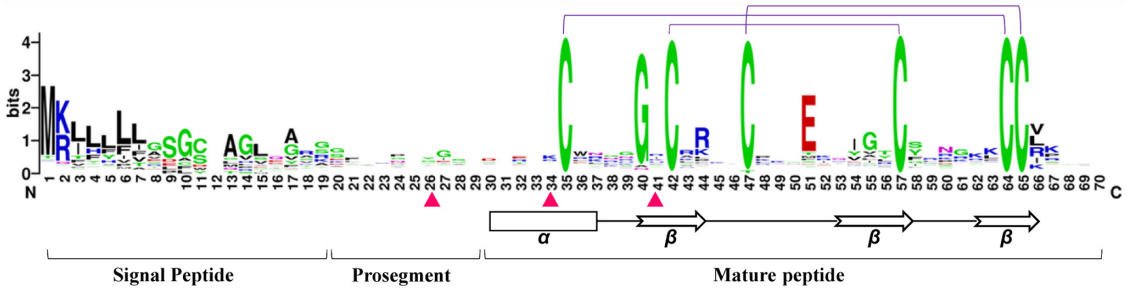

Figure 3. Sequence logo of vertebrate β-defensins. Peptide sequence logo of all 490 β-defensins related to this study. Sites indicated by '▲' have been found to be under positive natural selection, which shows selection [significant ($P < 0.05$) d_N/d_S at these amino acid sites]. These sites are detected by both M2 and M8 models (see Table 1 for detail). The positions of the α-helix, β-strand, signal/prosegment peptide, and mature peptide are all indicated below the sequence logo.

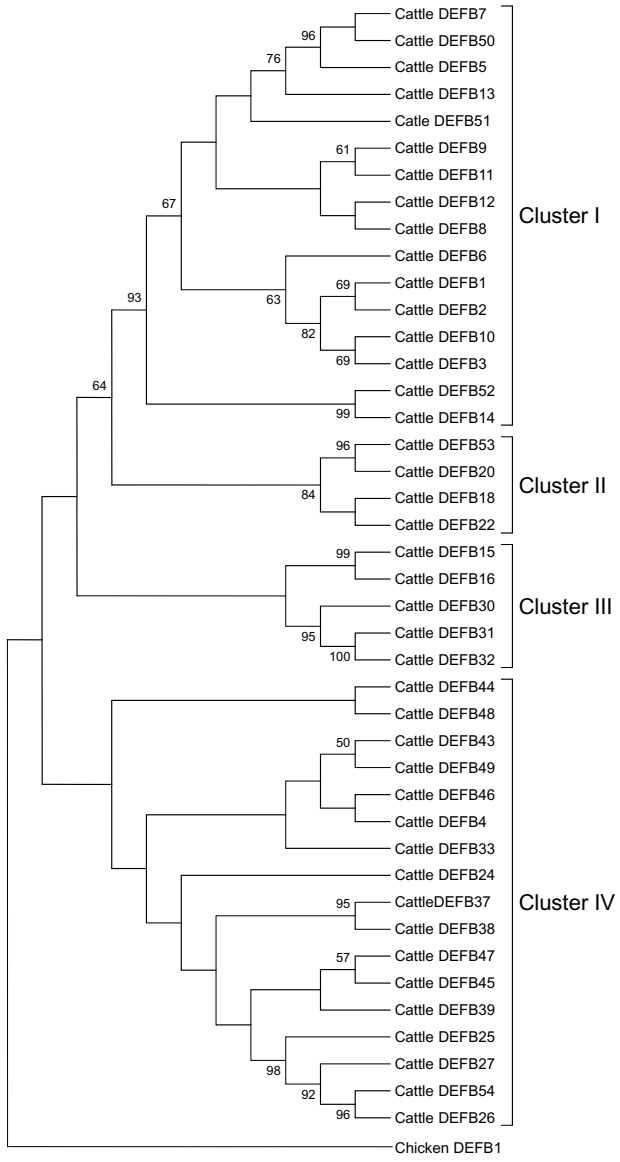

Figure 4. Phylogenetic tree of cattle β-defensin peptides. The full-length protein sequences of the cattle β-defensins were used to construct a phylogenetic tree using the neighbor-joining method. In this tree, a chicken β-defensin sequence is used as out-group. Bootstrap values were obtained by testing the tree 1,000 times and those greater than 50% are shown.

To identify whether there are specific residues or motifs in each of the four phylogenetic clusters in cattle, we analyzed the peptide sequences of cattle β-defensins. The β-defensin peptides share some common characteristics: cationic properties, small size, and six conserved cysteine residues. However, β-defensins are a rapidly evolving gene family with relatively less sequence similarity between paralogs. To understand the structural features of cattle β-defensin peptides, we analyzed the amino acid sequences of signal peptide, prosegment, and mature peptide regions. From the alignment of the peptide sequences of all intact β-defensin genes in cattle (Fig. 5), we identified five potential residues or motif markers that can distinguish the four clusters of cattle β-defensin sequences. All five residues or motif markers are located in the signal peptide and prosegment region. The Cluster I sequences have comparative conserved LH, H(Y), A(X), F(L), and LSA(S) motifs (Fig. 5). Here, "X" represents any amino acid that appeared due to the substitution of conserved residue at a particular position. In contrast, the motifs present at the same positions in Clusters II, III, and IV cattle β-defensin sequences are relatively less conserved (Fig. 5). Although the identified markers are mostly cluster specific, in certain cases, the conserved amino acid residue or motif in a particular position are shared by two or three phylogenetic clusters. For example, cattle β-defensin sequences in Cluster I possess His residues at position 4 (numbering is according to the cattle sequence *DEFB7* in Fig. 5 excluding gaps), whereas conserved sequences at the same position in Clusters II, III, and IV contain Leu residues. To some extent, the variety of these residues or motifs indicates the diverse expression patterns of β-defensin genes in cattle.[49]

In the mature peptide, the conserved Gly-X-Cys^{IV} motif is signally attractive. The conserved Gly-X-Cys^{IV} motif forms a β-bulge region, which is thought to be responsible for forming a twist in the β-sheets and to be essential for the correct folding.[50] In cattle β-defensins, we found two conserved Gly-X-Cys motifs (GXC^{II} and GXC^{IV}). Clusters I, II, and III have two highly conserved motifs (GXC^{II} and GXC^{IV}), while Cluster IV has only one conserved motif (GXC^{II}).

The net charge and the hydrophobicity determine the antimicrobial activity of β-defensin peptides.[51] As showed in Figure 5, most of the positively charged residues, such as Arg and Lys, are mainly located within the C-terminus of mature peptide in all four cattle β-defensin clusters. There are more negatively charged amino acids Asp/Glu in Cluster IV cattle β-defensins. This is rare in the other three cattle clusters. In cattle β-defensins, some hydrophobic residues, such as Ala, Val, Leu, and Phe, are all rich and reasonably well conserved in the signal/prosegment peptide.

Protein structural characteristics of vertebrate β-defensins. Functional research revealed that β-defensins not only perform diverse functions in protecting against pathogens but are also involved in regulation of the immune response and reproduction.[49] The functional diversity may depend on the structural variation in the β-defensins. To resolve protein structure of vertebrate β-defensins, we have built theoretical models by homology (Fig. 6).

In the mature β-defensin peptides, the overall fold of the β-defensins is composed of three β-strands arranged into an antiparallel β-sheet. Different β-defensins share a remarkable similarity at the level of secondary and tertiary structure, in spite of very low similarity in the amino acid sequence. The structural framework does tolerate a substantial variability in amino acid substitution. These models suggest that different β-defensins display the canonical β-defensin disulfide arrangement and a similar fold.

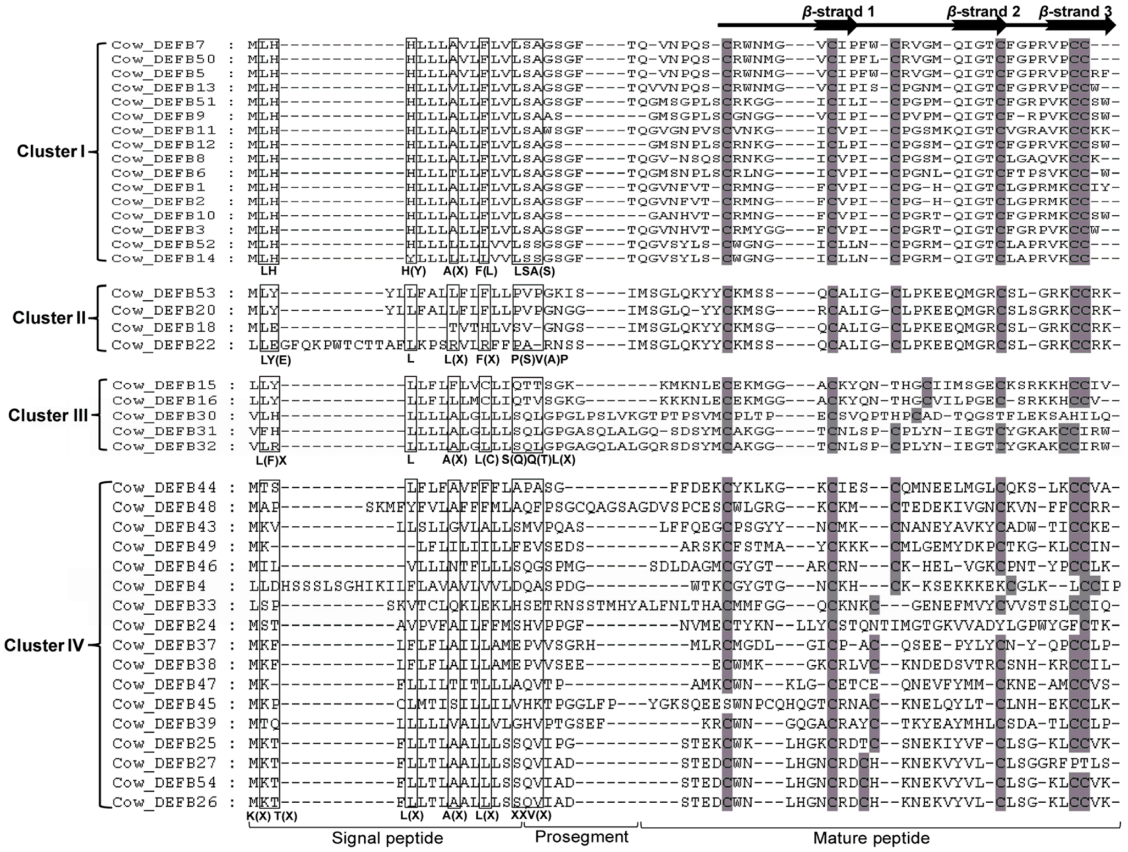

Figure 5. Multiple sequence alignment of the cattle β-defensin amino acid sequences. The six conserved Cys residues are highlighted in gray and the amino acid residues or motifs that distinguish the four phylogenetic clusters (Clusters I, II, III, and IV) are marked with boxes. The positions of the β-strands, signal peptide, prosegment, and mature peptide are all indicated.

As indicated in Figure 6, the α-helix flanking the β-sheet in human and other vertebrate β-defensin peptides is believed to be involved in anchoring β-defensin to the cell wall and to play important roles in killing the pathogens. Compared with β-defensins in other vertebrates, the classic α-helix region is particularly absent in all four cattle β-defensin peptide clusters (Fig. 6).

It is known that most AMPs show cationic properties, which is essential for their biological activity.[52] The surface charge distribution endows the β-defensin peptides with amphipathicity, which allows them to insert into the cell membrane of pathogens and act as antibacterials.[53] The most significant difference between β-defensins is that they differ markedly in surface charge distribution (Fig. 6). The surface electrostatic potential distribution is the determinant of functional specificity and difference between vertebrate β-defensins. These variations in electrostatic surface distribution indicate that these β-defensin proteins have distinct mechanisms for the pattern of antimicrobial activity. Another remarkable difference among β-defensins is the variation in loop sizes and orientations (Fig. 6). Subtle variation in loop sizes and orientations can manipulate the fold structure and protein conformation. Thus, β-defensins perform specific antibiotic activity and diverse functions.

Discussion

Frequent changes in the number of β-defensin genes in vertebrate evolution. In this study, we obtained and analyzed intact β-defensin gene sequences from a wide range of vertebrate taxa. Data mining methods based on high-coverage genome sequences are considered as a reliable method to detect β-defensin genes. Because of genome quality and searching difficulties, partial sequences and pseudogenes of β-defensins were not included in this study. The results indicated that there are frequent changes in the number of intact β-defensins within vertebrate lineages. Previous studies showed that primate genomes encode α, β, and θ-defensins, but the cattle genome contains only the β-defensin subfamily.[52] In this study, we found 42 β-defensins in the cattle genome representing a large expansion compared to other mammals. We hypothesized that this extensive duplication and divergence of β-defensins involved in innate immunity may be due to the substantial load of microorganisms present in the rumen of cattle. This hypothesis runs parallel with the so-called "niche adaptation hypothesis", which suggests that the evolution of the rumen led to a requirement for more sophisticated immune mechanisms to manage the interface between microbes and the animal host.[49] The large number of microorganisms increases the risk of infections at mucosal

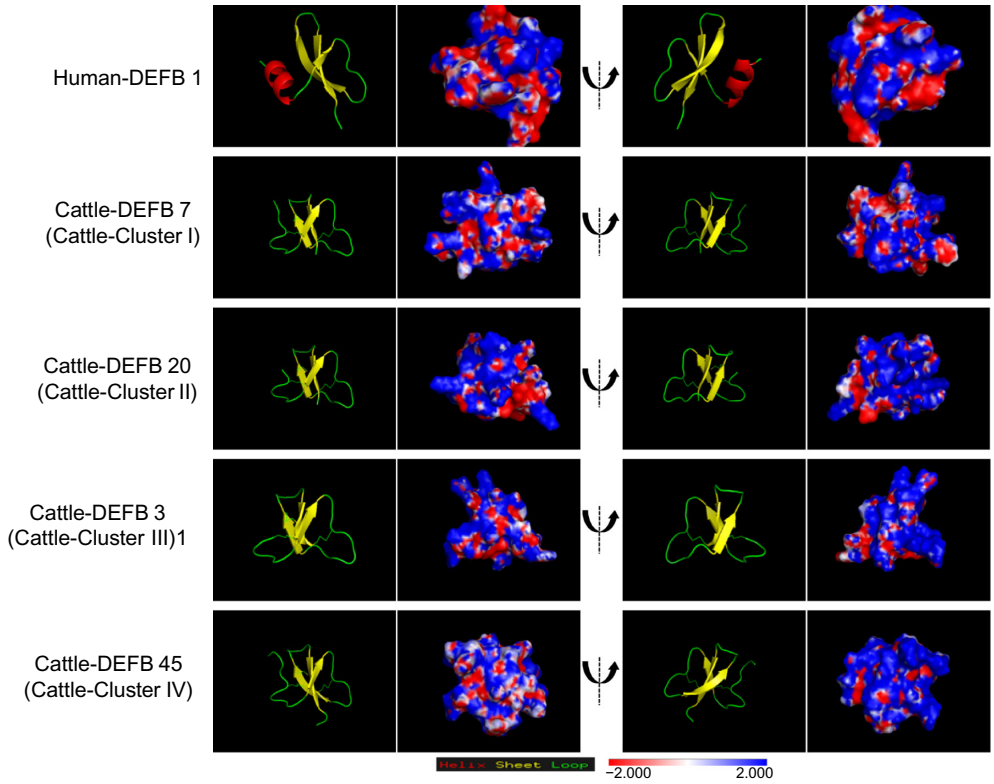

Figure 6. Theoretical three-dimensional models and surface electrostatic potential for the vertebrate β-defensins. The left and right views of each β-defensin structure are all shown. The electrostatic potential (±2 kcal/mol·e) is colored with red (−) or blue (+). Existing templates (cattle-DEFB7: Bovine neutrophil beta-defensin 12; cattle-DEFB20: BETA-DEFENSIN 2; cattle-DEFB31: Resistin; cattle-DEFB45: Beta-defensin 106) were used for modeling these proteins.

surfaces and enhances positive selection for the traits that enable stronger and more diversified innate immune responses at these locations.[28] The β-defensin gene number in a species is relevant to the ever-changing microbial challenges in the ecological niches which they inhabit. In addition to cattle, some mammals such as rat, mouse, and especially microbats have extensive β-defensin gene repertoires. This is presumably due to a unique and diverse microbial environment in their habitats. It is apparent that there has been a rapid evolution of β-defensin genes in mammals through gene or genome duplication and sequence diversification. The rapid evolution and diversity of the β-defensin gene family, considered in the context of their varied antimicrobial and immune regulatory activities, indicates myriad functions for β-defensins in mammalian host defense.

Extensive gene and genome duplications[54] have been regarded as an important raw material for the evolution of acquired immunity. It has long been assumed that β-defensins evolved as an AMP to oppose potentially harmful microorganisms in the environment.[55] Species-specific β-defensins may be required for animals to better deal with the species-specific microbial challenges that they face. Therefore, cattle would be expected to develop a greater level of β-defensins. Marine mammals have fewer β-defensin genes (dolphin and manatee have 8 and 17 β-defensins, respectively) compared with most land mammals, and this may be due to their aquatic habitat since there are fewer prokaryotes in freshwaters and saline lakes.[56] Platypus has a relatively small number of β-defensins in mammals, even less than chicken with 14 β-defensins. Platypus has a blend of mammalian and reptilian features. It is the most remarkable mammal, not only because it lays eggs but also because it is venomous. A previous study identified three *Ornithorhynchus* venom defensin–like peptide (OvDLP) genes, which produce the major components of platypus venom.[57] Among amphibians, only one salamander and one Chinese brown frog β-defensin has been identified in previous studies.[17,18] AMPs dermaseptins and magainins have only been identified from frog skin.[58,59] We suppose that by compensating for species-specific AMPs such as OvDLP and magainins, platypus and frogs develop less β-defensins during the evolution of vertebrate. The presence of multiple, divergent subsets of β-defensins in each species may help animals to better cope with different microbial challenges in the ecological niches they inhabit.

Selective pressures on the evolution of β-defensin genes. A previous study has showed that gene duplication followed by positive selection has indeed been observed in β-defensin gene families involved in immune responses.[21] To further understand the driving force for sequence divergence of β-defensins during evolution, we tested whether sites under positive Darwinian selection occur in the vertebrate β-defensin mature peptide domain by estimating selection

for (d_N/d_S) for each codon in β-defensin using the PAML 4.7 package.[43] We found three sites subject to positive selection and that two of the three positively selected sites are located in the mature peptide. These results support the hypothesis that natural selection has acted to diversify the functionally active mature β-defensin region.[32] Selective pressure analysis revealed that the N-terminal of the mature peptide and prosegment peptide are all important for vertebrate β-defensins.

Positive selection can greatly accelerate the rate of amino acid change. Divergence of these β-defensin genes often leads to either an additional layer of functional redundancy or acquisition of functional novelties, both of which conceivably help the host cope more effectively with a broader range of pathogens. Differential production of species-specific copies of β-defensins may help species occupy different ecological niches.

Coherence between the structure and activity of β-defensins. Analysis of the structural and functional characteristics of the β-defensins highlights the ability to engineer these peptides to gain a better understanding of their function. The β-defensin peptides share some common characteristics: cationic properties, small size, and three disulfide linkages.

The classic α-helix in β-defensins is particularly absent in cattle β-defensin peptides. In consideration of the fact that the α-helical region may anchor the β-defensins to the bacterial cell wall and play important roles in the immune specificity of β-defensins, we inferred that there are still uncharted mechanisms for immune specificity of β-defensins in cattle.

The biological activities of the β-defensins result solely from the changes in the specific mutation sites, characterized either by the alterations to the geometry of molecular surface or to its physicochemical properties (such as electric charge, hydrophobicity, etc.). It is believed that positively charged amino acid residues are universally toward the C-terminus and play antimicrobial function.[60] So variations in single–amino acid substitutions and N-terminal deletions do not affect the charge or adequately alter the hydrophobicity. However, these changes can alter the bacterial susceptibility and the overall rate of killing.[61]

Conclusions

Investigation of the β-defensin genes in vertebrates has revealed extensive gene gains and gene losses in this study. The number of intact β-defensin genes varies from 1 in the western clawed frog to 42 in the cattle. Multiple species-specific gene gains and gene losses have occurred throughout the evolution of vertebrates. Selective pressure tests show that there are three amino acid sites under significant positive selection and highlight the important value of prosegment/mature peptide regions for antibiotic activity. Structural characteristics analysis suggested that structural diversity determines diverse functions performed by β-defensins.

Acknowledgments

We thank Prof. Douglas Armstrong for his comments and corrections of the manuscript.

Author Contributions

Conceived and designed the experiments: JT, DL, MY, QZ. Performed the experiments: JT, DL, QL, LZ, XF. Analyzed the data: JT, DL, LZ, XZ, HX, YY. Contributed materials/analysis tools: DL, MY. Wrote the paper: JT, DL, MY, UG. Agree with manuscript results and conclusions: JT, DL, QL, LZ, QZ, UG, XF, HX, YY, XZ, MY. Jointly developed the structure and arguments for the paper: JT, DL, QL, LZ, QZ, UG, XF, HX, YY, XZ, MY. All authors reviewed and approved of the final manuscript.

Supplementary Files

Supplementary Figure 1. Neighbor-joining tree with species and gene names of the β-defensins from 29 vertebrates. The NJ tree was constructed in MEGA 5 by calculating the proportion difference (p-distance) of aligned amino acid sites of the β-defensin sequences. Zebrafish preprohepcidin1 (GenBank: AY363452.1) and preprohepcidin2 (GenBank: AY363453.1) genes were used as out-groups. The reliability of each branch was tested using 1,000 bootstrap replications. Branch lengths are drawn to scale, which is measured by the number of amino acid substitutions per site.

Supplementary Table 1. Primary structure of selected animal and plant cysteine-rich AMPs

Notes. Cysteine residues are underlined and printed in bold. The "*" represents stop codon. Cysteines paired in disulfide linkages are noted by common numerical subscripts. C-terminal amides are noted by *a*.

Supplementary File 1. Nucleotide sequences of all intact 490 β-defensins.

REFERENCES

1. Martin E, Ganz T, Lehrer RI. Defensins and other endogenous peptide antibiotics of vertebrates. *J Leukoc Biol*. 1995;58(2):128–36.
2. Rolff J. Why did the acquired immune system of vertebrates evolve? *Dev Comp Immunol*. 2007;31(5):476–82.
3. Cormican P, Lloyd AT, Downing T, Connell SJ, Bradley D, O'Farrelly C. The avian toll-like receptor pathway – subtle differences amidst general conformity. *Dev Comp Immunol*. 2009;33(9):967–73.
4. Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol Rev*. 2002;190(1):95–122.
5. Radhakrishnan Y, Fares MA, French FS, Hall SH. Comparative genomic analysis of a mammalian beta-defensin gene cluster. *Physiol Genomics*. 2007;30(3):213–22.
6. Zaiou M, Gallo RL. Cathelicidins, essential gene-encoded mammalian antibiotics. *J Mol Med (Berl)*. 2002;80(9):549–61.
7. Selsted ME, Ouellette AJ. Mammalian defensins in the antimicrobial immune response. *Nat Immunol*. 2005;6(6):551–7.
8. Zou J, Mercier C, Koussounadis A, Secombes C. Discovery of multiple beta-defensin like homologues in teleost fish. *Mol Immunol*. 2007;44(4):638–47.
9. Bulet P, Stocklin R. Insect antimicrobial peptides: structures, properties and gene regulation. *Protein Pept Lett*. 2005;12(1):3–11.
10. Castro MS, Fontes W. Plant defense and antimicrobial peptides. *Protein Pept Lett*. 2005;12(1):13–8.
11. Ganz T. Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol*. 2003;3(9):710–20.

12. Zhang L, Yu W, He T, et al. Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science*. 2002;298(5595): 995–1000.
13. Lynn DJ, Higgs R, Gaines S, et al. Bioinformatic discovery and initial characterisation of nine novel antimicrobial peptide genes in the chicken. *Immunogenetics*. 2004;56(3):170–7.
14. Xiao Y, Hughes AL, Ando J, et al. A genome-wide screen identifies a single beta-defensin gene cluster in the chicken: implications for the origin and evolution of mammalian defensins. *BMC Genomics*. 2004;5(1):56.
15. Liu L, Zhao C, Heng HH, Ganz T. The human beta-defensin-1 and alpha-defensins are encoded by adjacent genes: two peptide families with differing disulfide topology share a common ancestry. *Genomics*. 1997;43(3):316–20.
16. Lehrer RI, Ganz T. Defensins of vertebrate animals. *Curr Opin Immunol*. 2002;14(1):96–102.
17. Meng P, Yang S, Shen C, Jiang K, Rong M, Lai R. The first salamander defensin antimicrobial peptide. *PLoS One*. 2013;8(12):e83044.
18. Zhang Z, Zhang B, Nie X, Liu Q, Xie F, Shang D. Transcriptome analysis and identification of genes related to immune function in skin of the Chinese brown frog. *Zoolog Sci*. 2009;26(1):80–6.
19. Dalla Valle L, Benato F, Maistro S, Quinzani S, Alibardi L. Bioinformatic and molecular characterization of beta-defensins-like peptides isolated from the green lizard Anolis carolinensis. *Dev Comp Immunol*. 2012;36(1):222–9.
20. Hellgren O, Ekblom R. Evolution of a cluster of innate immune genes (beta-defensins) along the ancestral lines of chicken and zebra finch. *Immunome Res*. 2010;6(1):3.
21. Hughes A. Evolutionary diversification of the mammalian defensins. *Cell Mol Life Sci*. 1999;56(1–2):94–103.
22. Aldred PM, Hollox EJ, Armour JA. Copy number polymorphism and expression level variation of the human α-defensin genes DEFA1 and DEFA3. *Hum Mol Genet*. 2005;14(14):2045–52.
23. Tang YQ, Yuan J, Osapay G, et al. A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated α-defensins. *Science*. 1999;286(5439):498–502.
24. Conibear AC, Rosengren KJ, Daly NL, Henriques ST, Craik DJ. The cyclic cysteine ladder in θ-defensins is important for structure and stability, but not antibacterial activity. *J Biol Chem*. 2013;288(15):10830–40.
25. Nguyen TX, Cole AM, Lehrer RI. Evolution of primate θ-defensins: a serpentine path to a sweet tooth. *Peptides*. 2003;24(11):1647–54.
26. Zhu S, Gao B. Evolutionary origin of β-defensins. *Dev Comp Immunol*. 2013;39(1):79–84.
27. Huang Y, Li Y, Burt DW, et al. The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet*. 2013;45(7):776–83.
28. Bovine Genome Sequencing and Analysis Consortium; Elsik CG, Tellam RL, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*. 2009;324(5926):522–8.
29. Schutte BC, Mitros JP, Bartlett JA, et al. Discovery of five conserved β-defensin gene clusters using a computational search strategy. *Proc Natl Acad Sci U S A*. 2002;99(4):2129–33.
30. Huttner KM, Bevins CL. Antimicrobial peptides as mediators of epithelial host defense. *Pediatr Res*. 1999;45(6):785–94.
31. Selsted ME, Tang Y-Q, Morris W, et al. Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils. *J Biol Chem*. 1993;268(9):6641–8.
32. Taylor K, Barran PE, Dorin JR. Structure-activity relationships in beta-defensin peptides. *Biopolymers*. 2008;90(1):1–7.
33. Rádis-Baptista G, Oguiura N, Hayashi MA, et al. Nucleotide sequence of crotamine isoform precursors from a single South American rattlesnake (*Crotalus durissus terrificus*). *Toxicon*. 1999;37(7):973–84.
34. Ouellette A, Selsted M. Paneth cell defensins: endogenous peptide components of intestinal host defense. *FASEB J*. 1996;10(11):1280–9.
35. Yeh R-F, Lim LP, Burge CB. Computational inference of homologous gene structures in the human genome. *Genome Res*. 2001;11(5):803–16.
36. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 1997;268(1):78–94.
37. Chenna R, Sugawara H, Koike T, et al. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*. 2003;31(13):3497–500.
38. Zuckerkandl E, Pauling L. *Evolutionary Divergence and Convergence in Proteins*. New York: Academic Press; 1965:97–165.
39. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985;39:783–91.
40. Desriac F, Jégou C, Brillet B, Le Chevalier P, Fleury Y. *Antimicrobial Peptides from Fish*. Boca Raton: CRC Press; 2013:106–41.
41. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28(10):2731–9.
42. Nicholas KB, Nicholas HB Jr. *GeneDoc: Analysis and Visualization of Genetic Variation* [e-book]; 1997. Available at: http://www.nrbsc.org/gfx/genedoc/ebinet.htm.
43. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
44. Kelley LA, Sternberg MJ. Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc*. 2009;4(3):363–71.
45. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014;42:gku340.
46. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*. 2008;29(11):1859–65.
47. Arvidsson PI, Ryder NS, Weiss HM, Hook DF, Escalante J, Seebach D. Exploring the antibacterial and hemolytic activity of shorter- and longer- chain β-, α,β-, and γ- peptides, and of β- peptides from β^2- 3- aza- and β^3- 2- methylidene- amino acids bearing proteinogenic side chains – a survey. *Chem Biodivers*. 2005;2(3):401–20.
48. Semple CA, Taylor K, Eastwood H, Barran PE, Dorin JR. Beta-defensin evolution: selection complexity and clues for residues of functional importance. *Biochem Soc Trans*. 2006;34(pt 2):257–62.
49. Meade KG, Cormican P, Narciandi F, Lloyd A, O'Farrelly C. Bovine β-defensin gene family: opportunities to improve animal health? *Physiol Genomics*. 2014;46(1):17–28.
50. Xie C, Prahl A, Ericksen B, et al. Reconstruction of the conserved β-bulge in mammalian defensins using D-amino acids. *J Biol Chem*. 2005;280(38):32921–9.
51. Kluver E, Schulz-Maronde S, Scheid S, Meyer B, Forssmann WG, Adermann K. Structure-activity relation of human beta-defensin 3: influence of disulfide bonds and cysteine substitution on antimicrobial activity and cytotoxicity. *Biochemistry*. 2005;44(28):9804–16.
52. Bulet P, Stocklin R, Menin L. Anti-microbial peptides: from invertebrates to vertebrates. *Immunol Rev*. 2004;198(1):169–84.
53. Sahl H-G, Pag U, Bonness S, Wagner S, Antcheva N, Tossi A. Mammalian defensins: structures and mechanism of antibiotic activity. *J Leukoc Biol*. 2005;77(4):466–75.
54. McLysaght A, Hokamp K, Wolfe KH. Extensive genomic duplication during early chordate evolution. *Nat Genet*. 2002;31(2):200–4.
55. Hancock RE, Scott MG. The role of antimicrobial peptides in animal defenses. *Proc Natl Acad Sci USA*. 2000;97(16):8856–61.
56. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA*. 1998;95(12):6578–83.
57. Whittington CM, Papenfuss AT, Bansal P, et al. Defensins and the convergent evolution of platypus and reptile venom genes. *Genome Res*. 2008;18(6):986–94.
58. Zairi A, Tangy F, Bouassida K, Hani K. Dermaseptins and magainins: antimicrobial peptides from frogs' skin – new sources for a promising spermicides microbicides – a mini review. *Biomed Res Int*. 2009;2009:452567.
59. Bevins CL, Zasloff M. Peptides from frog skin. *Annu Rev Biochem*. 1990; 59(1):395–414.
60. Bauer F, Schweimer K, Klüver E, et al. Structure determination of human and murine β-defensins reveals structural conservation in the absence of significant sequence similarity. *Protein Sci*. 2001;10(12):2470–9.
61. Krishnakumari V, Singh S, Nagaraj R. Antibacterial activities of synthetic peptides corresponding to the carboxy-terminal region of human β-defensins 1–3. *Peptides*. 2006;27(11):2607–13.