BOVINE NK-LYSINS: GENOMIC EXPANSION AND FUNCTIONAL DIVERSIFICATION

A Dissertation

by

JUNFENG CHEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	James E. Womack
Committee Members,	Sara D. Lawhon
	Penny Riggs
	Friedhelm Schroeder
	Loren Skow
Interdisciplinary Faculty Chair,	Dorothy Shippen

May 2016

Major Subject: Genetics

ABSTRACT

NK-lysin is a cationic antimicrobial peptide (AMP) of the host innate immune system and active against a broad spectrum of targets, including bacteria, viruses, fungi and cancer cells. NK-lysin has been well-studied in humans and pigs, in which each genome contains a single copy of *NK-lysin*. However, the bovine genome has received much less attention with only one study, in which two 405-bp *Bo-lysin* fragments with 94% nucleotide identity were detected among four experimental donors. These two sequences likely represent two different bovine *NK-lysin* genes. The purpose of the present study was to characterize the gene number and the genomic organization of bovine *NK-lysins* and their roles in host resistance to pathogens, including pathogens involved in bovine respiratory diseases (BRD).

Two overlapping BAC clones (CH240-372P1 and CH240-27G22) covering the whole bovine *NK-lysin* region were sequenced by PacBio (Seattle, WA) and the assembled supercontig revealed four *NK-lysin* genes on cattle chromosome 11. *NK2A, NK2B* and *NK2C* are tandemly arrayed as three copies in 30 ~ 35 Kb segments, located 41.8 Kb upstream of *NK1*. All four genes are functional, albeit with differential tissue expression. *NK1, NK2A* and *NK2B* exhibited the highest expression in intestine Peyer's patch while *NK2C* was expressed almost exclusively in lung. Four peptides corresponding to the functional helices 2 & 3 of each gene product were synthesized. Circular dichroism (CD) spectroscopy demonstrated that peptides adopted a more helical secondary structure

upon binding to an anionic model membrane, and a liposome leakage assay suggested that these peptides disrupt the model membrane. To test the potential role in host response to BRD pathogens, we analyzed RNA-seq data to determine expression of each *NK-lysin* gene in bronchial lymph node and lung in healthy animals and animals challenged with BRD pathogens. The expression of some *NK-lysins*, especially *NK2C*, was significantly elevated in most of the challenged animals, indicating potential functions in BRD resistance. Antimicrobial effects of the synthetic peptides against *Escherichia coli*, *Staphylococcus aureus*, *Pasteurella multocida* and *Mannheimia haemolytica* were further confirmed with bacterial killing assays, and their lytic influences on cell membranes were confirmed by transmission electron microscopy (TEM).

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Womack, for his guidance and support throughout the course of this research. His intelligence and kindness made my 5-year overseas study the most wonderful experience, both in science and in life. Thanks to my committee members, Dr. Lawhon, Dr. Riggs, Dr. Schroeder, and Dr. Skow, for pushing me to the best. Special thanks go to Dr. Leif Andersson for his insightful advice on my research and scientific writings.

Thanks also go to my friends and colleagues for making my time at Texas A&M University unforgettable. I also want to extend my gratitude to the Lawhon, Schroeder and Eichler labs (University of Washington) for their expertise and participation in this study.

My family deserves the most credit for their love and encouragements. Sincere thanks to my sister, Junli Chen, for being my best friend and always supporting me.

NOMENCLATURE

AMP	Antimicrobial Peptide
CNV	Copy Number Variation
SD	Segmental Duplication
SNP	Single Nucleotide Polymorphism
CD	Circular Dichroism
INDEL	Insertion/Deletion
BRD	Bovine Respiratory Disease
BVDV	Bovine Viral Diarrhea Virus
BRSV	Bovine Respiratory Syncytial Virus
IBR	Infectious Bovine Rhinotracheitis

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iv

NOMENCLATURE v

TABLE OF CONTENTS vi

LIST OF FIGURES ix

LIST OF TABLES xiii

CHAPTER I INTRODUCTION 1

Cattle: domestication and agricultural importance 1
Antimicrobial peptides 3
Granulysin/NK-lysin 9
Types of genetic variations 11
Mechanism of CNV and SD formation 14

CHAPTER II GENOMIC ORGANIZATION OF THE BOVINE NK-LYSIN GENE FAMILY 18

Introduction 18
Materials and methods 21
Confirmation of NK-lysin inclusion in the BAC clones 21
BAC sequencing with SMRT technology 22
Repeat element analysis 23
Results 24
NK-lysin has expanded to a four-gene family in cattle 24
Bovine NK-lysin gene family arose by tandem segmental duplications 28
Repetitive sequences analysis within bovine NK-lysin gene family 32
Discussion 38

CHAPTER III BIOLOGICAL FUNCTION OF BOVINE NK-LYSIN GENES 40

Introduction 40
Materials and methods 42

Expression profiles 42
Peptide synthesis 44
Circular dichroism (CD) assay 45
Liposome leakage assay (Fluorescence quenching assay) 46
Bacterial killing assay 47
Transmission electron microscopy (TEM) 47
Results 48
Tissue expression of the bovine NK-lysin genes 48
Conformational changes of bovine NK-lysin peptides upon liposome binding 50
Bovine NK-lysin peptides disrupt model membranes 52
Bovine NK-lysin peptides exhibit antimicrobial effects on both Gram-positive and Gram-negative bacteria 54
Synthetic NK1 peptide disrupted the E. coli membrane 56
Discussion 58

CHAPTER IV RESPONSE OF BOVINE NK-LYSINS TO BRD-ASSOCIATED PATHOGENS 60

Introduction 60
Materials and methods 62
RNA-seq data analysis 62
Antimicrobial killing assay 63
Transmission electron microscopy (TEM) 63
Results 64
Elevated expression of bovine NK2C in animals challenged with BRD pathogens 64
Bovine NK-lysin peptides exhibit antimicrobial effects on BRD-causing bacteria P. multocida and M. haemolytica 66
Bovine NK1 peptide lyses Pasteurella multocida cell membranes 68
Discussion 70

CHAPTER V GENETIC VARIATION WITHIN THE BOVINE NK-LYSIN REGION 77

Introduction 77
Materials and methods 80
Deletion of a 9-bp fragment in the third exon of NK1 80
Copy number variation of NK2B in Holstein cattle homozygous for the NK-lysin region 81
Analysis of gene conversions within bovine NK-lysin gene family 82
Results 86
A 9-bp deletion in the third exon of NK1 causes a 3-aa deletion in the peptide 86
Gene conversions within the bovine NK-lysin gene family 89
Copy number variation of NK2B in Holstein cattle 91

Deletion of NK2B in Holstein cattle is in linkage disequilibrium with a SNP from the 770K HD SNP array 94
Discussion 94

CHAPTER VI DISCUSSION AND CONCLUSIONS 96

Discussion 96
Conclusion 99

REFERENCES 102

APPENDIX PHOSPHORUS ASSAY PROTOCOL 137

LIST OF FIGURES

Page

Figure 1. Phylogenetic analysis of seven different bovine *NK-lysin* related mRNA sequences (mRNA-1 – 7) from NCBI nucleotide database. Bootstrap values are shown at 21

Figure 2. BAC clone information. Two overlapping BAC clones (CH240-372P1 and CH240-27G22, highlighted by red box) were placed against the Bos_taurus_UMD_3.1.1 assembly to cover the whole *NK-lysin* region. Two annotated bovine *NK-lysin* genes, *LOC100300483* and *LOC616323* are indicated by black arrows. 25

Figure 3. Sequence comparison between the Bo-NK supercontig and the genome assembly (Bos_taurus_UMD_3.1.1). Mismatches (vertical blue lines), internal duplications (grey box) and four *NK-lysin* gene loci (blue arrows) are indicated. 26

Figure 4. Dotplot analysis of the Bo-NK supercontig against itself. Three segmental duplicate fragments are revealed: SD-NK2A: 62.1-97.1 Kb, SD-NK2B: 97.1-130.1 Kb and SD-NK2C: 130.1-160.3 Kb. 27

Figure 5. Genomic organization of the bovine *NK-lysin* gene family and identified breakpoints. The flanking sequence of JP-2 was used as the reference sequence. 29

Figure 6. Genomic structure and predicted amino acid sequence were compared among four bovine *NK-lysin* genes. (A) Size comparison of five exons and four introns (B) Comparison of the predicted amino acid compositions. The amino acid sequence of *NK2A* was used as the reference, six conserved cysteine residues were indicated. 30

Figure 7. Phylogenetic analysis of the full coding sequences of four bovine *NK-lysins* and *NK-lysin* orthologs in humans, pig, horse, sheep and goat. The accession number for each sequence in the NCBI nucleotide database is indicated and the bootstrap values are shown at branch points. 31

Figure 8. Repeat element analysis within bovine *NK-lysin* region. (A) Distribution of repeat classes within the assembled supercontig Bo-NK. Four junction points and genes are indicated. (B) Distribution of SINE elements within 5 Kb upstream and downstream of each junction point. The portion of each element relative to its consensus sequence is shown on the *y* axis. 36

Figure 9. Comparison of the repeat densities between whole genome assembly (UMD_3.1.1) and the assembled Bo-NK supercontig.....37

Figure 10. Expression of four bovine *NK-lysin*s in lung (L), thymus (T), spleen (S), respiratory lymph node (RLN) and intestinal Peyer's patch (IPP). The expression of each gene in the tissue that exhibited the lowest expression level was set at 1. The average expression levels and standard deviations were calculated from three healthy individuals. 49

Figure 11. Secondary structural changes of four synthetic bovine NK-lysin peptides upon liposome binding. CD spectra of NK-lysin peptides in lipid-free (A) and lipid-bound states (B) are compared. Estimated secondary structural contents, including alpha-helices, beta-sheet, beta-turn and the total secondary structure in lipid-free and lipid-bound states are shown in (C) and (D), respectively. 51

Figure 12. Intensities of the released fluorescent dye (ANTS) from liposome plotted against the concentration of four bovine *NK-lysin* peptides. 53

Figure 13. Antimicrobial activities of four bovine NK-lysin peptides against gram-negative *E. coli* (A) and gram-positive *S. aureus* (B). Cell viabilities were analyzed by comparing the surviving cells after peptide treatment with the control. Error bars represented the standard deviations calculated from four biological replications. 55

Figure 14. Transmission electron micrographs of *E.coli* cells with and without 5 uM *NK1* peptide treatment. (A)(C) Control cells. (B)(D) Cells treated with 5 uM *NK1* peptide for 20 mins. (E) Comparison of the average electron intensities of thirty cells between the control and *NK1*-treated cell groups. 57

Figure 15. Comparisons of the expressions of four bovine *NK-lysin* genes in bronchial lymph node (BLN, A) and lung (LNG, B) among healthy animals and animals challenged with *P.multocida*, *M.bovis*, *M.haemolytica*, BRSV, BVDV and IBR. The Y axis shows the FPKM value, and each black dot represents the FPKM value of an individual. Three or four individuals were included in each control and challenged group. 65

Figure 16. Antimicrobial effects of bovine NK-lysin peptides on BRD-causing pathogens *P. multocida* strains ATCC 43019 (A) and ATCC 43137 (B), *M. haemolytica* ATCC BAA-410 (C) and ATCC 33396 (D). Surviving cell numbers after peptide treatment are shown on the Y axis. Error bars represent the standard deviations calculated from four biological replications. 67

Figure 17. Influence of 20 μM of bovine NK1 peptide on the cell membrane of *Pasteurella multocida* (ATCC 43019) examined by transmission electron microscopy. (A) Control cells. (B) and (C) Cells treated with 20 μM *NK1* peptide for 30 mins. (D) Statistical analysis of the average electron intensity of control cells versus *NK1*-treated cells. Thirty cells from each group were used for statistical analysis. 69

Figure 18. Helical wheel of four synthetic bovine NK-lysin peptides: (A) NK1, (B) NK2A, (C) NK2B and (D) NK2C. Hydrophilic residues (circles), hydrophobic residues (diamonds), negatively charged residues (triangles) and positively charged residues (pentagons) are indicated. Color indicates the hydrophobicity of a residue, in which green represents the most hydrophobic residue, and the amount of green is decreasing proportionally to the hydrophobicity with yellow representing zero hydrophobicity. Hydrophilic residues are coded red with pure red being the most hydrophilic residue, and the amount of red decreasing proportionally to the hydrophilicity. Light blue indicates the potentially charged amino acids. 71

Figure 19. Sequence comparisons of four bovine NK-lysin peptides (NK1, NK2A – C) with pig NK-lysin (Pig-NKL). The four bovine NK-lysin peptide sequences were predicted from the genomic sequence of L1 Domino 99375, donor for the CHORI-240 Bovine BAC Library and PacBio sequencing. Red color indicates alignment with high consensus and blue represents alignment with low consensus. 78

Figure 20. Full coding sequence comparison between two referenced *NK1* mRNAs extracted from the NCBI database (NM_001046578 and XM_010810081) and the sequence (NK1(PacBio)) predicted from the PacBio sequencing result. Red color indicates alignment with high consensus and blue represents alignment with low consensus. 87

Figure 21. A 9-bp INDEL (TGGTGCTCC) in the coding region of bovine *NK1* gene. (A) Individual homozygous for the deletion. (B) Individual heterozygous for the deletion. (C) Individual homozygous for the presence of this 9-bp fragment. Black arrow (↑) indicates the INDEL site and horizontal black arrow (↔) represents the 9-bp INDEL (TGGTGCTCC). 88

Figure 22. Gene conversion events within the bovine *NK-lysin* gene family. (A) (B) Amplicons of *NK2A* in individuals 92, 70, 3850, 93 and 44 were compared with the corresponding regions of *NK2B* and *NK2C* from L1 Domino 99375. All sequences were aligned to the NK2A (Domino). (C) Amplicons of *NK2B* in individuals 43, 47, 26, 23, 93, 91 and 70 were compared with the corresponding regions of *NK2A* and *NK2C* from L1 Domino 99375. All

sequences were aligned to the NK2B (Domino). Red boxes represent the minimum gene conversion tracts. 90

Figure 23. *NK2A, NK2B and NK2C* nucleotide sequence analysis in four homozygous individuals (2527, 2796, 2822 and 3850). Five different clone sequences from four individuals (Seq 1 - 5) were phylogenetically analyzed with four bovine *NK-lysin* reference sequences (*NK1, NK2A, NK2B* and *NK2C*) and corresponding pig (Pig-NKL) and horse (Horse-NKL) orthologs by MEGA 6.0. Bootstrap values are shown at branch points. 93

Figure 24. Phylogenetic analysis of the full coding sequences of four bovine *NK-lysins* and *NK-lysin* orthologs in humans, pig, horse, sheep, goat, water buffalo (WB) and bison. The accession number for each sequence in the NCBI nucleotide database is indicated and the bootstrap values are shown at branch points. 97

LIST OF TABLES

Page

Table 1. List of seven different variants of bovine *NK-lysin* sequences and the corresponding accession numbers from the NCBI nucleotide database. 20

Table 2. Primer information 24

Table 3. List of the repeat classes within the assembled Bo-NK contig and the corresponding frequencies**.** Totals in each category are underlined in the Frequency column. ERV, endogenous retrovirus; hAT, histone acetyltransferase; RTE, recombinational telomere elongation. 33

Table 4. Primer and probe information 44

Table 5. Sequences and properties of four synthetic bovine NK-lysin peptides 45

Table 6. Primer and amplicon information 83

Table 7. Individuals used in the genetic study of bovine *NK-lysin* gene family 84

Table 8. Haplotypes of Holstein cattle homozygous across the entire *NK-lysin* region. Different colors represent different haplotype structures. 92

Table 9. Number of sequenced clones and different sequences obtained from each individual in the analysis of homozygous cattle. 93

CHAPTER I

INTRODUCTION

Cattle: domestication and agricultural importance

Cattle are one of the most common domesticated animals and have significant roles in agriculture throughout the world. Different breeds of cattle have evolved for specialized purposes, including beef cattle for meat, dairy cattle for milk and draft cattle for drawing loads. Other secondary products from cattle include clothing from hides, fertilizer from dung and tools from bones and hoofs. Two major lineages of cattle, the humpless taurine cattle *B. taurus* and the humped zebu cattle *B. indicus*, share a most recent common ancestor approximately 1.7 – 2.0 million years ago [1]. Most commercial cattle breeds in developed countries are descended from the taurine lineage, however, it is a common practice to hybridize purebred taurine cattle and zebu to enhance host resistance to pathogens or performance in harsh environments [2-4].

The domestication of animals and plants was an important step in human history, because it allowed a shift from nomadic to semi-settled or settled lifestyle [5, 6]. Archaeological and genetic analysis suggested that the taurine and zebu cattle were independently domesticated around 10,500 years ago from the same ancestor species, the auroch (*Bos primigenius*) which has been extinct since 1627 [7]. Domestication of cattle occurred in two major regions, the Near East (Turkey) where *Bos taurus* was domesticated and the Indian subcontinent (Pakistan) giving rise to the zebu cattle [8, 9].

Some scholars have proposed a third domestication event from the North African aurochs giving rise to the African taurine cattle [10, 11].

The first cattle in the Americas originated from Portugal and Spain and were brought to the Caribbean island of Hispaniola by Christopher Columbus in 1493. Cattle continued to be imported to the Americas by the Spanish colonists until 1512. In 1512, Caribbean cattle were introduced to Mexico and later moved to Texas, and these Spanish cattle are believed to be the ancestors of the current New World cattle breeds [12], such as the Texas Longhorns which are resistant to cattle tick fever, the Mexican Corriente cattle and the Colombian Romosinuano cattle. Spanish cattle were the only cattle in the North America for several hundred years until 1860s, when the indicus cattle were imported via Jamaica, and gene flow from these indicus cattle into the New World cattle began [13]. Some British breeds were brought to the United States in the late 1700s through the late 1800s and developed into the most common commercial beef breeds in the United States, including the Angus and Hereford. These British breeds generally reach the mature size at an earlier age with a smaller mature size than the Continental European breeds, have less growth potential and yield carcasses with a lower percentage of saleable products, but excel in fertility and calving ease as well as attain higher quality grades. To improve the growth rate and leanness of existing breeds, Continental European breeds were imported into the U.S. in the late 1960s and early 1970s, including the Charolais, Chianina, Gelbvieh, Limousin, Maine Anjou, Salers, and Simmental. In comparison to the British breeds, Continental European breeds are

generally later maturing with larger mature size, produce carcasses with less fat and a higher percentage of saleable products, but have lower quality carcass grades and more calving difficulty when mated to cattle of the British breeds.

Antimicrobial peptides

Antimicrobial peptides (AMPs), also called host defense peptides (HDPs), are small cationic molecules consisting of 10 ~ 50 amino acids. These peptides usually contain a large proportion of positively charged and hydrophobic residues, which are two major determinants of the antimicrobial activities. AMPs are effector molecules in the host innate immune system and widespread in both plant and animal kingdoms, suggesting evolutionarily conserved roles in multicellular organisms [14-16]. The reservoir of the identified AMPs is so diverse that it is difficult to categorize all the molecules into completely distinct groups, but they can be broadly divided into five classes on the basis of their secondary structures [17]. The first class consists of the anionic peptides, which are usually present in the surfactant extracts, bronchoalveolar lavage fluid and airway epithelial cells. An example is dermcidin in humans [18]. The second class mainly contains the linear cationic peptides without disulfide bonds, exemplified by the silk moth's cecropin and the African frog's magainin [19, 20]. The third class is composed of linear peptides characterized by a predominance of one or two specific residues, such as the proline-arginine-rich peptide (PR-39) [21]. Peptides of this group are very flexible in solution because they lack of cysteine residues. Cationic peptides that are fragments of larger AMP molecules are members of the fourth class. For example, cathelicidins are

mature antimicrobial peptides which are part of the C-termini of larger precursors whose N-termini share a homology with porcine cathelin [22]. The fifth class includes the peptides with six conserved cysteine residues which form three intramolecular disulfide bonds. These molecules usually display a defined anti-parallel β-sheet structure, such as the mammalian β-defensins [23, 24]. Some researchers suggested an additional class of AMPs, most of which belong to the SAPLIP (saposin-like protein) family, that share the conserved feature of six invariant cysteine residues and three disulfide bridges and a conserved globular fold with five helices, the saposin fold. Members in this class include the saposins [25], pulmonary surfactant proteins B [26], acid sphingomyelinases(ASMs) [27], acyloxyacyl hydrolases(AOAH) [28], plant aspartic proteases(AP) [29], amoebapores [30], countin [31] and the NK-lysin molecules [32, 33] which are the focus of this study.

AMPs were discovered in 1939, when Dubos demonstrated that an extract from a soil *Bacillus* strain protected mice from *Pneumococci* infection [34]. An AMP molecule was identified from this extract and named as gramicidin [35]. The first reported AMP molecule in animals was a defensin, which was isolated from rabbits in 1956 [36]. Since then, a substantial number of natural AMPs have been identified in both prokaryotes and eukaryotes on the basis of their antimicrobial activities. Most of these natural AMPs exhibit a wide spectrum of antimicrobial activities against Gram-positive and Gram-negative bacteria [21], fungi [37], viruses [38], protozoa [39] and cancer cells [40]. The significance of AMPs in host immune systems is not limited to their broad antimicrobial

effects, they are also important regulators of immune responses. The first non-microbicidal activity of AMPs was reported in 1989 when the neutrophil derived α-defensins were found to be chemotactic towards human monocytes [41]. Subsequent studies demonstrated that these α-defensins were also chemoattractive to (CD4+/CD45RA+) CD4+ and CD8+ T cells, immature dendritic cells (iDC), but not to memory (CD4+/CD45RO+) T cells [42]. On the other hand, an enhanced IgG antibody response was observed when co-administering ovalbumin (OVA) with α-defensins HNP1-3 in mice compared to OVA alone [43], indicating a role in the modulation of adaptive immunity. Other non-microbicidal activity of an AMP molecule includes its ability to inhibit pro-inflammatory responses induced by LPS through suppressing the LPS- induced gene expressions in macrophages [44]. AMPs have also been suggested to contribute to wound healing. The expression of an antimicrobial gene LL-37 was observed to be significantly enhanced in skin wounds *in vivo*, and decreased to the lowest level upon wound closure [45].

The potent antimicrobial activity of an AMP molecule is attributed to its specific amino acid composition, including a net positive charge, hydrophobicity and amphipathicity. AMPs are proposed to exhibit their antimicrobial effects by forming pores in target membranes following three general steps. The first step of every AMP-mediated killing is the attraction of the molecule to a target membrane, by the electrostatic interaction between the cationic AMP and an anionic target membrane. The negatively charged lipopolysaccharide (LPS) and other anionic lipids in the outer membranes of Gram-

negative bacteria and the teichoic acids of Gram-positive bacteria membranes are all targets of cationic AMP molecules. Cancer cells are also the targets of an AMP due to an increased fraction of anionic molecules in their cell membranes, such as the phosphatidylserine (PS) [46] and O-glycosylated mucins [47]. In contrast, AMPs exhibit low cytotoxicity to normal host cells due to the overall neutral charge of their cell membranes which are mainly comprised of zwitterionic phospholipids [48].

The second step involves the attachment of an AMP to the target membrane and the ensuing occurrence of peptide-membrane interaction, which contains two physically distinct states. The target membrane will be stretched at a low peptide/lipid ratio and peptides begin to orientate and insert into the bilayers at a high peptide/lipid ratio.

The third step is the pore formation in target membranes caused by AMPs after insertion. Three models are proposed to explain this mechanism: the ‘carpet model’, the ‘barrel-stave model’ and the ‘toroidal-pore model’ [21]. In the ‘carpet model’, peptides accumulate in parallel on the membrane, covering the cell surface in a carpet-like manner. When the peptide concentration reaches a critical threshold, they begin to disrupt the bilayers in a detergent-like manner, resulting in the formation of micelles. In the ‘barrel-stave model’, peptide helices form a bundle in the membrane with a central lumen, being like a barrel with the peptide helices as the staves. The hydrophobic face aligns with the lipid core region of the bilayer and the hydrophilic face form the interior region of the pore. In the ‘toroidal-pore model’, the polar faces of the peptides are

always associated with the polar head groups of the lipids even when peptides are perpendicularly inserted in the lipid bilayers, resulting in the water core lined by both the inserted peptides and the lipid head groups. Despite the existence of three proposed models, they are related to each other. The interaction between an AMP molecule and a target membrane is actually a continuous graduation between three models rather than a simple process that can be explained by only one individual model [49]. It is well-known that AMPs are multifunctional molecules. In addition to the well-studied killing mechanism of pore formation, AMPs can also exhibit their influences on targets through other mechanisms, such as inhibition of the synthesis of cell walls, nucleic acids and proteins. Peptidoglygan is the main component of cell walls, and an essential structure responsible for cell shape in bacteria, especially in the Gram-positive bacteria. In contrast, peptidoglygan is not found in eukaryotic cells. Therefore, compounds that can inhibit the synthesis of peptidoglygan are strong candidate therapeutic agents for bacterial infections without causing damage to host cells. Several AMP molecules including the lantibiotics and Lcn972 have been demonstrated to inhibit cell wall synthesis by targeting lipid II, an important precursor of peptidoglycan synthesis [50]. On the other hand, some AMPs are able to penetrate cell membranes and disrupt the synthesis of nucleotide acids or proteins. A 21-residue AMP molecule, Buforin II, was shown to cross the cell membrane and bind to DNA and RNA due to sequence identity between the N-terminal of histone H2A and buforin II [51]. Indolicidin, puroindoline and cathelicidins have also been shown to inhibit the synthesis of nucleic acid and protein [52-54]. Besides the mechanism of pore formation in killing cancer cells because

of the increased negative charge in the cell membrane, AMPs can also target different signaling pathways to provoke cell deaths. For example, magainin can induce cell deaths by increasing the levels of ROS and the caspase-3 activity in cancer cell line HL-60, while tachyplesin induces cell apoptosis by regulating the intracellular potassium concentration in the same cancer cell line [55, 56].

Different from the antibiotics which induce the development of resistance in microbes within a short application period and cause potential threats to the public health [57], the electrostatic interaction between cationic AMPs and anionic target membranes reduce the development of resistance while preserving the efficacy of antimicrobial effects. Therefore, AMPs are considered as candidate antimicrobial drugs. However, relatively few AMPs actually have proceeded into clinical trials [58]. Only two peptides, pexiganan and omiganan [59], demonstrated efficacy in Phase III clinical trials, but are still not approved by the US Food and Drug Administration (FDA). Several factors inhibit the development of AMPs as antibacterial drugs, including the cost of peptide synthesis, resistance to proteolytic degradation and toxicology to host cells [60]. Efforts have been made to address each of the factors. For example, a high-throughput system was developed to synthesize peptides at low cost and screen large numbers of peptides for improved antimicrobial activities [61]. On the other hand, an enhanced resistance to proteolytic degradation can be achieved by replacing the natural residues with D-amino acids or other non-natural amino-acid analogues. The hemolytic toxicity of a newly identified natural AMP or synthetic peptide is usually tested in vitro, and the risk of

hemolytic toxicity can be further alleviated by masking the peptides with liposomal formulation [62].

Granulysin/NK-lysin

Human granulysin and pig NK-lysin are cationic antimicrobial peptides in the innate immune system and present in the granules secreted by the activated cytotoxic T lymphocytes (CTLs) and natural killer (NK) cells [32, 63]. Since the first identification of human granulysin by subtractive hybridization in a search for genes expressed by the activated T lymphocytes [64] and the isolation of NK-lysin from pig small intestine based on its antibacterial activity [63], extensive studies have focused on the characterization of these two novel molecules. Single copies of *granulysin* and *NK-lysin* genes are contained in the genomes of humans and pig. Two splicing isoforms of mRNAs for *granulysin* are identified in humans, the 519 and NKG5, where the NKG5 lacks the 243-bp second exon of the 519 [65, 66]. In contrast, there is only one reported transcript of *NK-lysin* in pig. The mature porcine NK-lysin molecule is approximately 9 kDa and consists of 78 residues, which is processed from a 129-residue precursor (Accession number: Q29075). The intact NK-lysin molecule contains six conserved cysteine residues which form three intrachain disulfide bonds. NMR structure of the porcine NK-lysin reveals a saposin fold, which consists of five alpha-helices, corresponding to the residues 3-18, 24-37, 42-51, 54-61 and 66-72 respectively. Among five helices, helix 1 is located in the core while helices 2 & 3 face one side and helices 4 & 5 face the other side [67].

Consistent with the broad spectrum of targets of other AMPs, NK-lysin molecules and their derivatives are also active against a wide range of microorganisms, including both Gram-positive and Gram-negative bacteria [63], fungi [68], protozoa [69], viruses [70] and cancer cells [71]. One of the most interesting antimicrobial activities of NK-lysin molecules is the capacity to directly kill extracellular *Mycobacterium tuberculosis*, which is particularly resistant to the human immune response [72, 73]. They also exhibit potent effects on intracellular *Mycobacterium tuberculosis* following permeation of the cellular membrane by the pore-forming protein perforin [74]. A 22-residue domain spanning the helix 3 and a disulphide-constrained loop was suggested to be responsible for this anti-mycobacterial activity [73]. In addition to the well-known antimicrobial activities, NK-lysin molecules play several other significant roles in maintaining host health. NK-lysin can stimulate the insulin secretion without changes in cytosolic free calcium concentration, indicating its function in maintaining blood glucose homeostasis [75]. Due to the interaction between cationic NK-lysin molecules and anionic lipopolysaccharides (LPS) [76], NK-lysin is indicated to alleviate endotoxemia, a disease caused by a high level of LPS or endotoxin in the blood. Similar to other AMP molecules, NK-lysin and its derivatives exhibit anti-tumor activity with little effects on normal host cells [71]. This selectivity of effects is attributed to the electrostatic interaction between cationic NK-lysin and anionic phosphatidylserine (PS), an increasing level of component on the surface of cancer cells as cancer aggravates [46] [77].

NK-lysin may possess important functions other than these currently identified. Understanding of the multifaceted molecule in host health will be necessary for designing therapeutic drugs. However, a *NK-lysin* counterpart has not been identified in the mouse genome [78], precluding the gene silencing study in mice. Identification and characterization of this molecule in other animals is essential. *NK-lysin* orthologs have been identified in horse [79], chicken [80] and ruminant species including goat [81], water buffalo [82] and cattle [83]. In addition, *NK-lysin* sequences in two other ruminant species, sheep and bison, can be identified in public databases. Bovine *NK-lysin* was first reported a decade ago [83], when two bovine cDNA fragments, *Bo-lysin 89* and *Bo-lysin 62*, were obtained from each of four different cows. It was unclear whether the detected sequences were from two different *NK-lysin* genes or alleles of a single gene. The aim of this study is to characterize the genomic structure, including the gene number and genomic organization, of the *NK-lysin* gene family in cattle, as well as the biological function of each family member.

Types of genetic variations

Genetic variation refers to the diversity of genomes among individuals within a population or among different populations. Nucleotide diversity is the percentage of the total nucleotides in a genome that differ on average between two individuals in a population and is estimated to be 0.1% - 0.4% in the humans [84, 85]. Genetic variants can be broadly divided into different forms, including small-scale sequence variation (single nucleotide polymorphism and insertion or deletion of a few base pairs) and large-

scale structural variation (copy number variation, duplication, deletion, translocation and inversion).

Single nucleotide polymorphisms (SNPs) are the most common form of genetic variants and refer to single nucleotides that vary within a population of individuals. There are 10 to 30 million SNP sites in the human genome. SNPs within coding regions can be divided into three categories on the basis of the functional influences on gene products, synonymous SNPs, nonsynonymous SNPs and nonsense mutations. Synonymous SNPs are the variants that code for the same amino acids and therefore have no effects on the protein sequences. In contrast, a nonsynonymous SNP is the point mutation that codes for a different amino acid and displays potential effects on the function of the translated protein. For example, sickle-cell disease is caused by a nonsynonymous SNP of A – T which results in the substitution of the 6^{th} residue glutamic acid by valine, known as the E6V mutation. Another type of point mutation is the nonsense mutation, which codes for a premature stop codon leading to a truncated protein. This type of point mutation usually shows a strong impact on the protein and results in a nonfunctional gene product. A nonsense mutation of C – T within the LGR4 gene causes the termination of LGR4 protein at the amino acid position 126 and is found to fully disrupt the protein function. This mutation is strongly associated with several human traits, including the low bone mineral density (BMD), electrolyte imbalance, late onset of menarche, reduced testosterone levels and an increased risk of squamous cell carcinoma of the skin and biliary tract cancer [86].

Another type of small-scale sequence variation is the insertion or deletion of a few base pairs, also called an INDEL. An INDEL with a length that is not a multiple of 3 in the coding region of a gene produces a frameshift mutation, which will completely change the amino acid composition downstream of the INDEL site and thus is very deleterious to the gene function. A cytosine insertion within the coding region of *NOD2* is associated with human Crohn's disease due to the frameshift and the resulting premature stop codon [87]. Therefore, natural selection will purify these deleterious INDELs from the coding region of functional genes. As a result, INDELs occur more frequently within the intergenic and non-coding regions of genes.

In contrast to the small-scale sequence variation which involves an individual nucleotide or a few nucleotides, structural variation usually involves a large genomic region. One of the most common structural variations in the human genome is copy number variation (CNV), which is defined as the difference in the copy number of a large genomic region, ranging from 1 Kb to several megabases, among the genomes of different individuals [88]. A CNV can be tandemly organized within the same chromosome region or at multiple sites on different chromosomes. Duplications (>1 Kb) that are highly identical (90%) are known as low copy repeats (LCRs) or segmental duplications (SDs) based on genome sequence analysis between individuals. SDs frequently lead to local genomic instability and therefore serve as one of the principal mechanisms of gene family expansion [89] and provide substrates for new gene function and development [90, 91].

Both CNVs and SDs have been associated with phenotypic diversity and disease susceptibility [92, 93].

Other structural variations include inversions and translocations. In contrast to the gain or loss of genetic material of a CNV, an inversion or translocation is considered as a balanced structural variation that rearranges the orientation or location of a DNA segment rather than changes its copy number. An inversion occurs when a piece of genomic fragment is broken twice, flipped 180 degrees, and rejoined. A reciprocal translocation is a type of rearrangement in which two nonhomologous fragments exchange the locations. Both forms of structural variations can lead to phenotypic diversity including genetic diseases. For example, a recurrent 400 Kb inversion in the factor VIII gene is associated with human Hemophilia A [94], and a Robertsonian translocation of the chromosome 21 long arm to the long arm of chromosome 14 is responsible for 2 – 4% of the Down syndrome [95].

Mechanism of CNV and SD formation

Homologous recombination (HR) is a biological mechanism for DNA repair. Deletions or duplications will not occur if a damaged DNA fragment is repaired using the homologous sequence at the same chromosomal position on the homologous chromosome as the template. However, recombination between homologous sequences at different chromosomal positions or non-allelic homologous recombination (NAHR) sometimes occurs. NAHR is one of the major contributors to genome rearrangement and

is mediated by highly homologous pre-existing repeats or previously duplicated regions such as the SDs [96-98]. A large number of the identified breakpoints associated with recurrent rearrangements are embedded within these repeat elements or SDs [99]. During meiosis, misalignment and subsequent crossing-over can occur between DNA fragments with high sequence identity at different chromosomal positions on each chromosome, producing gametes with different copies of the fragment between the two NAHR sites. The minimal nucleotide tract length with high sequence identity to each other required for an efficient initiation of recombination is defined as the minimal efficient processing segments (MEPS). The length of MEPS for an occurrence of NAHR during meiosis ranges from 300 – 500 bp, thereby usually resulting in large CNVs [100].

Non-homologous end joining (NHEJ) is another common DNA repair mechanism, which rejoins the ends of a double-strand break (DSB) without the requirement of a homologous template [101]. In general, NHEJ repairs a broken DNA fragment accurately with gain or loss of very little sequences at both broken ends, maintaining the genomic stability [102]. However, mutations in some of the genes that are involved in DSB repair pathways including *RAD50*, *XRS2* and *MRE11* can significantly increase the frequency of extensive deletions at the broken ends in *Saccharomyces cerevisiae* [101]. On the other hand, insertions of extrachromosomal DNA sequences, including the mitochondrial DNA fragments and retrotransposons, at the broken ends are also reported [103, 104]. In humans, NHEL is suggested as the predominant contributor to the interchromosomal recombination and segmental duplication in subtelomeres, associated

with primate evolution [105]. In contrast to NHEJ which does not require sequence homology to guide the repairing, microhomology-mediated end joining (MMEJ) repairs a DNA break by identifying a microhomology of 5- 25 complementary base pairs on both strands and leads to a deletion of the sequence between the annealed microhomologies [106].

The presence of microhomology at the non-homologous recombination site has usually been regarded as the signature of non-replicative MMEJ repair pathway. Recently, increasing evidence implicates that the formation of microhomology at a DNA broken site is also linked to DNA replication pathway and thus a novel mechanism of CNV formation involving the fork stalling and template switching (FoSTeS) has been proposed [107, 108]. Since a double-strand DNA break is more likely to induce a recombination-based DNA repair, including the NAHR, NHEJ and MMEJ, a single-strand DNA (ssDNA) lesion rather than a double-strand break is proposed as the initiating damage in the FoSTeS model. This ssDNA break can cause the stalling of replication forks and successive template-switching to repair the DNA breaks, but results in structural variations due to the errors at replication forks.

Although multiple models have been proposed for the occurrence of a copy number change variant, a single model is usually not enough to account for a given event. Instead, multiple mechanisms work together at different timelines to lead to a duplication or deletion event. For example, a NAHR event might trigger the breakage-

fusion-bridge cycle by forming a dicentric chromosome, and the fusion step might be mediated by an end-joining mechanism. Also, end-joining mechanisms are important in repairing the broken ends that result from other events [106].

CHAPTER II

GENOMIC ORGANIZATION OF THE BOVINE NK-LYSIN GENE FAMILY [*]

Introduction

Copy number variation (CNV) is a form of structural genomic variation, usually as a deletion or a duplication, ranging from 1 Kb to several Mb. CNV is very common in animal genomes, where it covers ~12% of the human genome and is a major contributor to phenotypic diversity [88, 109-112]. In contrast to the single nucleotide polymorphisms (SNPs), CNVs involve large chromosomal regions with the potential for substantial impact on phenotypes and have been associated with a large number of genetic disorders [113-115]. Duplications (>1 Kb) that are highly identical (90%) and exist at multiple locations interchromosomally or intrachromosomally are defined as segmental duplications (SDs). Since SD segments are large in size and share high sequence identity to each other, they often result in chromosomal rearrangements and thus genome instability. Many multigene families are proposed to arise by segmental duplications and subsequent functional differentiations, including those with uniform members such as *ribosomal RNA* genes and those with variable genes like *immunoglobulins*. Several whole genome CNV distribution analyse have been performed among different breeds of cattle, and two independent studies suggested that the bovine *NK-lysin* gene is located in a copy number variation region (CNVR) [112,

[*] Reprinted with permission from "Bovine NK-lysin: Copy number variation and functional diversification" by Chen J, et al, 2015. *Proc Natl Acad Sci U S A*, 112(52):E7223-9, Copyright [2015] by the National Academy of Sciences of the United States of America.

116]. A search of the National Center for Biotechnology Information (NCBI) bovine nucleotide database identified seven different *NK-lysin* related sequences (Table 1), and a phylogenetic analysis of these sequences showed four clades that potentially represented four different bovine *NK-lysin* genes. We designated these genes, *NK1, NK2A*, *NK2B* and *NK2C* (Fig. 1). *NK2A, NK2B* and *NK2C* are closely related to each other while *NK1* is more divergent. The genes corresponding to *NK1* and *NK2A* have previously been annotated as *uncharacterized LOC616323* [Gene ID: LOC616323] and *Bovine GNLY* [Gene ID: 404173], respectively in the bovine reference genome assembly UMD 3.1 (UCSC genome browser). These two genes are tandemly arranged on chromosome 11, while *NK2B* and *NK2C* are not included in the current genome assemblies. The assembly of the bovine *NK-lysin* region in the current genome assemblies is likely collapsed due to the duplications. Resequencing this region with deep coverage and long sequencing reads is necessary to resolve the correct genomic organization of this gene family.

The advent of next-generation sequencing (NGS) technology significantly reduces the sequencing cost and thus advances the de novo sequencing of a new species as well as improving the genome assembly of a sequenced species. However, one of the biggest disadvantages of this widespread technology is the difficulty in assembling a repetitive region due to the short length of the reads that are not long enough to span the repetitive region. As a result, the inability of these short reads to scaffold across repetitive regions results in gaps and incomplete assemblies [117]. Single molecule real time sequencing

(SMRT) is one of the latest sequencing technologies and was commercialized by the Pacific Biosciences in 2011. In contrast to the short reads from NGS technologies, SMRT sequencing can generate extraordinary long reads with most reads > 14 Kb, which enables the correct assembly of a complex repetitive region of the genome [118, 119].

Table 1. List of seven different variants of bovine *NK-lysin* sequences and the corresponding accession numbers from the NCBI nucleotide database.

Sequences	Accession	Cluster
mRNA-1	XM_005192449	NK2A
mRNA-2	XM_005192450	NK2A
mRNA-3	BC114176	NK2B
mRNA-4	AY245798	NK2A
mRNA-5	BC114178	NK2C
mRNA-6	AY245799	NK2C
mRNA-7	NM_001046578	NK1

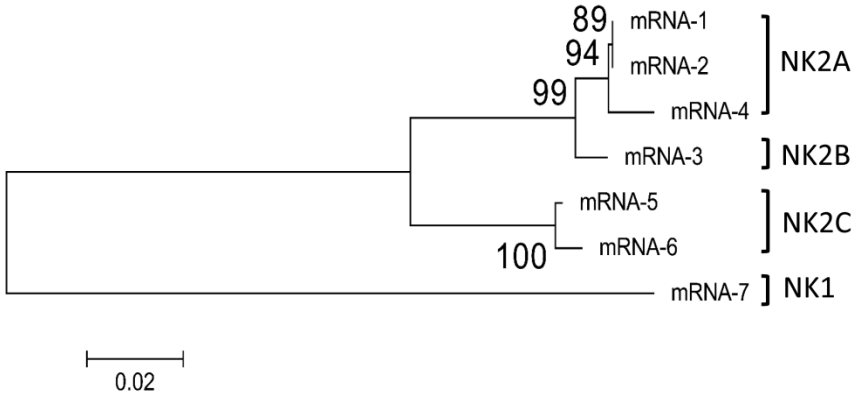


Figure 1. Phylogenetic analysis of seven different bovine *NK-lysin* related mRNA sequences (mRNA-1 – 7) from NCBI nucleotide database. Bootstrap values are shown at branch points.

Materials and methods

Confirmation of NK-lysin inclusion in the BAC clones

Two overlapping BAC clones covering the bovine NK-lysin region were identified from the NCBI clone database and selected from the bovine CHORI-240 BAC library (http://bacpac.chori.org/bovine240.htm). Confirmation of the *NK-lysin* inclusion in the BAC clones was conducted with the Bo-lysin primers (Table 2) which were designed within the conserved region of *NK-lysin* reference sequences. Briefly, DNA of each BAC clone was extracted using the ZR BAC DNA Miniprep Kit (*Zymo Research*). The extracted BAC DNA was then used as the template in a PCR reaction with the mixture of reagents including 5 * PCR buffer, 0.2 μM each of the forward and reverse Bo-lysin primer, 0.8 mM dNTPs, 1 μL (1.25 U/μL) PrimeSTAR GXL DNA polymerase (*Takara*), 25 ng BAC DNA and water to bring the volume to 50 μL. Amplification was carried out

in the condition that includes 30 cycles of 10 s at 98 ℃, 15 s at 62 ℃, 20 s at 68 ℃, and a final extension of 5 min at 68 ℃. Amplicons were purified using QIAquick PCR Purification Kit (*Qiagen*) and cloned into pCR™4Blunt-TOPO® vector (*Life Technologies*). Fifteen colonies from each BAC were inoculated and cultured overnight in 700 μL LB broth with 50 μg/mL ampicillin. Plasmid DNA was extracted using the PureLink® Quick Plasmid Miniprep Kit (*Invitrogen*) and sequenced with the BigDye Terminator v1.1 Cycle Sequencing Kit (*Applied Biosystems*).

BAC sequencing with SMRT technology

BAC clone sequencing was carried out with the SMRT technology (Pacific Biosciences, Seattle, U.S.A), as described previously [119]. Briefly, BAC DNA was sheared to generate fragments of approximately 20 Kbp, which were then ligated with hairpin adaptors at both ends to construct the sequencing libraries using the PacBio DNA Template Prep Kit 2.0 (10 Kbp), followed by the purification with (0.405X) Agencourt® AMPure® beads. The sequencing reactions were performed with P4/C2 chemistry on a PacBio RS. Each clone was sequenced twice in two separate SMRT cells. De novo assembly of the data from each SMRT cell and from the two combined SMRT cells from each clone was performed following the standard SMRT Analysis (v. 2.0.1) pipeline. A further de novo assembly was attempted using combined data from all four SMRT cells, and the final contigs were joined into a single supercontig using Sequencher. The supercontig was then compared with the reference sequence (UMD 3.1.1 assembly) using the miropeats alignment in Parasight [120], and further dotplot analysis of the

supercontig was implemented by UniproUGENE [121, 122]. Four pairs of primers specific for each putative junction point (JP-1, JP-2, JP-3 and JP-4) were tested in the genomic DNA of L1 Domino 99375 to validate the BAC assembly.

Repeat element analysis

Repeat elements within the Bo-NK supercontig and the UMD_3.1.1 assembly were identified and annotated using CENSOR with a bovine specific library downloaded from Repbase that included ancestral sequences [123, 124]. To estimate the density of each repeat family within the whole genome assembly (UMD_3.1.1), the assembled chromosomes were broken into different bins of the same size as the Bo-NK contig (~ 227 Kb) and those consisting of > 10% gaps were excluded from the analysis. Repeat density for each repeat family with > 5 copies in a bin was represented by the repeat coverage per 1000 bp. Ambiguous repeat elements at boundaries were assigned to bins based on a minimum 50% repeat length overlap threshold. Overlaps between repeats and bins were identified using the Genomic Ranges package from Bioconductor [125, 126]. Repeat densities across all bins were used to estimate the empirical cumulative distribution function of each repeat family using the "ecdf" command in R and Bioconductor [127], which was then used to estimate the probability of sampling a bin with repeat density greater than the repeat density of the Bo-NK supercontig ($P(X > x)$). A repeat family was overrepresented in the Bo-NK supercontig if $P(X > x)$ was < 0.05. Finally, repeat annotation plots were generated using the base graphics system in R [127].

Table 2. Primer information

Primers	Forward (5' to 3')	Reverse (5' to 3')
Bo-lysin	ACCCAGCACTCCCACTG	ACATACCTGGCTTGCTTTTG
JP-1	CTAAGTGGCCGGATTGTTGT	CAGGGTCTTCTCCTCTGACG
JP-2	GAAATGCTCTCACAGCAACA	AATAGCAATGAAATGATGATGGT
JP-3	AAAATGCTCTCACAGCAATGAA	AATAGCAATGAAATGATGGTAGCTG
JP-4	GATAGTCTCCCCAACCAGTCAG	GAATTGCTGAGCTGGAAGAAGT
BP-1	GCCTGCCTTCATGGAGTTTA	TGGCACAGGTAATGGGATAA

Results

NK-lysin has expanded to a four-gene family in cattle

To avoid complexity resulting from allelic variation, BAC clones were used in this study. The precise number of genes in the bovine *NK-lysin* family and their genomic organization was determined by sequencing two overlapping BAC clones (CH240-372P1 and CH240-27G22) covering the *NK-lysin* region (Fig. 2). Despite a sequencing coverage depth of > 700 X for both BACs after the first-round of sequencing, each BAC was assembled into six contigs because of the presence of highly repetitive sequences. After a second-round of sequencing, the average coverage was increased to 1310 ~ 1551 X, however, three contigs were still generated for CH240-372P1 and two contigs were generated for CH240-27G22. Since these two BAC clones were overlapping, a final *de novo* assembly of all sequencing data was performed. This analysis produced a two-contig assembly where the two contigs were overlapping by ~ 2 Kb at 100% identity and were subsequently joined into a single contig. This resulted in a linear supercontig of 227,063 bp covering the whole bovine *NK-lysin* region. Overall, the assembled contig

(Bo-NK) was longer than the current genome assembly by ~ 38 Kb, where the corresponding reference sequence was 189,124 bp (Bos_taurus_UMD_3.1 Chr. 11: 48,986,139 – 49,175,262 bp). The difference in length was primarily due to misassemblies in the reference genome, in which repetitive regions containing the *NK2B* and *NK2C* genes, were collapsed (Fig. 3).

Figure 2. BAC clone information. Two overlapping BAC clones (CH240-372P1 and CH240-27G22, highlighted by red box) were placed against the Bos_taurus_UMD_3.1.1 assembly to cover the whole *NK-lysin* region. Two annotated bovine *NK-lysin* genes, *LOC100300483* and *LOC616323* are indicated by black arrows.

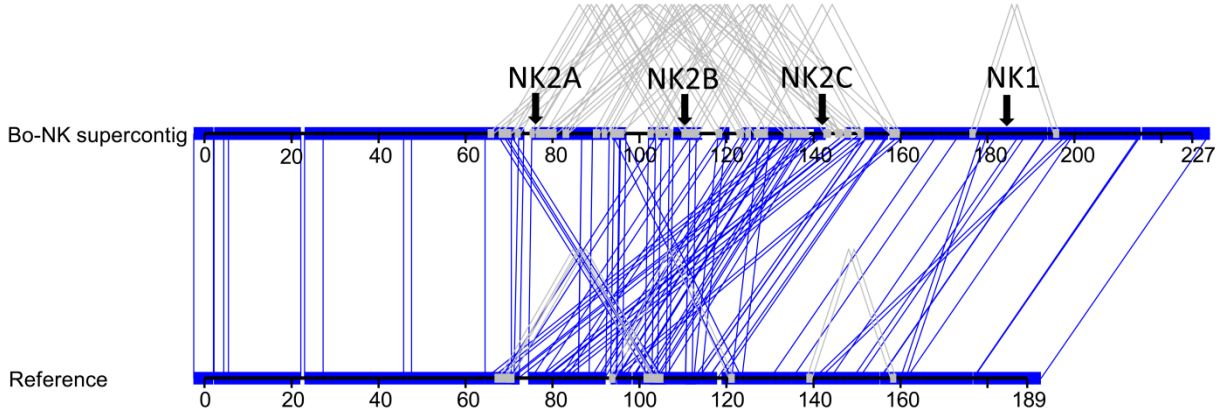


Figure 3. Sequence comparison between the Bo-NK supercontig and the genome assembly (Bos_taurus_UMD_3.1.1). Mismatches (vertical blue lines), internal duplications (grey box) and four *NK-lysin* gene loci (blue arrows) are indicated.

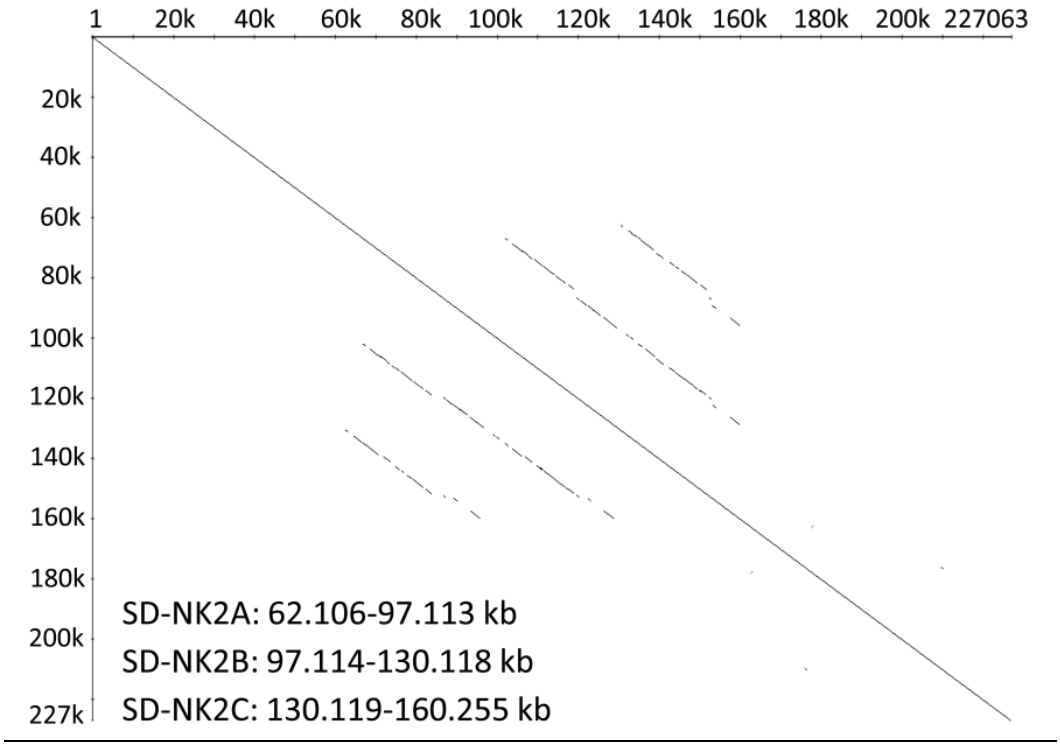
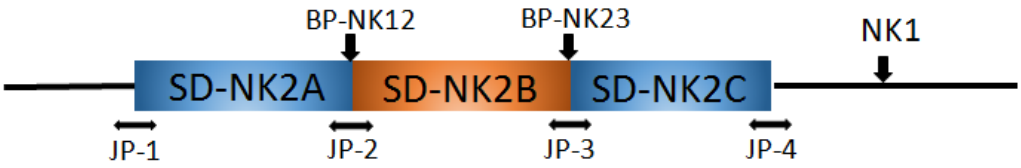


Figure 4. Dotplot analysis of the Bo-NK supercontig against itself. Three segmental duplicate fragments are revealed: SD-NK2A: 62.1-97.1 Kb, SD-NK2B: 97.1-130.1 Kb and SD-NK2C: 130.1-160.3 Kb.

Bovine NK-lysin gene family arose by tandem segmental duplications

Dotplot analysis of the Bo-NK supercontig against itself revealed three segmental duplications (SDs) with ~ 95% sequence identity (SD-NK2A: 62.1-97.1 Kb, SD-NK2B: 97.1-130.1 Kb and SD-NK2C: 130.1-160.3 Kb), each containing one *NK-lysin* gene, and *NK1* was 41.8 Kb downstream from the *NK2C* gene (Fig. 4). Since the SD-NK2C lacked the right end of the duplicated fragment and was shorter than SD-NK2A and SD-NK2B, the flanking sequence of junction point-4 (JP-4) was different from the other three breakpoints, JP-1 – JP-3 (Fig. 5). To confirm the accuracy of the Bo-NK contig, four primer pairs at each junction point were tested using genomic DNA of L1 Domino 99375 (donor for the CHORI-240 Bovine BAC Library). Sanger sequencing showed that JP-1, JP-3 and JP-4 PCR products were perfectly aligned with the Bo-NK contig, while there were six mismatches out of 567 nucleotides between the JP-2 PCR product and the Bo-NK contig. Another primer pair (BP-1) was tested in order to determine whether these six mismatches were due to error in the Sanger sequencing or PacBio sequencing. Sanger sequencing verified six sequencing errors at the BP-NK12 breakpoint in the Bo-NK contig. The Bo-NK contig therefore represented the correct assembly of the bovine *NK-lysin* region, and demonstrated that four *NK-lysin* genes are located in this region on cattle chromosome 11. Complete genomic sequences of four *NK-lysin* genes were compared to determine genetic organization and structure (Fig. 6). All four bovine *NK-lysin* genes contain five exons consistent with the human and pig orthologs. The exon sizes were comparable among four genes whereas the sizes of the introns of *NK1* were larger than introns from the other genes, accounting for a larger genomic size of *NK1*

(Fig. 6A). *NK2A*, *NK2B* and *NK2C* are about 95% identical to each other, but only 85% identical to *NK1*. The predicted amino acid compositions of the four bovine *NK-lysin*s show high sequence identity and include six invariant cysteine residues, conserved among *NK-lysin* molecules in other animals (Fig. 6B). Phylogenetic analysis of the full coding sequences of four bovine *NK-lysins* with *NK-lysin* orthologs in humans, pig, horse, sheep and goat revealed that *NK-lysin* gene family expansion is only in the ruminants, suggesting the divergence of *NK1* and *NK2* cluster in the ancestor of cattle, sheep and goats (Fig. 7).




```
JP-2    GAGA--GCTT GGGTTTGATT CCTGGACAGG GAACTAAGAT CCCACAGGC- TGTGGGGCAA
JP-3    ....--.... .......... .......... .......... .........- ..........
JP-1    ACT.CTA... CCAACAC.GC GG.T.C.... .......... ......T..- ....T.....
JP-4    TG.CAT...- ..AGC.CTGC TGCT.CT--. CTG.....T. G.TTT..T.A ...CC.A.TC
```

Figure 5. Genomic organization of the bovine *NK-lysin* gene family and identified breakpoints. The flanking sequence of JP-2 was used as the reference sequence.

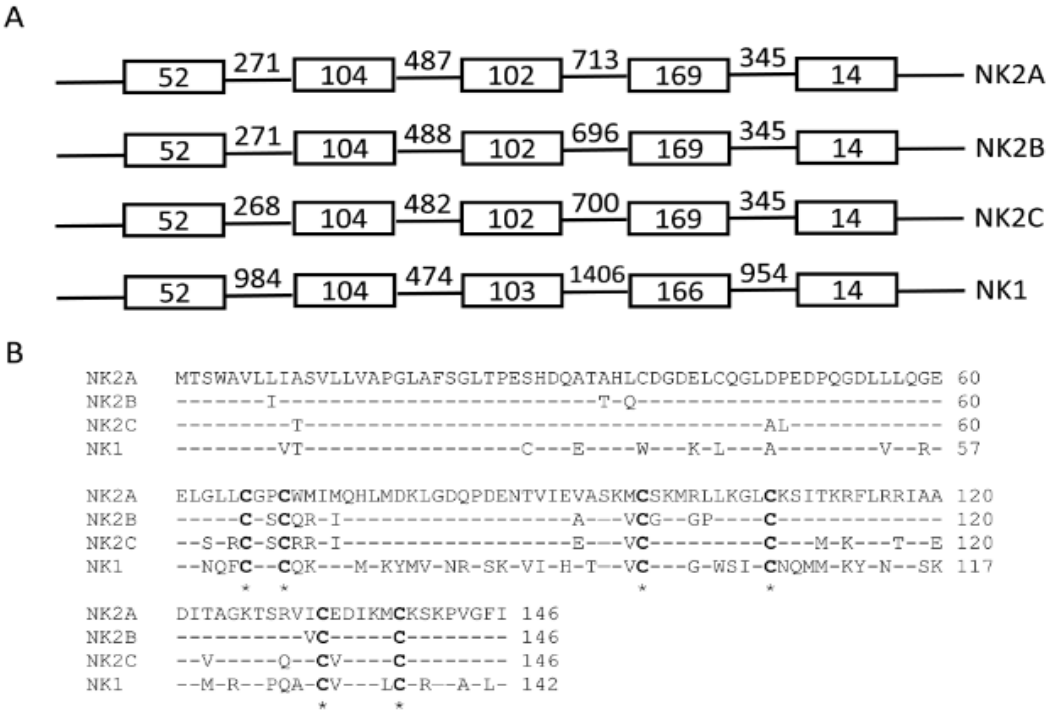


Figure 6. Genomic structure and predicted amino acid sequence were compared among four bovine *NK-lysin* genes. (A) Size comparison of five exons and four introns (B) Comparison of the predicted amino acid compositions. The amino acid sequence of *NK2A* was used as the reference, six conserved cysteine residues were indicated.

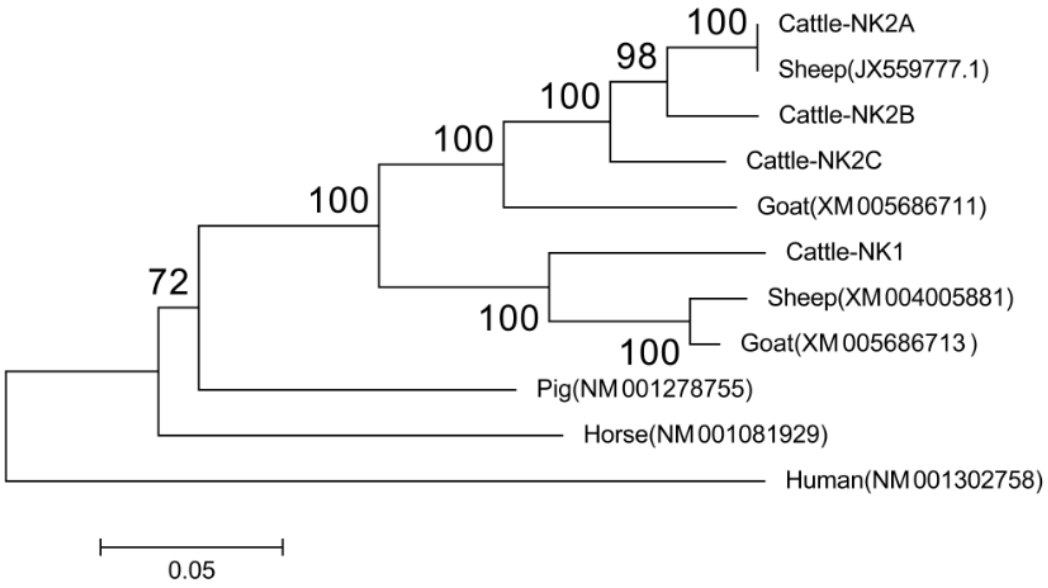


Figure 7. Phylogenetic analysis of the full coding sequences of four bovine *NK-lysins* and *NK-lysin* orthologs in humans, pig, horse, sheep and goat. The accession number for each sequence in the NCBI nucleotide database is indicated and the bootstrap values are shown at branch points.

Repetitive sequences analysis within bovine NK-lysin gene family

Repetitive sequences are usually associated with recombination hotspots in human [128], and chromosomal instability caused by mispairing between such repeats at breakpoints is responsible for several diseases [129, 130]. To gain more insights into the mechanism of *NK-lysin* expansion in cattle, we analyzed the distribution of repeat elements (REs) within this region. The distributions of different repeat classes within the assembled contig are shown in Figure 8A and summarized in Table 3. Overall, the downstream region of each breakpoint is more repetitive than upstream, and the flanking sequences of *NK1* are highly repetitive, consisting of a large percentage of LINES, which is distinct from the rest of the region within this gene family. Several repeat families are overrepresented within the *NK-lysin* region, including two ancient mammalian L1 families, two LTR families and four ruminant/bovine specific SINEs (BOVTA, BTALUL2, CHR-2_BT, and CHR-2A families) (Fig. 9). Due to the enrichment of SINEs around junction points, we plotted the distributions of several ruminant/bovine specific repeat families between 5 Kb upstream and downstream of each junction point (Fig. 8B). The adjacent downstream regions of JP-1, JP-2 and JP-3 are enriched with SINES, especially the BOVTA element. BOVTA elements form a bovine specific repeat family analogous to the primate repeat ALU family, which are usually associated with segmental duplications in humans [96]. These results demonstrate that the fragments flanking breakpoints share high homology, and could contribute to unequal crossover during meiosis and structural instability within the bovine *NK-lysin* gene family.

Table 3. List of the repeat classes within the assembled Bo-NK contig and the corresponding frequencies**.** Totals in each category are underlined in the Frequency column. ERV, endogenous retrovirus; hAT, histone acetyltransferase; RTE, recombinational telomere elongation.

Class	Superfamily	Family	Frequency
LINEs			**147**
	L1		**102**
		HAL1	3
		L1_BT	3
		L1-2_BT	22
		L1-3_BT	13
		L1-BT	1
		L1MA10	12
		L1MB6_5	4
		L1MB7	13
		L1MC3	4
		L1MC4B	15
		L1MC5	1
		L1ME3C_3end	1
		L1ME3D_3end	3
		L1ME3E_3end	1
		L1ME5	1
		L1P_MA2	5
	L2		**9**
		L2	8
		L2B	1
	RTE		**36**
		BovB	36
SINEs			**225**
	SINE		**24**
		BCS	1
		BOVA2	23
	SINE2/tRNA		**189**
		Bov-tA1	28
		Bov-tA2	10
		Bov-tA3	4

Table 3. Continued

Class	Superfamily	Family	Frequency
		Bovc-tA2	11
		BOVTA	40
		BTALUL1	1
		CHR-2_BT	13
		CHR-2A	9
		CHR-2B	15
		CHRL	9
		CHRL1_BT	1
		MIR	11
		MIR3	3
		MIRb	13
		MIRc	5
		SINE2-1_BT	6
		SINE2-2_BT	8
		SINE2-3_BT	1
		THER1	1
	RTE		**12**
		BTALUL2	**12**
ERV			**44**
	ERV1		**15**
		BtERVF2_I	1
		ERV1-2-I_BT	2
		LTR1_BT	6
		LTR11_BT	3
		LTR39B_BT	1
		LTR39D_BT	1
		MER41_BT	1
	ERV2		**1**
		ERV2-1-LTR_BT	1
	ERV3		**28**
		LTR33C	1
		LTR67B	5
		MLT1E2	9

Table 3. Continued

Class	Superfamily	Family	Frequency
		MLT1F	1
		MLT1F1	2
		MLT1J	2
		MLT1J1	1
		MLT1J2	3
		MLT1M	4
LTR			**3**
	Gypsy		**3**
		LTR88b	3
DNA Transposon			**8**
	DNA transposon		**1**
		X25_DNA	1
	hAT		**3**
		Charlie13a	1
		MER91B	1
		UCON52	1
	Mariner/Tc1		**4**
		MER47B	2
		TIGGER5_B	1
		TIGGER5A	1

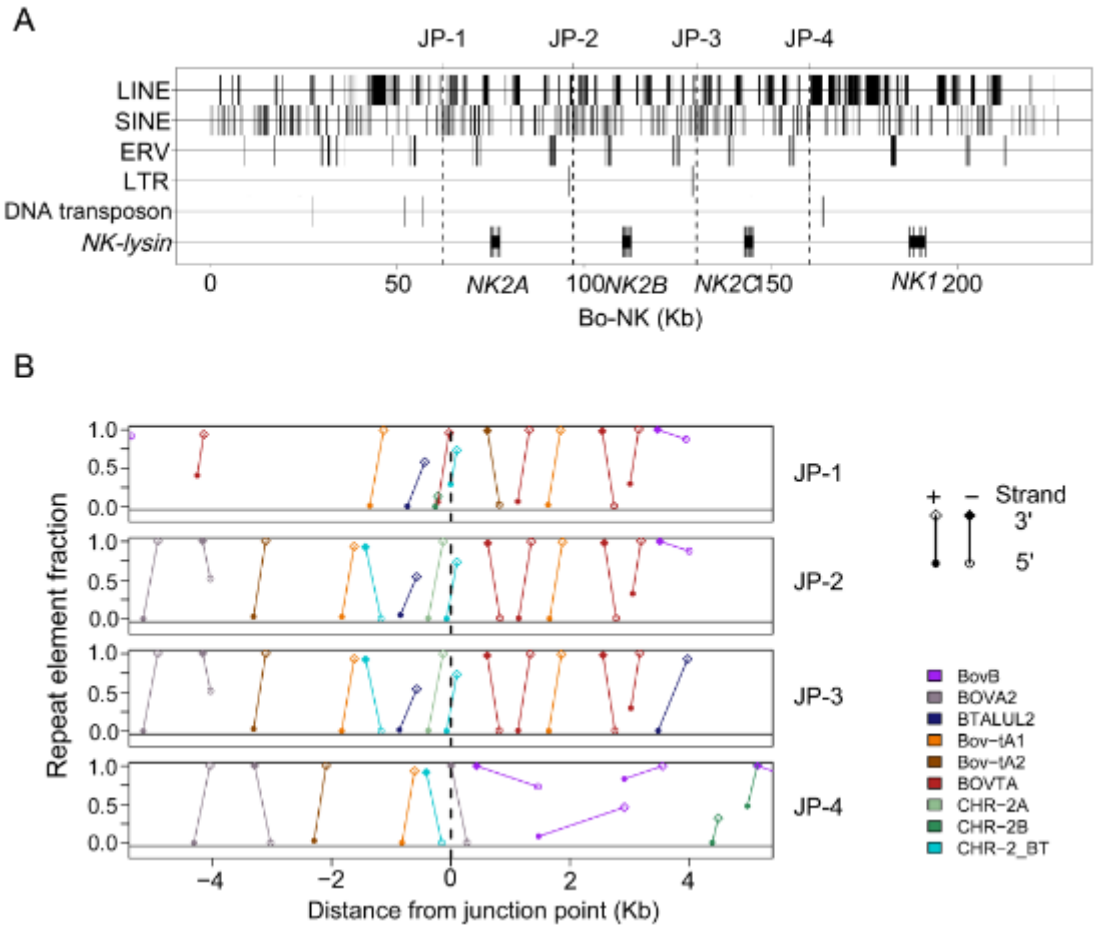


Figure 8. Repeat element analysis within bovine *NK-lysin* region. (A) Distribution of repeat classes within the assembled supercontig Bo-NK. Four junction points and genes are indicated. (B) Distribution of SINE elements within 5 Kb upstream and downstream of each junction point. The portion of each element relative to its consensus sequence is shown on the *y* axis.

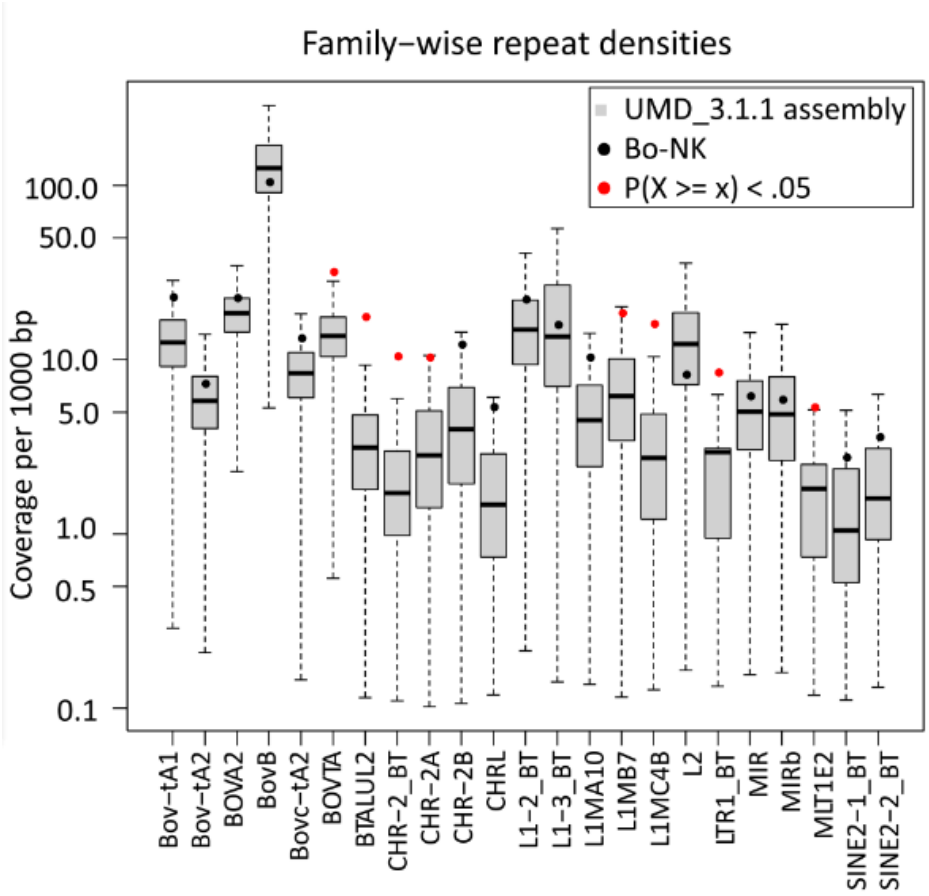


Figure 9. Comparison of the repeat densities between whole genome assembly (UMD_3.1.1) and the assembled Bo-NK supercontig.

Discussion

By sequencing the BAC clones, we provided evidence for tandem duplications of three *NK-lysin* genes, likely derived from an ancestral fourth copy located ~ 41.8 Kb downstream on cattle chromosome 11. Conserved features of *NK-lysin* orthologs exist in all four bovine *NK-lysin* genes, including the presence of five exons/four introns, six well-conserved cysteine residues, and a high proportion of positively charged amino acids. Also, the genome context flanking the bovine *NK-lysin* gene family demonstrated conserved syntenies with the *granulysin* region of human and most other mammalian genomes. The human *granulysin* gene maps to chromosome 2 centromeric to *SFTPB* and *USP39* and telometric to *ATOH8* and *ST3GAL5*. Similarly, the bovine *NK-lysin* gene family maps centromeric to *SFTPB* and *USP39* and telometric to *ATOH8* and *ST3GAL5* on chromosome 11. The conserved genome context implies that no major inter-chromosomal genomic reorganization occurred in this region since the divergence of the ancestors of cattle and humans.

The arrangement of *NK2A, NK2B* and *NK2C* as head to tail in tandem triplicate replicates is consistent with predominate duplication pattern observed in cattle and other mammals including mouse, rat and dog. This is in contrast to the archetypical organization of interspersed duplications in higher primates [131-136]. Segmental duplication (SD) with subsequent differentiation is the major mechanism of gene family expansion. Acting as the substrates of genome evolution, SD regions are also particularly unstable and hotspots of copy number variation (CNV) [88, 134, 135, 137,

138]. Further studies are necessary to test the copy number polymorphism of each *NK-lysin* gene within this family. In order to characterize the mechanism of this gene family expansion, we investigated the features of sequences flanking each breakpoint, and found that the fragments downstream of each breakpoint were highly repetitive. These highly repetitive regions share high sequence homology and potentially drive rearrangements among the genetic elements flanked by these repeats, which can result in deletions or duplications of genomic fragments.

In contrast to the single copy of *NK-lysin* gene in most species including human, pig, chicken and horse, four *NK-lysin* genes are clustered in a region with highly repetitive sequences in the cattle genome. Cattle are the first mammals in which multiple *NK-lysin* genes have been found, and this is consistent with the observation of gene family expansions in cattle for several other genes related to innate host immunity, such as the *defensins*, *cathelicidins* and *interferons* [139-142]. Perhaps reflecting an evolutionary strategy to deal with the substantial number of pathogens and the increased risk of infections in the rumen of cattle, the enlarged gene families encoding the antimicrobial peptides may be selected to meet an increased demand [143]. Further studies are needed to investigate the function of each *NK-lysin* gene.

CHAPTER III

BIOLOGICAL FUNCTION OF BOVINE NK-LYSIN GENES [*]

Introduction

A multi-gene family is a group of genes that arise from a common ancestral gene and share high sequence identity and some conserved features to each other. Three models have been proposed for the evolutionary process of a multi-gene family, "divergent evolution", "concerted evolution" and "birth-and-death evolution" [144]. The "divergent evolution" model proposes that duplicated genes diverge gradually from each other since the duplication event. While the ancestor gene keeps its original function, the duplicates are free from purifying selection and can accumulate mutations and eventually acquire new gene functions. In contrast to the divergence among gene family members from the "divergent evolution" model, members in a gene family could evolve in a concerted manner and share high sequence identity to each other due to the gene conversion events proposed by the "concerted evolution" model. Another model is the "birth-and-death evolution" model, in which some duplicated genes are functional and maintained in the genome for a long time, whereas others are deleted or become pseudo-genes. Although four bovine *NK-lysin* genes are identified in the cattle genome and each has a complete genomic structure with five exons and four introns, their expression pattern and biological functions are unknown.

---

[*] Part of this chapter is reprinted with permission from "Bovine NK-lysin: Copy number variation and functional diversification" by Chen J, et al, 2015. *Proc Natl Acad Sci U S A*, 112(52):E7223-9, Copyright [2015] by the National Academy of Sciences of the United States of America.

Cationic AMP molecules can interact with both anionic phospholipid model membranes and bacterial membranes, presenting an unordered structure in lipid-free state while adopting a more helical structure upon binding to lipid membranes [145]. Circular dichroism (CD) spectroscopy is a valuable tool to study the conformational change of a peptide from lipid-free to lipid-bound states based on the differences in the far-UV (180-260 nm) CD spectrum. CD is the dichroism resulting from the differential absorption of left and right circularly polarized light by an asymmetrical and optically active molecule, including peptides. Secondary structure of a peptide refers to the general three dimensional form of a local segment within the peptide, which is determined by the pattern of hydrogen bonds formed between amine hydrogen and carbonyl oxygen atoms contained in the backbone of peptide bonds. Common secondary structures include alpha helices, beta sheets, beta turn and random coils, and each structure presents a characteristic CD spectrum in the far-UV region. The fraction of each secondary structure in an unknown peptide can be estimated by comparing its CD spectrum to a set of reference molecules whose secondary structures have been previously determined by the high-resolution X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.

Cationic AMPs are proposed to kill targets by forming pores in the membranes. In this study, liposome leakage assay was performed to investigate the disruptive effects of bovine NK-lysin peptides on the model membranes mimicking bacterial membranes, and antimicrobial effects on cell membranes were further confirmed with transmission

electron microscopy (TEM). In the liposome leakage assay, large unilamellar vesicles (LUV) are prepared with entrapped fluorophore/quencher dyes. The fluorescence of the fluorophore is quenched by the quencher when the liposome is free from damage, resulting in the low detected fluorescence intensity. However, the detected fluorescence intensity is increased when the liposome is disrupted due to the leakage of entrapped fluorophore dye, which is free from the quenching influences of the quencher dyes. The detected fluorescence intensity is associated with the disruptive effects of peptides before the liposome is completely disrupted [146].

Materials and methods

Expression profiles

Total RNA was extracted from intestine Peyer's patch (IPP), lung, thymus, spleen and respiratory lymph nodes (RLN) from three mixed breed cattle using the RNeasy Mini kit (Qiagen). RNA was then reverse transcribed into cDNA with a SuperScript® II Reverse Transcriptase kit (Invitrogen). Real-time PCR with Taqman-MGB chemistry was employed to study the expression pattern of each bovine *NK-lysin* gene in the prepared tissues. Specific Taqman-MGB probes for each of the *NK1*, *NK2A* and *NK2C* genes were designed using Primer Express v.2 software (Applied Biosystems) and the corresponding primer pairs were designed within 50 bp upstream and downstream from the probe using Primer3. The specificity of each probe was validated if the probe worked only for the plasmid containing its target gene template in a qPCR test reaction, but not for the plasmid containing other *NK-lysin* templates. As for the *NK2B* gene, no feasible

specific probe was achieved, and therefore a pair of gene-specific PCR primers was designed where 3' end of both forward and reverse primers were placed at the gene specific loci. Specificity of *NK2B* primers was confirmed by sequencing fifteen clones from the cloned qPCR product. Primer Express v.2 was then utilized to search for a feasible Taqman-MGB probe within the *NK2B* amplicon. All primer and probe information is summarized in Table 4. Primers were synthesized by Sigma and Taqman-MGB probes plus the bovine *GAPDH* gene expression master mix were purchased from Applied Biosystems. Quantitative PCR was performed in triplicate reactions with a 20 μL mixture containing equivalent amount of cDNA from each tissue, 2 × Taqman Universal master mix (Applied Biosystems), 0.3 μM of each forward and reverse primer and 100 nM Taqman-MGB probe. The reaction was carried out following the program: 50 ℃ for 2 min, 95 ℃ for 10 min, followed by 95 ℃ for 15 s and 60 ℃ for 1 min for 40 cycles. The mean threshold cycle value (Ct) of each sample was normalized to the internal control, *GAPDH*, and the expression profile for each gene was obtained by comparing its normalized Ct value to the calibrator sample, where the gene exhibited the lowest expression level.

Table 4. Primer and probe information

Primer	Forward (5' to 3')	Reverse (5' to 3')	Probe (5' to 3')
Ex-NK1	CCAGCAAGAATGTCATCATCC	GTCCTTAGAGATGCGATTGAGATAC	CTTTGCAACCAGATGA
Ex-NK2A	AGGAGAAGAGCTGGGCCTAC	GCTGATCTCCCAACTTGTCC	TCCTTGTTGGATGATAATG
Ex-NK2B	GAGAATACCGTCATCGAGGC	TTGCACAGACCTTTCAGCG	TCCAAGGTGTGCGGC
Ex-NK2C	AATTTCTCCGTACCATCGCT	ATGAAACCTACTGGCTTGCTT	AGGACATCGTAGCTGG

Peptide synthesis

Four 30-aa peptides corresponding to the functional region helices 2 and 3 of each *NK-lysin* gene product were synthesized with > 95% purity by Peptide 2.0 Inc (Chantilly, VA). Amino acid compositions and properties of peptides are summarized in Table 5 Lyophilized peptides were dissolved and aliquoted in 10 mM potassium phosphate buffer (pH 7.4) and stored at −20 ℃ before use. Concentrations of the stock peptides were determined by amino acid assay in the Texas A & M University Protein Chemistry Lab.

Table 5. Sequences and properties of four synthetic bovine NK-lysin peptides

Peptide	Sequence	Length (aa)	Charge	Net Charge (pH 7)	Hydrophobicity (pH 6.8)
NK1	VIIHVTSKVCSKMGLWSILCNQMMKKYLNR	30	+6	4.93	37.1
NK2A	TVIEVASKMCSKMRLLKGLCKSITKRFLRR	30	+8	7.82	31.43
NK2B	TVIEAASKVCGKMGPLKGLCKSITKRFLRR	30	+7	6.82	26.1
NK2C	TVIEEASKVCSKMRLLKGLCKSIMKKFLRT	30	+6	5.82	30.57

Circular dichroism (CD) assay

Phospholipids 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (POPG) and 1',3'-bis[1,2-dimyristoyl-*sn*-glycero-3-phospho]-*sn*-glycerol (Cardiolipin) were purchased from Avanti Polar Lipids (U.S.A.). Lyophilized lipids were dissolved in chloroform to a concentration of 20 mg/mL and stored at −20 ℃ before use. To prepare the negatively charged liposome containing 35% POPE, 50% POPG and 15% cardiolipin, the appropriate amounts of the lipid stock solutions were mixed and the chloroform was evaporated under N_2 with constant rotation, so that the dried lipid mixture formed a thin film on the glass wall, which was further dried in a vacuum environment overnight. The dried mixture was re-suspended in potassium phosphate buffer (10 mM, pH 7.4) to the concentration of 10 mM, vortexed for ten mins, bath-sonicated for fifteen mins and subjected to five freeze-thaw cycles. The solution was subsequently extruded through a polycarbonate membrane (100 nm), back and forth, twenty times and stored at 4 ℃

before use. The CD spectrum was obtained in the same phosphate buffer containing 20 μM of each peptide with or without liposome at the working concentration of 1 mM at room temperature with a JASCO J-815 CD Spectrometer (JASCO, Easton, MD). Each sample was scanned five times at wavelengths ranging from 190 to 250 nm with the step resolution of 1 nm. All data were expressed as the mean molar ellipticity ($deg.cm^2.dmol^{-1}$), background (buffer or liposome only) subtracted and the content of each secondary structure including alpha-helix, beta-sheet and beta-turn was estimated with the analysis software provided by the manufacturer of the CD spectrometer using CONTIN with SDP48 as the reference set.

Liposome leakage assay (Fluorescence quenching assay)

Liposome containing 35% POPE, 50% POPG and 15% cardiolipin and the entrapped fluorophore/quencher (ANTS/DPX) dye pair were prepared by a method similar to that described above, except that the potassium phosphate buffer was replaced by a dye-containing Pipes buffer (5 mM ANTS/50 mM DPX/20 mM Pipes/27.5 mM NaCl, pH 7.4) to suspend the dried lipids. The liposome with entrapped ANTS/DPX was subjected to a G-50 Sephadex chromatography column to eliminate the free dye, and the total lipid concentration of the collected dye-free fractions was determined by a phosphorus assay [147]. Dye-free liposome was mixed with or without peptides in a Pipes buffer (20 mM Pipes/ 85 mM NaCl, pH 7.4) to a final lipid concentration of 300 μM and peptides at a serial dilutions of 0.5, 1, 2, 5 and 10 μM. The fluorescence intensity was measured using

a BioTek Synergy 2 microplate reader, with excitation filter 330/80 and emission filter 540/35. Fluorescence intensity was measured before and after the addition of peptides.

Bacterial killing assay

Overnight cultures of gram-positive *S. aureus* (ATCC 25923) and gram-negative *E. coli* (ATCC 25922) grown in lysogeny broth (LB) at 37℃ with aeration were sub-cultured to fresh LB at a ratio of 1:50 and grown at 37℃ with aeration for another 2.5 hours to mid-exponential phase, washed and re-suspended in potassium phosphate buffer (10 mM, pH 7.4) to a concentration of 3×10^6 CFU/mL. An aliquot of 110 μL of prepared bacterial cells was incubated with 10 μL buffer or buffer plus peptides at working concentrations of 0.05, 0.1, 0.5, 1 μM at 37 ℃ for 2 h, and then plated onto LB agar plates. Colonies of the surviving bacteria were manually counted after overnight incubation at 37 ℃. The results were collected from four biological replicates and two independent experiments.

Transmission electron microscopy (TEM)

One hundred and ten μL of the *E.coli* cells (ATCC 25922) (3×10^8 CFU/mL) were incubated with 10 μL buffer or buffer plus 5 μM *NK1* peptide at 37 ℃ for 20 mins. Cells were fixed with equal volume of 2.5 % glutaraldehyde at room temperature for 2 hours, washed and placed in 0.1 M sodium cacodylate buffer. The fixed cells were postfixed in 1% OsO4 with 1% $K_4[Fe(CN)_6]$ for 1 hour at 4 ℃, rinsed with 0.1 M sodium cacodylate buffer, followed by dehydration in an ascending ethanol gradient (50%, 70%, 80%, 90%, 95% and 100%) and embedded in epoxy resin. Ultrathin sections were obtained

with a Leica EM UC6 ultramicrotome and poststained with uranyl acetate and lead citrate, and examined with a Morgagni 268 transmission electron microscope (FEI). Additional image analyses were performed with ImageJ [148]. Statistical analysis of the mean electron intensities of thirty cells from both the control and *NK1*-treated groups was performed with student *t*-test (paired, two-tailed, unequal variances).

Results

Tissue expression of the bovine NK-lysin genes

To test whether all the identified bovine *NK-lysin* genes are expressed and display the same expression profile, we compared the mRNA levels of each gene among five tissues, including lung, thymus, spleen, respiratory lymph node (RLN) and intestine Peyer's patch (IPP). Real time PCR analysis demonstrated that all four bovine *NK-lysin* genes are expressed but each exhibits a tissue specific expression profile (Fig. 10). *NK1* and *NK2A* genes are highly expressed in the IPP, but expressed at extremely low level in the lung. The difference was greater than 100-fold. *NK2B* is more generally expressed with highest levels in the IPP and lung. A distinct expression pattern was observed for *NK2C*, where it was expressed at the highest level in the lung, indicating a potential novel function.

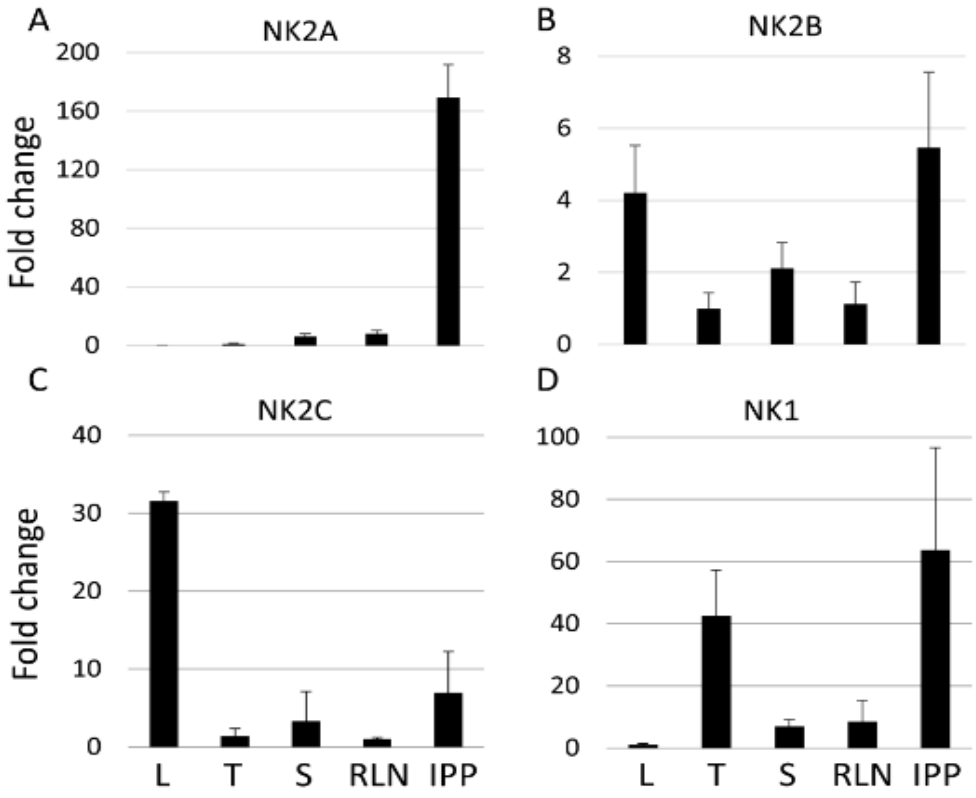


Figure 10. Expression of four bovine *NK-lysin*s in lung (L), thymus (T), spleen (S), respiratory lymph node (RLN) and intestinal Peyer's patch (IPP). The expression of each gene in the tissue that exhibited the lowest expression level was set at 1. The average expression levels and standard deviations were calculated from three healthy individuals.

Conformational changes of bovine NK-lysin peptides upon liposome binding

To determine the interactions between bovine NK-lysin peptides and bio-membranes in target microorganisms, we employed circular dichroism (CD) spectroscopy to study the potential conformational changes of these peptides upon their interaction with anionic liposome mimicking bacterial membranes. The CD spectrum of each of the peptides in buffer presented a single negative band at 200 nm, which indicated an unordered structure (random coil) (Fig. 11A). However, two negative bands at 208 nm and 222 nm along with a positive band at 192 nm were exhibited when mixed with the negatively charged liposome (35% POPE + 50% POPG + 15% Cardiolipin), suggesting the conformational transition of the peptides from random coils to a more ordered structure (Fig. 11B). The proportional contents of the alpha–helix, beta-sheet and beta-turn of each peptide in both lipid-free and lipid-bound states were also compared (Fig. 11C,D). The proportions of the total ordered secondary structures, especially the alpha helices, were enhanced in the presence of liposome for all peptides. The fractions of each secondary structure for *NK2A*, *NK2B* and *NK2C* were comparable upon interaction with liposome, while those for *NK1* were different with a lower degree of helicity and a higher proportion of beta-sheet in the lipid-bound state. This result was consistent with the behavior of most cationic AMPs, which exhibit an unordered structure in aqueous solution but adopt a more helical conformation upon interaction with anionic phospholipid membranes [149].

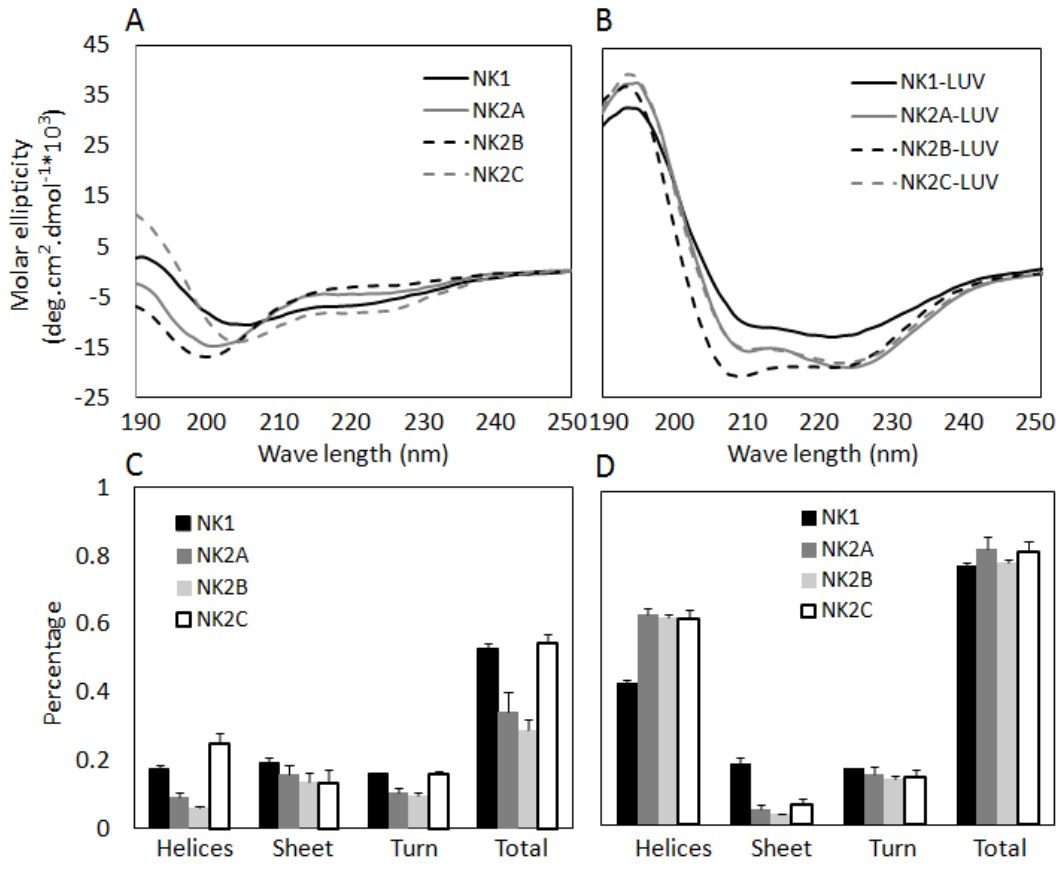


Figure 11. Secondary structural changes of four synthetic bovine NK-lysin peptides upon liposome binding. CD spectra of NK-lysin peptides in lipid-free (A) and lipid-bound states (B) are compared. Estimated secondary structural contents, including alpha-helices, beta-sheet, beta-turn and the total secondary structure in lipid-free and lipid-bound states are shown in (C) and (D), respectively.

Bovine NK-lysin peptides disrupt model membranes

A liposome leakage assay was performed to investigate the influences of the synthetic bovine *NK-lysin* peptides on a model membrane. The peptides began to disrupt the liposome at a concentration of 0.5 μM, resulting in the release of entrapped fluorescence dye ANTS (Fig. 12). As the concentrations were increased to 1 μM and subsequently to 2 μM, the released fluorescence intensities were correspondingly elevated and the leakage of entrapped dye caused by *NK1* peptide was remarkably greater than that caused by the other peptides. However, the detected fluorescence intensity for the four peptides was comparable at the peptide concentration of 5 μM, and was maintained at this level when the concentration was increased to 10 μM, indicating the complete disruption of the vesicles at a peptide concentration of 5 μM.

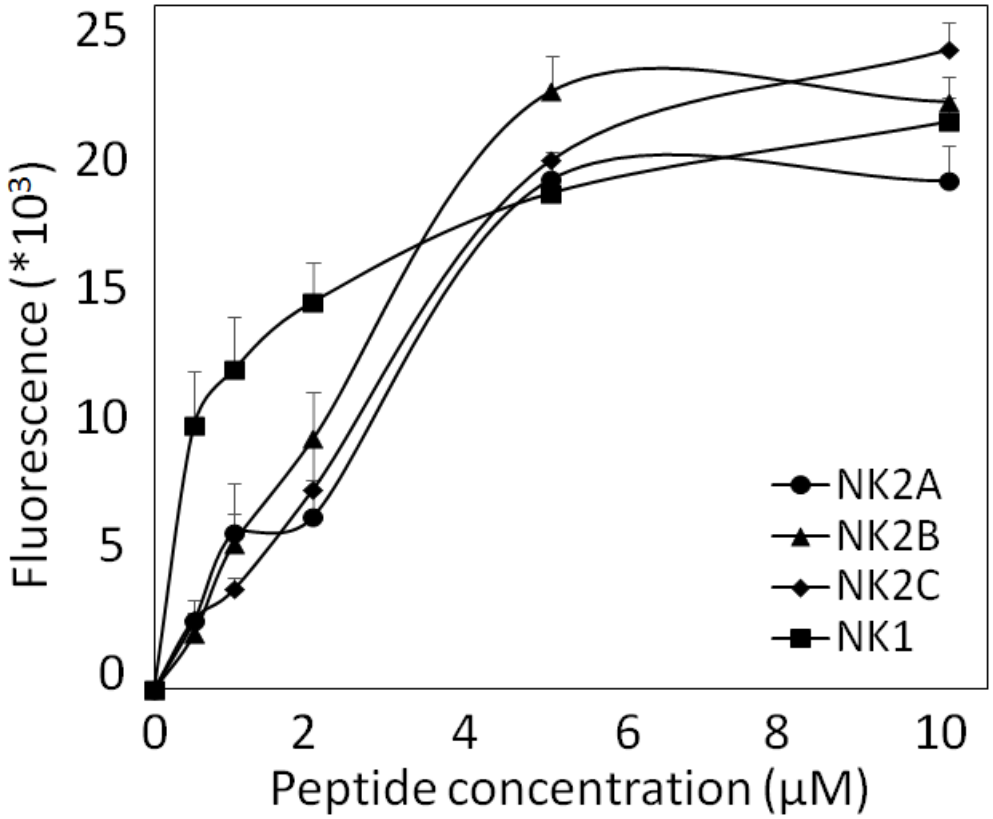


Figure 12. Intensities of the released fluorescent dye (ANTS) from liposome plotted against the concentration of four bovine *NK-lysin* peptides.

Bovine NK-lysin peptides exhibit antimicrobial effects on both Gram-positive and Gram-negative bacteria

Antimicrobial capacities of the synthetic forms of four bovine *NK-lysin* peptides were tested against both the gram-positive bacteria *S. aureus* and the gram-negative bacteria *E. coli*. All peptides were effective against both bacterial strains at nanomolar concentrations, although gram-negative *E. coli* was more susceptible (Fig. 13A). At the lowest peptide concentration of 0.05 μM, an approximately 10-fold decrease in viable *E.coli* cells was observed, and bacterial numbers were reduced from initial 10^6 CFU/mL to less than 10^4 CFU/mL after incubating with 1 μM of *NK2A*, *NK2B* or *NK2C* molecules for 2 hr. Even fewer cells (400 CFU/mL) survived incubation with 1 μM *NK1* peptide. All peptides were less active against the gram-positive *S. aureus* (Fig. 13B). Bacterial numbers were not significantly reduced when incubated at peptide concentrations up to 0.1 μM for any of the four peptides. At 0.5 μM, all peptides produced approximately 10-fold cell loss. At the concentration of 1 μM, the *NK1* molecule reduced *S. aureus* numbers by approximately 100-fold, which was stronger than the other three peptides. Although there was a difference in the ability of the peptides to kill gram-positive and gram-negative bacterial strains, the *NK1* peptide showed strongest antimicrobial effects against both strains.

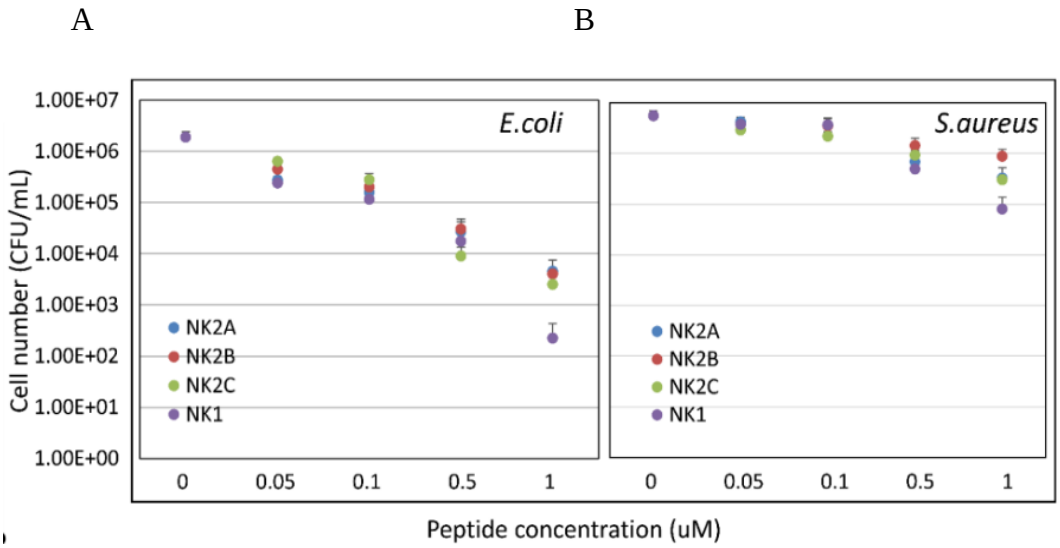


Figure 13. Antimicrobial activities of four bovine NK-lysin peptides against gram-negative *E. coli* (A) and gram-positive *S. aureus* (B). Cell viabilities were analyzed by comparing the surviving cells after peptide treatment with the control. Error bars represented the standard deviations calculated from four biological replications.

Synthetic NK1 peptide disrupted the E. coli membrane

The effects of bovine *NK-lysin* molecules on the *E. coli* cell membrane were investigated by transmission electron microscopy (TEM) (Fig. 14). Specifically, the membrane integrity and intracellular structure were compared and analyzed between untreated *E.coli* cells and cells treated with 5 μM *NK1* peptide. Most of the untreated cells maintained a normal cell shape with an intact cytoplasmic membrane and full cytoplasmic contents (Fig. 14A), whereas treated cells had characteristic expansion of the periplasmic space with shrinkage of the cytoplasmic compartment (Fig. 14B). The cytoplasm of treated cells was less electron dense with clear zones, indicating the disruption of the cell membranes and leakage of intracellular contents. Protruding bubbles were observed from the membrane of treated cells (Fig. 14D) whereas outer membranes of untreated cells displayed a uniform appearance with slightly waved membranes (Fig. 14C). Statistical analysis confirmed that the average electron density of untreated cells was significantly ($P < 0.001$) stronger than the treated ones (Fig. 14E). The results from this assay demonstrated the lytic action of bovine *NK-lysin* peptides, which may directly cause pore formation in the cell membrane.

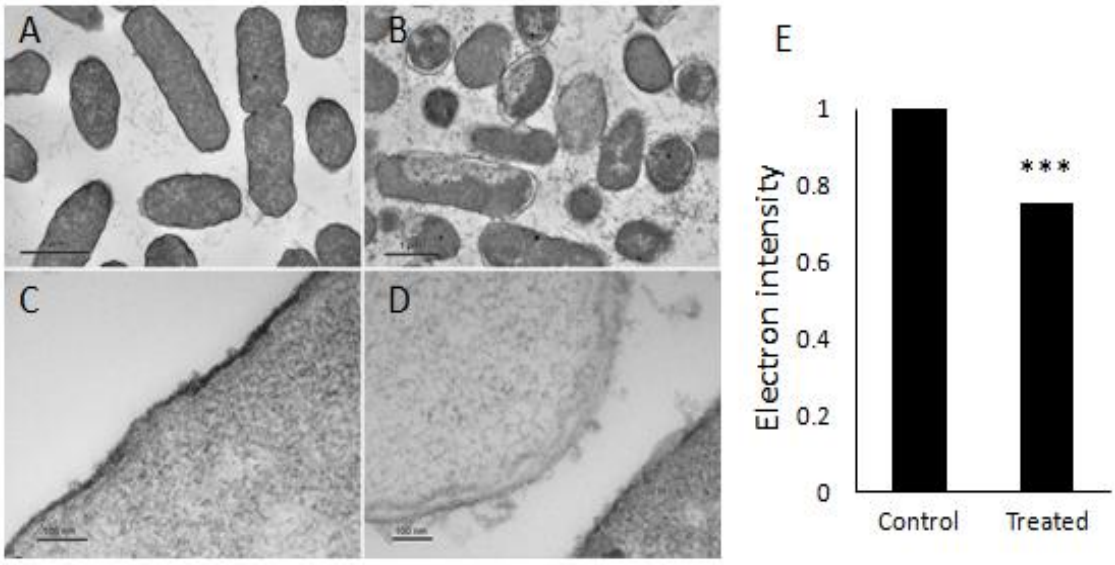


Figure 14. Transmission electron micrographs of *E.coli* cells with and without 5 uM *NK1* peptide treatment. (A)(C) Control cells. (B)(D) Cells treated with 5 uM *NK1* peptide for 20 mins. (E) Comparison of the average electron intensities of thirty cells between the control and *NK1*-treated cell groups.

Discussion

During the evolution of a gene family, it is common for the ancestor gene copy to maintain its original biological function while the duplicates can accumulate mutations and potentially evolve into genes with novel functions. For example, duplicates of bovine *lysozyme* genes, an immunity-related gene family, exhibit non-immune functions in the digestive systems [143]. In this study, gene expression profiles showed that three bovine *NK-lysin* genes *NK1*, *NK2A* and *NK2B* are predominately expressed in the intestine Peyer's patch which is consistent with the *NK-lysin* orthologs in most other species, while the bovine *NK2C* gene is expressed at the highest level in lung, indicating its potential novel function in the bovine respiratory system. It will be interesting to measure the expression of each bovine *NK-lysin* gene in more tissue types to investigate their potential novel functions, including rumen, stomach, liver, heart, brain, muscle and kidney.

Cationic AMPs are important molecules in the innate immune system and are widespread in both plant and animal kingdoms [14]. One of the conserved characteristics of AMPs is their cationic and hydrophobic composition, which enables them to be potent killers of microbes with cytoplasmic membranes rich in anionic phospholipids and selectively safe to host cells with neutral charged membranes. Several factors are proposed to affect the antimicrobial capacities of AMPs, including the net positive charge, hydrophobicity and amphipathicity. Increased positive charge and hydrophobicity are two major contributors to the enhancement of the antimicrobial effect

of an AMP molecule [150, 151]. The net charges (pH = 7) and hydrophobicities (pH = 6.8) differ among the functional regions of the four examined NK-lysin peptides, with NK1 possessing the highest hydrophobicity with the least positive charge and NK2A being the most positively charged peptide (Table 5). During gene family expansion, each paralog has evolved to encode a peptide with a specific amino acid composition, which might enable the bovine *NK-lysin* family to be active against a broad range of microbes. Although NK1 peptide showed the strongest killing ability against both *E. coli* and *S. aureus* cells, other NK-lysin peptides may exhibit stronger effects on different targets. Further studies are therefore suggested to compare the antimicrobial activities of each NK-lysin peptide against more bacterial strains.

CHAPTER IV

RESPONSE OF BOVINE NK-LYSINS TO BRD-ASSOCIATED PATHOGENS

Introduction

Bovine respiratory disease (BRD) or shipping fever is the most common infectious disease affecting both the upper and lower respiratory tracts of cattle and is a major cause of economic loss in North America through treatment costs, reduced performance and mortality [152-154]. BRD is multi-factorial with a variety of stressors, including host factors (age, genetics and host immunity) [155-157], environmental factors (temperature, transport, commingling and ventilation) [158-160] and pathogens (bacteria and viruses) leading to disease. Several microorganisms have been implicated in the pathogenesis of BRD including bacterial agents, such as *Mannheimia haemolytica* [161, 162], *Pasteurella multocida* [161], *Mycoplasma bovis* [163] and *Histophilus somni* [164], and viral agents, such as *bovine viral diarrhea virus* (BVDV) [163], *bovine respiratory syncytial virus* (BRSV) [165], *bovine herpesvirus-1* (BHV-1 or IBR) [165] and *bovine parainfluenza-3 virus* (PI-3) [166]. Interactions between environmental stressors and infectious agents are critical to the development of BRD. Environmental factors (such as transport or weaning) weaken the host's immune system and predispose animals to viral infections, which then precipitate secondary infections by bacterial pathogen, leading to the onset of BRD. Many efforts have been made to prevent and treat BRD, including feedlot management to reduce environmental stresses, vaccination

of animals to improve immune responses, breeding of cattle that are resistant to BRD pathogens [167] and anti-infectious agents (antibiotics and sulfas) to treat infected cattle.

RNA sequencing (RNA-seq), also called whole transcriptome shotgun sequencing (WTSS), is a powerful technology in which RNA is reverse transcribed into cDNA, which is then sequenced by the next-generation sequencing technology. RNA-seq allows scientists to profile the transcriptome of any tissues at any developmental stage. Since its inception, RNA-seq has been under active development and applied to many studies related to gene expression, mainly due to its advantages over other existing technologies such as microarrays [168]. The first advantage of RNA-seq is its unbiased detection of novel transcripts. Unlike the microarray technology which limits researchers to transcripts that correspond to existing genomic sequences, RNA-seq does not require transcript-specific probes and therefore can explore novel transcripts and enable transcriptome profiling for non-model organisms. Variations such as SNPs, INDELs and other previously unknown changes can also be detected by RNA-seq through transcriptome comparisons among different individuals. The second advantage of RNA-seq over microarrays is its low background signal, because sequencing reads can be unambiguously mapped to unique regions of the genome. Thirdly, RNA-seq can quantify a broad range of expression levels by analyzing the number of sequencing reads, which makes it feasible to detect low-abundance transcripts. RNA-seq can also deliver increased specificity and sensitivity in quantifying gene expression with strict parameters set for the analysis when compared to microarrays. Since four bovine *NK*-

lysin genes share high sequence identity to each other, RNA-seq analysis becomes a valuable method to compare the expression of each *NK-lysin* gene among healthy animals and animals challenged with pathogens associated with bovine respiratory diseases.

Materials and methods

RNA-seq data analysis

RNA-seq data were generated and analyzed at the University of Missouri. Computations were performed on the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC). Animal challenge and whole transcriptome sequencing protocols were previously described [169] [170]. In this study, we analyzed the bovine *NK-lysin* expression in both the lung lesion and bronchial lymph node tissues [169, 170] collected from the same individual. Since the four bovine *NK-lysin*s share high sequence identity, especially *NK2A*, *NK2B* and *NK2C*, protocols were designed with extra care to remap the short (2 x 50 bp) reads specifically to each gene. Basically, all short reads from each sample were mapped with no allowed mismatches to a bowtie index built with the mRNA sequences of all four *NK-lysin*s using Bowtie 2 [171]. The mapping quality which measures the degree of confidence in mapping a read to a specific single locus was used to assess whether the reads were uniquely mapped to one of the four genes, and the number of these uniquely mapped reads was counted for each *NK-lysin* gene. Quality trimmed reads with a size of < 25 bp were excluded from this analysis.

Antimicrobial killing assay

Overnight cultures of four pathogenic bacterial strains (*P. multocida* ATCC 43019, ATCC 43137 and *M. haemolytica* ATCC BAA-410, ATCC 33396) were sub-cultured in brain-heart infusion (BHI) medium at 37 ℃ for an additional 2.5 hours to mid-exponential phase, washed and re-suspended in PBS (pH 7.4) to a cell concentration of 5 $\times 10^6$ CFU/mL. A 100-µL aliquot of cells was incubated with 20 µL PBS buffer or buffer plus each *NK-lysin* peptide prepared in the same buffer to the final peptide working concentrations of 1, 2, 5 and 10 µM at 37 ℃ for 1 h. After the 1 h incubation, a 100-µL aliquot of each mixture was diluted in PBS buffer to an approximate cell concentration of 3 $\times 10^3$ CFU/mL, from which another 100-µL aliquot was plated on Trypticase soy agar plates with 5% sheep blood. Colonies of the surviving cells were manually counted after overnight incubation at 37 ℃. Experiments were performed with four biological replicates.

Transmission electron microscopy (TEM)

50-µL overnight culture of *Pasteurella multocida* ATCC 43019 was sub-cultured in 5 mL BHI medium for 2 h. Four mL of the cell culture was subsequently washed and re-suspended into PBS buffer, and incubated with 20 µM NK1 peptide or an equal volume of PBS buffer for 30 mins at 37 ℃. The mixture was fixed with an equal volume of 3% glutaraldehyde and samples for TEM examination were prepared following the previously described protocol in CHAPTER III.

Results

Elevated expression of bovine NK2C in animals challenged with BRD pathogens

In this study, we investigated the potential contributions of bovine *NK-lysin*s to host resistance to BRD pathogens, especially the *NK2C* gene which is expressed in lung at a high level. The expression level of each *NK-lysin* gene was represented as the Fragments per Kilobase of transcript per Million (FPKM) value and compared in bronchial lymph node and lung lesion tissues among healthy animals and animals challenged with a set of BRD-causing pathogens. Overall, the expression of *NK1* gene was very low in these tissues while *NK2C* exhibited relatively high expression in both tissues (Fig. 15). When animals were challenged with the IBR virus, the expression of *NK2A*, *NK2B* and *NK2C* was significantly elevated in bronchial lymph nodes, and an increased expression of *NK2B* and *NK2C* in bronchial lymph nodes was also observed in most of the animals challenged with other pathogens (Fig. 15A). An elevated expression of *NK2A* and *NK2C* in lung was also observed in most of the pathogen-challenged animals, and the expression of *NK2C* was significantly higher than for the other three genes in the lungs (Fig. 15B). In contrast to the comparable gene expression levels in the four healthy control animals, the expression of *NK2C* showed large variation among animals within the same challenged group and the expression was elevated by > 20-fold in some of the experimentally challenged animals. All of these results suggest that *NK2C* can potentially play a significant role in host defense against specific infectious agents involved in BRD.

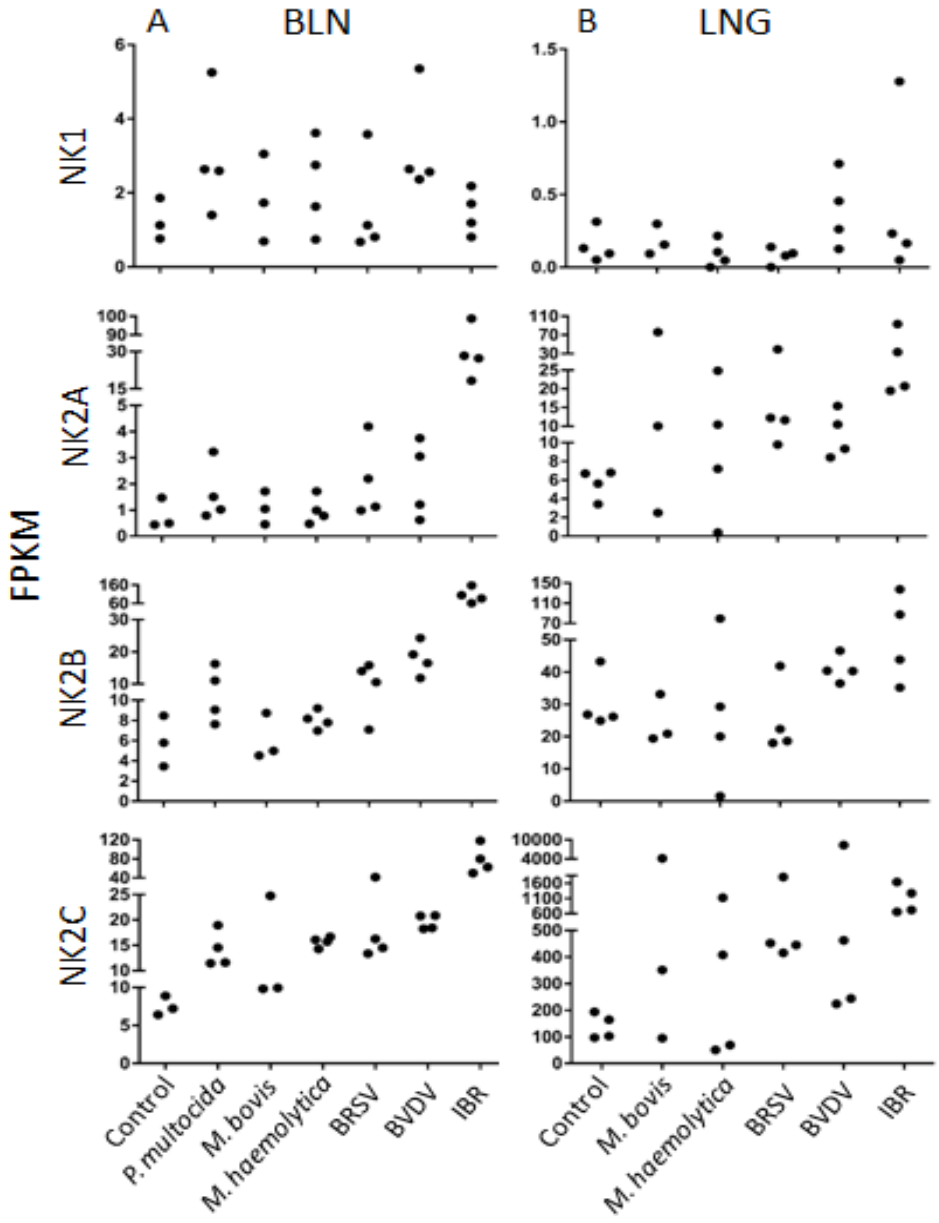


Figure 15. Comparisons of the expressions of four bovine *NK-lysin* genes in bronchial lymph node (BLN, A) and lung (LNG, B) among healthy animals and animals challenged with *P.multocida*, *M.bovis*, *M.haemolytica*, BRSV, BVDV and IBR. The Y axis shows the FPKM value, and each black dot represents the FPKM value of an individual. Three or four individuals were included in each control and challenged group.

Bovine NK-lysin peptides exhibit antimicrobial effects on BRD-causing bacteria P. multocida and M. haemolytica

The antimicrobial activities of bovine NK-lysin peptides were tested against two *P. multocida* bacterial strains (ATCC 43019 and ATCC 43137) and two *M. haemolytica* bacterial strains (ATCC BAA-410 and ATCC 33396). Overall, the *P. multocida* strains were less susceptible to the peptides than the *M. haemolytica* strains (Fig. 16 A, B). Significant cell number losses were not observed until the peptide concentration was increased to 10 μM for *NK1* and *NK2A* when an approximately 50-fold decrease in viable cells was produced. The *NK2B* and *NK2C* peptides did not display obvious killing abilities against both *P. multocida* strains. In contrast, the *NK2A* and *NK2C* peptides displayed potent antimicrobial activities against both *M. haemolytica* strains in a dose-dependent manner (Fig. 16 C, D). An approximately 5-fold decrease in cell numbers resulted from incubation with 1 μM of *NK2A* for 1 h, and the complete elimination of *M. haemolytica* cells was achieved with 5 μM of *NK2A* or 10 μM of *NK2C*. *NK1* and *NK2B* peptides exhibited weaker killing abilities against *M. haemolytica* and achieved an approximately 50-fold cell loss at the highest concentration of 10 μM. Surprisingly, *M. haemolytica* cells were very susceptible to the *NK2A* peptide but resistant to the *NK1*, which was the most potent peptide against *P. multocida* as well as *E. coli* and *S. aureus* in our previous study.

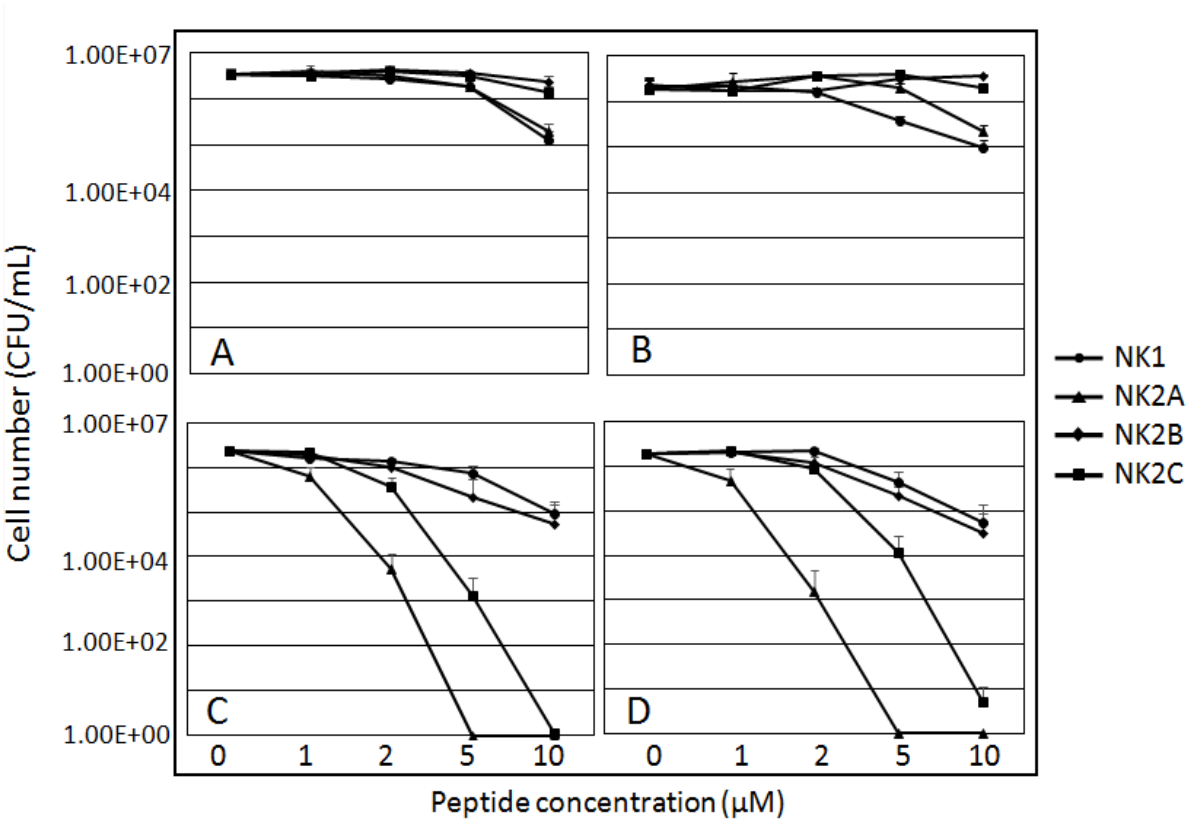


Figure 16. Antimicrobial effects of bovine NK-lysin peptides on BRD-causing pathogens *P. multocida* strains ATCC 43019 (A) and ATCC 43137 (B), *M. haemolytica* ATCC BAA-410 (C) and ATCC 33396 (D). Surviving cell numbers after peptide treatment are shown on the Y axis. Error bars represent the standard deviations calculated from four biological replications.

Bovine NK1 peptide lyses Pasteurella multocida cell membranes

The impacts of bovine NK-lysin peptides on the cell morphology and membrane integrity of *Pasteurella multocida* cells were examined by transmission electron microscopy (TEM) (Fig. 17). The untreated cells displayed intact outer and inner membranes with a clear periplasmic space, and the cytoplasm was homogeneously filled with electron dense material (Fig. 17A). Although the cell morphology was maintained, severe cellular damage with large clear zones in the cytoplasm indicating the leakage of cytoplasmic contents was observed when cells were treated with 20 μM of NK1 peptide for 30 mins (Fig. 17B). In addition, cytoplasmic constituents were coagulated into non-membrane-enclosed bodies within the areas near membranes. NK1 peptide treatment also caused the breakage of cell membranes (Fig. 17C arrows a & b) and the formation of protruded bodies in the cell membrane, resulting in the leakage of cytoplasm (Fig. 17C arrows c & d). Statistical analysis revealed that the overall electron density of an untreated *P. multocida* cell was significantly higher than that of a cell treated with bovine NK1 peptide for 30 mins, suggesting the leakage of cytoplasmic contents in NK1-treated cells (Fig. 17D). Therefore, bovine NK1 peptide can cause the release of cytoplasmic material from a *P. multocida* cell by damaging its cell membrane, which will eventually lead to cell death and the appearance of empty "shells" (ghost cells).

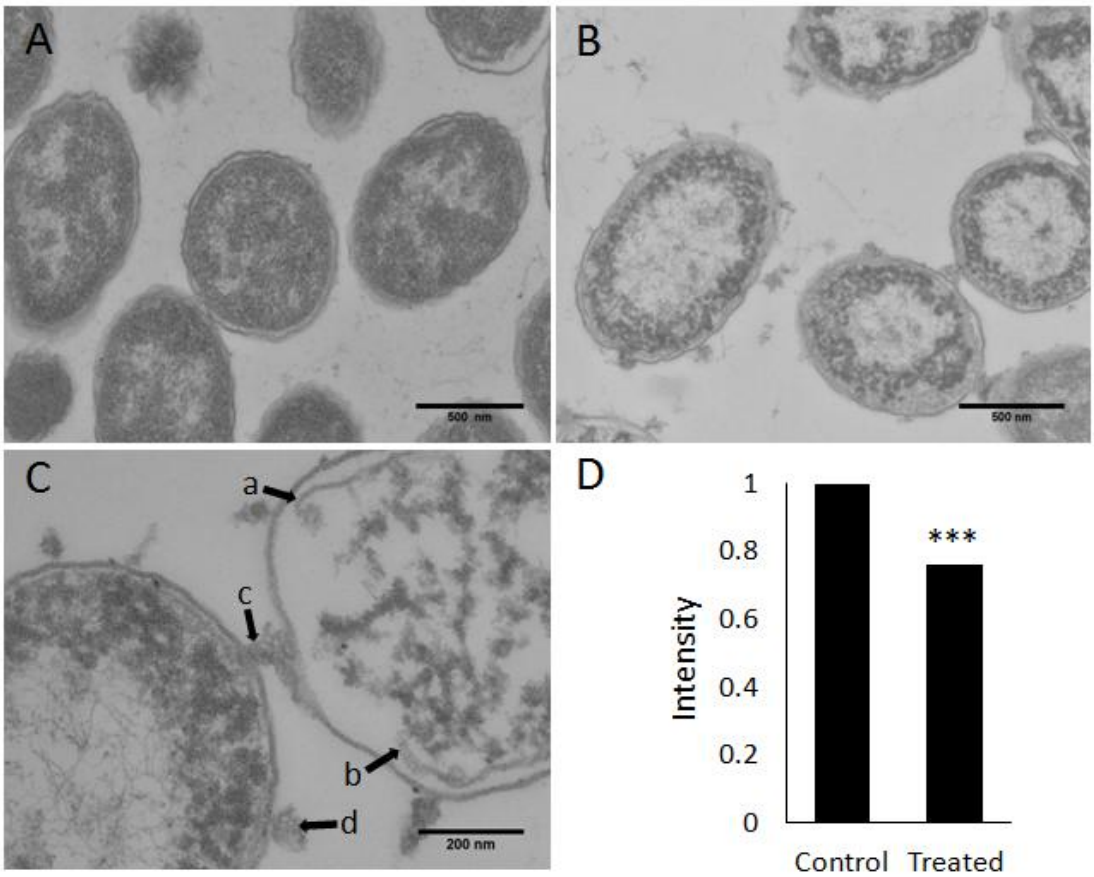


Figure 17. Influence of 20 µM of bovine NK1 peptide on the cell membrane of *Pasteurella multocida* (ATCC 43019) examined by transmission electron microscopy. (A) Control cells. (B) and (C) Cells treated with 20 µM *NK1* peptide for 30 mins. (D) Statistical analysis of the average electron intensity of control cells versus *NK1*-treated cells. Thirty cells from each group were used for statistical analysis.

Discussion

As discussed in the CHAPTER III, the synthetic NK1 peptide possesses the highest hydrophobicity and largest hydrophobic face with the least positive charge and NK2A is the most positively charged peptide (Fig. 18), which might enable the bovine NK-lysin family to be active against a broad range of microbes. Bacterial killing results revealed that NK1 exhibited the strongest antimicrobial effects on *E. coli*, *S. aureus* and *P. multocida* cells while NK2A was the most potent peptide against *M. haemolytica*, further demonstrating that each NK-lysin paralog has evolved to encode a peptide with a specific property and thus specific targets. The whole NK-lysin gene family cooperate together to enhance the host resistance to a wide range of infectious microbes.

(A)

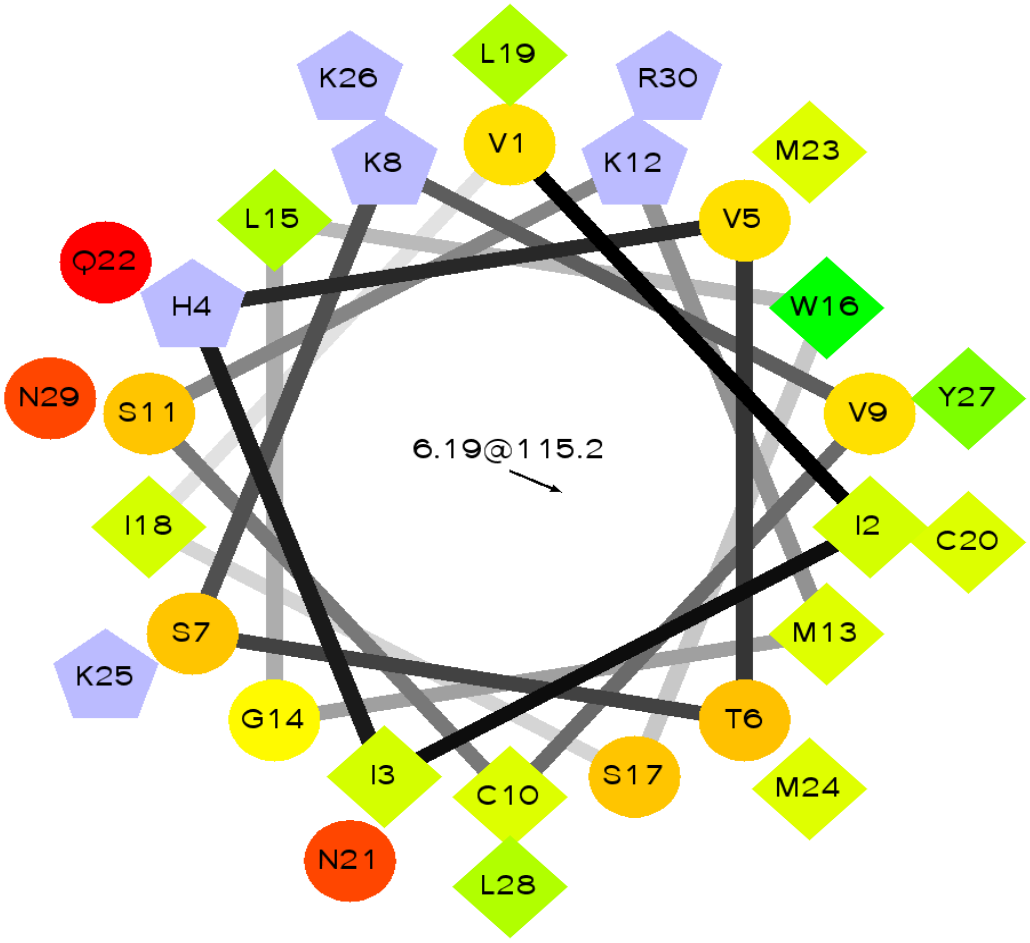


Figure 18. Helical wheel of four synthetic bovine NK-lysin peptides: (A) NK1, (B) NK2A, (C) NK2B and (D) NK2C. Hydrophilic residues (circles), hydrophobic residues (diamonds), negatively charged residues (triangles) and positively charged residues (pentagons) are indicated. Color indicates the hydrophobicity of a residue, in which green represents the most hydrophobic residue, and the amount of green is decreasing proportionally to the hydrophobicity with yellow representing zero hydrophobicity. Hydrophilic residues are coded red with pure red being the most hydrophilic residue, and the amount of red decreasing proportionally to the hydrophilicity. Light blue indicates the potentially charged amino acids.

Figure 18 Continued

(B)

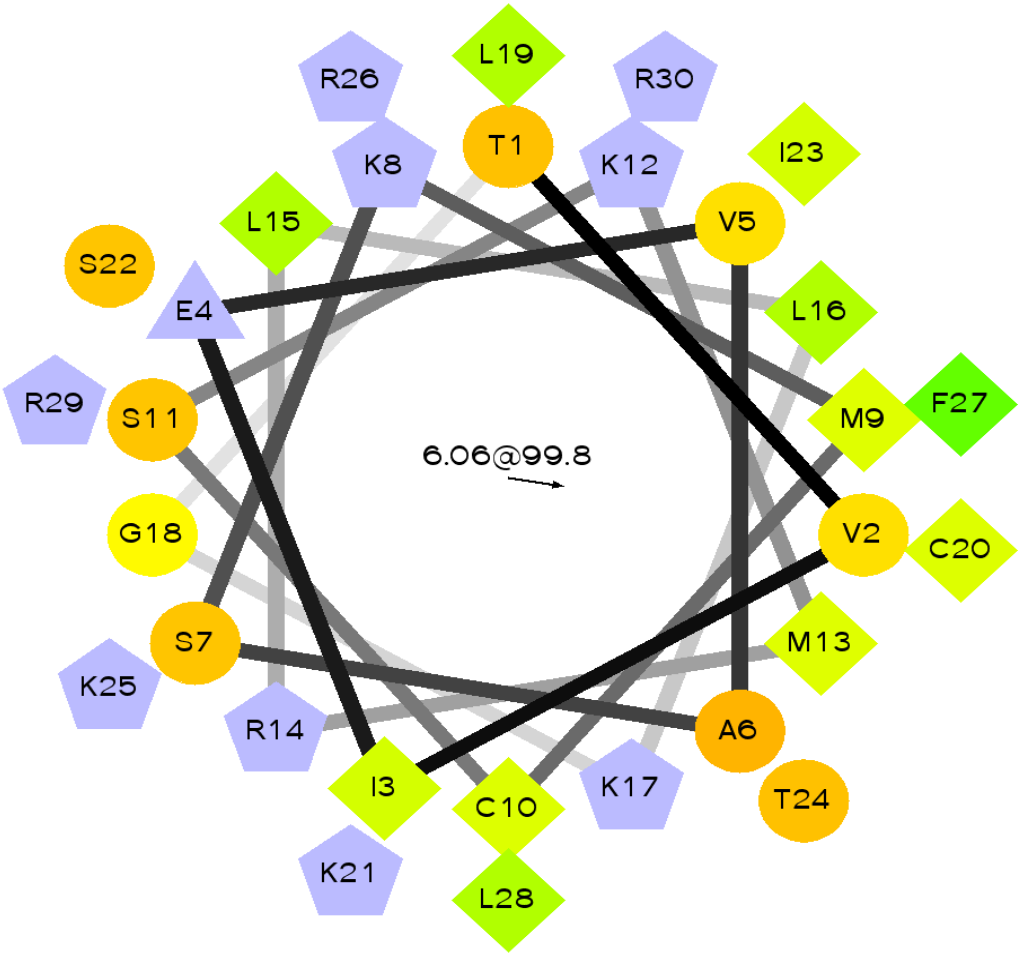


Figure 18 Continued

(C)

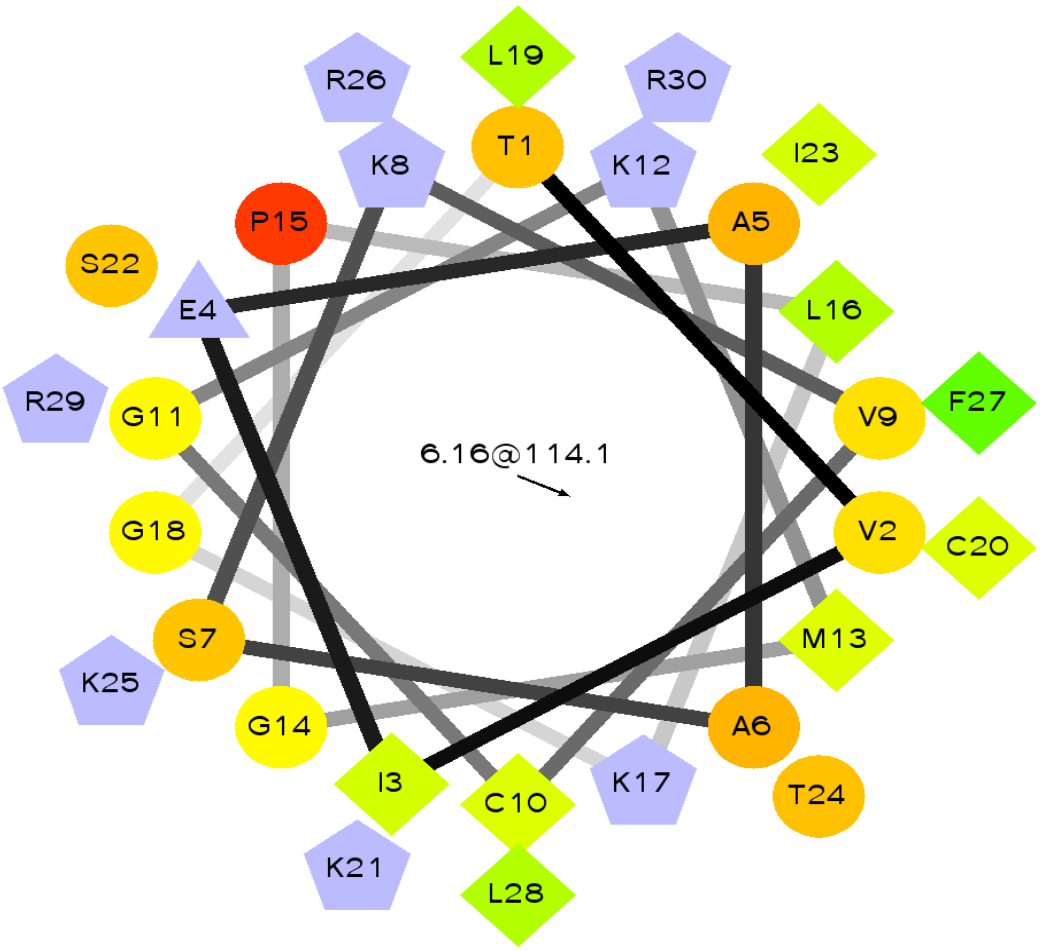
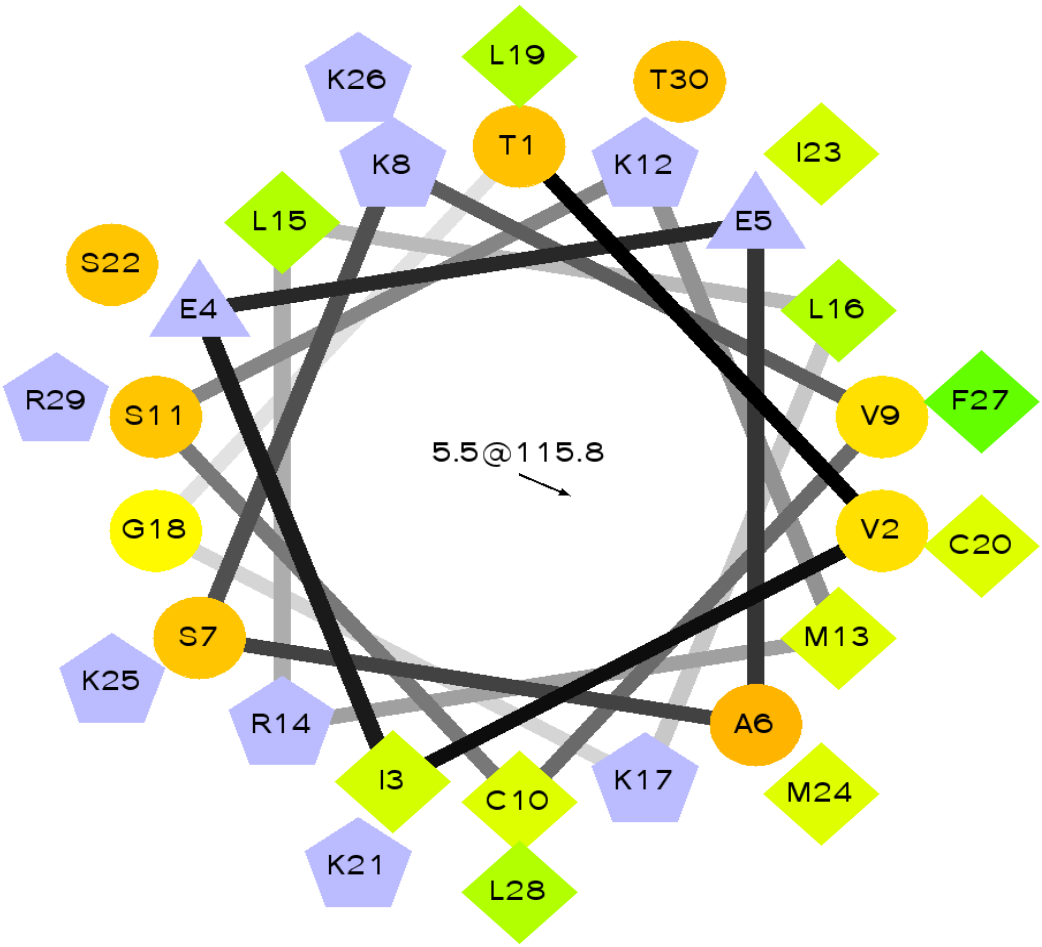


Figure 18 Continued

(D)



Identification of genes that influence the host response to BRD pathogens is an important step towards identifying the specific genetic variants which could be used in breeding cattle with an increased resistance to infections. Several studies have been undertaken to search for genes and associated genetic variants that contribute to host resistance to BRD pathogens or responses to vaccines, and the proposed genes or genomic regions include the *MHC* region, *TLRs*, *PVRL1* and *DST* [167, 172]. With the application of high density SNP genotyping technology, genome wide association studies become a valuable method for identifying genetic markers linked to phenotypic variation in BRD signs [167, 173, 174]. Another effective approach to the identification of genetic variants that could be beneficial to animal breeding relies on the identification of polymorphisms within suggested candidate genes based on the biological functions of their gene products. Since innate immunity is not only an essential component of the host immune response but also affects subsequent acquired immunity, genes that are expressed in the innate immune system are strong candidates for their effects on host resistance to infectious agents. Human granulysin/NK-lysin is an effector molecule in the innate immune system, and its expression is inducible by antigenic stimulation indicating its potential role in host responses to antigens [64]. Despite the existence of large individual variation in expression within individuals challenged with the same BRD-associated pathogen, the expression of bovine *NK-lysin* genes, especially *NK2C*, in both the bronchial lymph node and lung were elevated in most of the challenged animals. The synthetic peptides corresponding to the functional helices 2 and 3 of each gene product exhibited antimicrobial effects on the BRD-associated bacterial microbes,

P. multocida and *M. haemolytica*, and antimycobacterial activity has also been previously reported with some other derived bovine NK-lysin peptides [83]. All of these findings suggest that the bovine *NK-lysins* are potentially important in host resistance to BRD infections.

The large animal-to-animal variation within animals challenged with the same pathogen in the challenge study may be attributed to individual immunity, which at least partly results from genetic variation, such as gene copy number variations (CNVs), SNPs and INDELs. It will be important to investigate genetic variations within members of the bovine *NK-lysin* gene family and their regulatory regions to identify potential associations with host disease phenotypes. Further studies are therefore suggested to investigate the extent of copy number variation within and between breeds of cattle in all four bovine *NK-lysins* as well as the extent of nonsynonymous substitutions, especially in the region coding for the functional helices 2 and 3. Genotype-to-phenotype association studies could then be performed to test the effects of the identified genetic variants on host resistance to infectious agents, providing functional genetic makers for use in cattle breeding.

CHAPTER V

GENETIC VARIATION WITHIN THE BOVINE NK-LYSIN REGION

Introduction

A large number of proteins are synthesized as precursors in the form of pre-pro-proteins. The "pre-" prefix indicates a signal peptide at the N-terminal region of a precursor that transports the precursor into the appropriate secretory pathway and then is cleaved from the precursor. Proteins with signal peptides are defined as pre-proteins. However, some precursors require additional processing to become fully active, including the cleavage of another peptide termed as the pro-region. Pro-regions can be N-terminal extensions, C-terminal extensions or a combination of both, but N-terminal extensions are the most common. Precursors with both signal peptides and pro-regions are defined as the pre-pro-proteins. Pro-regions exist as a part of many pro-proteins, including α-lytic protease [175], subtilisin [176], carboxypeptidase Y (CPY) [177] and alkaline extracellular protease [178], and is required for the folding and maturation of the protein. Protein folding is considered as a competition process between the forward folding reaction and the reversed aggregation reaction, and the pro-region can facilitate the protein folding by either increasing the rate of forward reaction or decreasing the rate of aggregation reaction. Pro-regions are also involved in the intracellular protein transporting. Pro-region of the CPY protease, for example, has been shown to be responsible for targeting the protein to the yeast vacuole [179]. In addition, pro-region can be a potent inhibitor of its associated mature protein and regulate the protein activities in the host. For example,

pro-CPY has less than 0.1% of the biological activity of its mature CPY [179]. It is possible that pro-regions play other unidentified important roles, and variation within this region could have significant influences on protein functions. Pig NK-lysin precursor (UniProtKB: Q29075) consists of 129 amino acids, where the first six residues comprise the signal peptide and residues 47 – 124 are processed to be the mature peptide while the rest of the residues (7 – 46 and 125 – 129) form the N- and C-terminal pro-regions. Sequences of four bovine NK-lysin peptides were aligned with the pig NK-lysin to predict the peptide structure of the bovine NK-lysins (Fig. 19). The first twenty-two amino acids of bovine NK1 were predicted to comprise the signal peptide and residues 63 – 140 formed the mature peptide while the pro-regions were made up of 23 – 62 and 141 – 145 residues.

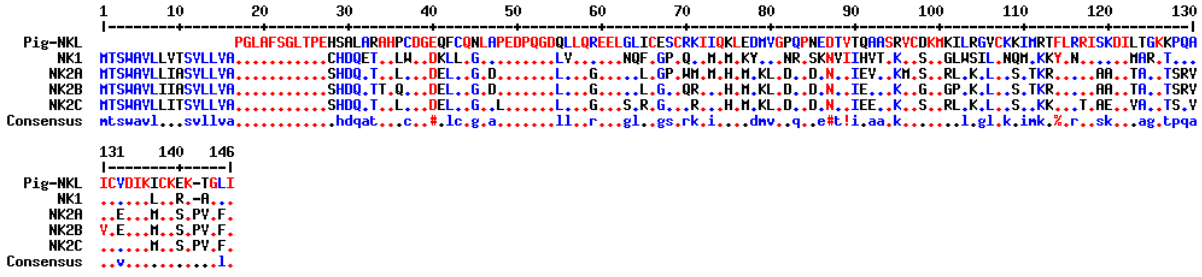


Figure 19. Sequence comparisons of four bovine NK-lysin peptides (NK1, NK2A – C) with pig NK-lysin (Pig-NKL). The four bovine NK-lysin peptide sequences were predicted from the genomic sequence of L1 Domino 99375, donor for the CHORI-240 Bovine BAC Library and PacBio sequencing. Red color indicates alignment with high consensus and blue represents alignment with low consensus.

Concerted evolution is a common pattern during the development of multigene families and explains the higher sequence identity shared by the paralogous genes within one species when compared to the ortholog in another species, even though the gene duplication event preceded the speciation event. One of the main proposed mechanisms responsible for the concerted evolution is gene conversion, which is involved in the genetic exchange, especially the unidirectional transfer of a fragment of sequence from a "donor" gene to the homologous "acceptor" gene, therefore converting the "acceptor" sequence [180]. The size of a gene fragment involved in a gene conversion event is variable, with some entire genes undergoing gene conversions and others displaying a mosaic evolution pattern. Under the mosaic evolution, some segments of the duplicated genes are homogenized by gene conversions while the rest of the gene sequences diverge with time, which usually complicates phylogenetic reconstructions and underestimates the evolution time [181-183]. There are four main methodologies to detect a gene conversion event [184], including the incompatibility between an estimated gene tree and the true duplication history, incompatibility of gene trees in different subregions, traces in the sequence alignment after gene conversion events and shared polymorphisms. Parallel mutations can create shared polymorphisms, but the probability that point mutations occur at the same locations of different duplicate genes is very low. Therefore, examination of shared polymorphisms is a powerful method to detect gene conversion events at the multisite variations (MSVs) [185-187]. MSV is a type of variation that refers to the polymorphism in paralogous genes within a gene family [188], including the MSV 1 and MSV 2, and is usually used to detect gene conversion sites. A SNP in

one of the duplicate genes and also a paralogous sequence variant is defined as a MSV 1, and a MSV 2 is a SNP shared by several duplicated genes [189]. A minimum gene conversion tract is a segment within an acceptor gene that spans at least two SNPs for which a potential donor gene could be identified. A maximum gene conversion tract is a segment that includes the minimum gene conversion tract plus the identical sequence between the donor and acceptor genes on both sides of the minimum gene conversion tract.

Materials and methods

Deletion of a 9-bp fragment in the third exon of NK1

Two annotated mRNA sequences of bovine *NK1* from the National Center for Biotechnology Information (NCBI) nucleotide database were aligned with the *NK1* mRNA sequence predicted from PacBio sequencing result using the Multiple sequence alignment by Florence Corpet (http://multalin.toulouse.inra.fr/multalin/). Total RNA was extracted from intestine Peyer's patch (IPP) of four mixed breed cattle using the RNeasy Mini kit (Qiagen). RNA was then reverse transcribed into cDNA with a SuperScript® II Reverse Transcriptase kit (Invitrogen). A pair of primers (NK1-CDS) was designed to amplify the whole coding region of the *NK1* gene using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/) (Table 6). The prepared IPP cDNA from four different individuals were amplified by the NK1-CDS primers and the amplicons were sequenced at the DNA Technologies Core Lab (Texas A&M University). To identify the genomic location of the 9-bp INDEL, another pair of primers (NK1-9bp) spanning the

exon 2 – intron 2 – exon 3 region of *NK1* were tested in the genomic DNA of nine individuals from different breeds (Table 7), and the sequences of amplicons were achieved by Sanger sequencing and analyzed with Multiple sequence alignment by Florence Corpet.

Copy number variation of NK2B in Holstein cattle homozygous for the NK-lysin region

Primer 3 was utilized to design a pair of primers (Bo-lysin) within the conserved region of *NK2A*, *NK2B* and *NK2C* genes, and PLINK was performed to identify individuals homozygous for all SNP sites across the entire *NK-lysin* region based on genotyping with the bovine 770K HD SNP array [167]. Genomic DNA from four homozygous Holstein cattle with different haplotypes was prepared for further analysis. Briefly, the Bo-lysin amplicons from each of the four selected cattle were cloned into the pCR™4 Blunt-TOPO® vector (*Life technologies*), and the clones were prepared for sequencing at Beckman Coulter Genomics (Danvers, MA, U.S.A). Only sequences present at least three times among the clones from a single individual were used for analysis. All sequences were analyzed phylogenetically with the corresponding reference sequences of *NK1* and *NK2A-C* by MEGA 6.0 [190], pig and horse *NK-lysin* sequences were included as outgroups. The absence of *NK2B* in individuals 2822 and 3850 were further confirmed by PCR with *NK2B* specific primers (Gs-NK2B).The same Gs-NK2B primers were also tested in individuals with different genotypes at the SNP (BovineHD1100014441, T/C) to determine the linkage of *NK2B* with this SNP.

Analysis of gene conversions within bovine NK-lysin gene family

Genomic sequences of bovine *NK2A*, *NK2B* and *NK2C* were aligned using the Multiple sequence alignment by Florence Corpet. A specific pair of primers was designed for each *NK2* gene (Gs-NK2A – C) by locating the 3' end of primers at the mismatched nucleotide loci, and the specificity of primers were confirmed in the genomic DNA of L1 Domino 99375 which contains single copies of each *NK2* gene. Genomic DNA of individuals from different breeds was amplified by the specific primers and amplicons were sequenced with Sanger sequencing at the DNA Technologies Core Lab (Texas A&M University). Sequences of all amplicons obtained by the same pair of primers were aligned to the *NK2A*, *NK2B* and *NK2C* sequences from L1 Domino 99375 using the Multiple sequence alignment by Florence Corpet.

Table 6. Primer and amplicon information

Primer	Forward (5' to 3')	Reverse (5' to 3')	Amplicon size	Utilization
NK1-CDS	ATGACCTCTTGGGCTGTTCT	CAGATGGCCTGAGGTGTCTT	395 bp	*NK1* coding region amplification
NK1-9bp	TGACTTCTCCTCTCCCTCTGTC	AGGGAATTCTCCAAAGTCCATT	873 bp	Identification of genomic region of 9-bp INDEL
Bo-lysin	ACCCAGCACTCCCACTG	ACATACCTGGCTTGCTTTTG	293 bp	Homozygotes analysis
Gs-NK2A	TGGTCCTTGTTGGATGATAATG	TTTCCAGCTGTGATGTCTGC	893 bp	Gene conversion test
Gs-NK2B	CTGTTCATGCTGTTTCTTCCAT	TTGCACAGACCTTTCAGCG	1479 bp	*NK2B* deletion test Gene conversion test
Gs-NK2C	CTACGCTGTGGTTCTTGTCG	GTTTTTCCAGCTACGATGTCCT	891 bp	Gene conversion test

Table 7. Individuals used in the genetic study of bovine *NK-lysin* gene family

Individual	Breed	Utilization	9-bp INDEL	SNP (BovineHD1100014441)	NK2B
Domino	Hereford		- /-		
23	Brahman	9-bp INDEL test	+/+		
26	Brahman	9-bp INDEL test	+/+		
43	Charolais	9-bp INDEL test	+/+		
44	Charolais	9-bp INDEL test	+/+		
70	Gelbvieh	9-bp INDEL test	+/-		
2822	Holstein	9-bp INDEL test	+/+		
3850	Holstein	9-bp INDEL test	+/+		
4171	Holstein	9-bp INDEL test	+/+		
4356	Holstein	9-bp INDEL test	+/+		
2527	Holstein	NK2B test		T	+
3069	Holstein	NK2B test		T	+
3084	Holstein	NK2B test		T	+
3153	Holstein	NK2B test		T	+
3489	Holstein	NK2B test		T	+
3532	Holstein	NK2B test		T	+
3540	Holstein	NK2B test		T	+
3611	Holstein	NK2B test		T	+
3612	Holstein	NK2B test		T	+
3803	Holstein	NK2B test		T	+
3856	Holstein	NK2B test		T	+
4454	Holstein	NK2B test		T	+
2796	Holstein	NK2B test		T	+
3443	Holstein	NK2B test		T	+
3022	Holstein	NK2B test		T	+

Table 7. Continued

Individual	Breed	Utilization	9-bp INDEL	SNP (BovineHD1100014441)	NK2B
3318	Holstein	NK2B test		T	+
4378	Holstein	NK2B test		T	+
2822	Holstein	NK2B test		C	-
4171	Holstein	NK2B test		C	-
4494	Holstein	NK2B test		T	+
3850	Holstein	NK2B test		C	-
4356	Holstein	NK2B test		C	-
23	Brahman	Gene conversion			
26	Brahman	Gene conversion			
43	Charolais	Gene conversion			
44	Charolais	Gene conversion			
47	Charolais	Gene conversion			
70	Gelbvieh	Gene conversion			
91	Red Angus	Gene conversion			
92	Red Angus	Gene conversion			
93	Red Angus	Gene conversion			
3850	Holstein	Gene conversion			

Results

A 9-bp deletion in the third exon of NK1 causes a 3-aa deletion in the peptide

Two referenced mRNA sequences (NM_001046578 and XM_010810081) were annotated for the gene *LOC616323* (Gene ID: 616323) in the NCBI database which corresponds to our newly annotated *NK1* gene. Full coding sequences of the two referenced mRNAs were compared with the coding sequence of *NK1* predicted from the PacBio sequencing result (Fig. 20), and two INDELs with the lengths of 9-bp and 27-bp were revealed. In order to confirm the authenticity of these two INDELs in the coding region of *NK1* gene, a pair of primers NK1-CDS whose amplicon spans both INDEL sites was tested in the cDNA prepared from four different individuals. Among four individuals, one was homozygous for the 9-bp deletion (Fig. 21A), two were heterozygous for this variant (Fig. 21B) and one was homozygous for the presence of this 9-bp on both chromosomes (Fig. 21C). Since we did not have the genomic information for the four tested individuals, we sequenced the genomic fragment spanning exon 2 – intron 2 – exon 3 of *NK1* in nine more individuals to identify the corresponding genomic location of this 9-bp variant (Table 7). Individual 70 was the only one carrying the 9-bp deletion while the rest were all homozygous for the presence of this 9-bp fragment. By comparing with the genomic sequence extracted from BAC assembly which contains the 9-bp deletion, we were able to identify the location of this INDEL in the third exon of *NK1*. However, the 27-bp deletion was not detected in any of the four tested individuals.

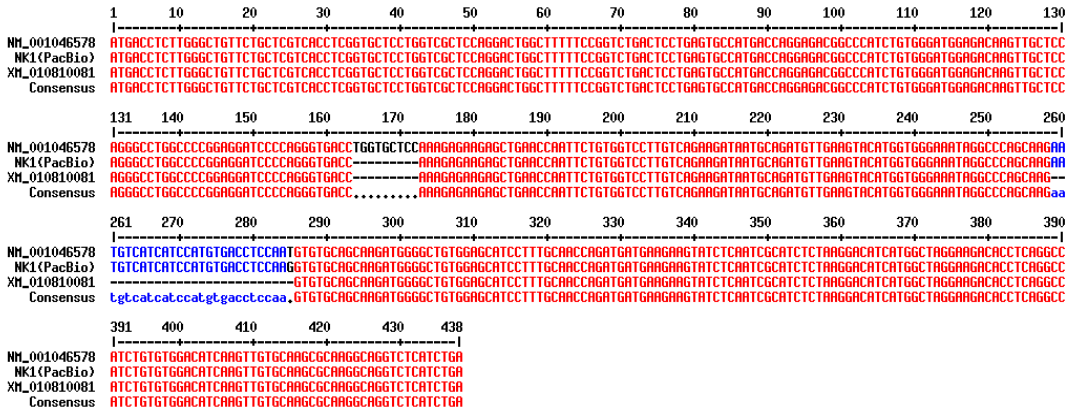


Figure 20. Full coding sequence comparison between two referenced *NK1* mRNAs extracted from the NCBI database (NM_001046578 and XM_010810081) and the sequence (NK1(PacBio)) predicted from the PacBio sequencing result. Red color indicates alignment with high consensus and blue represents alignment with low consensus.

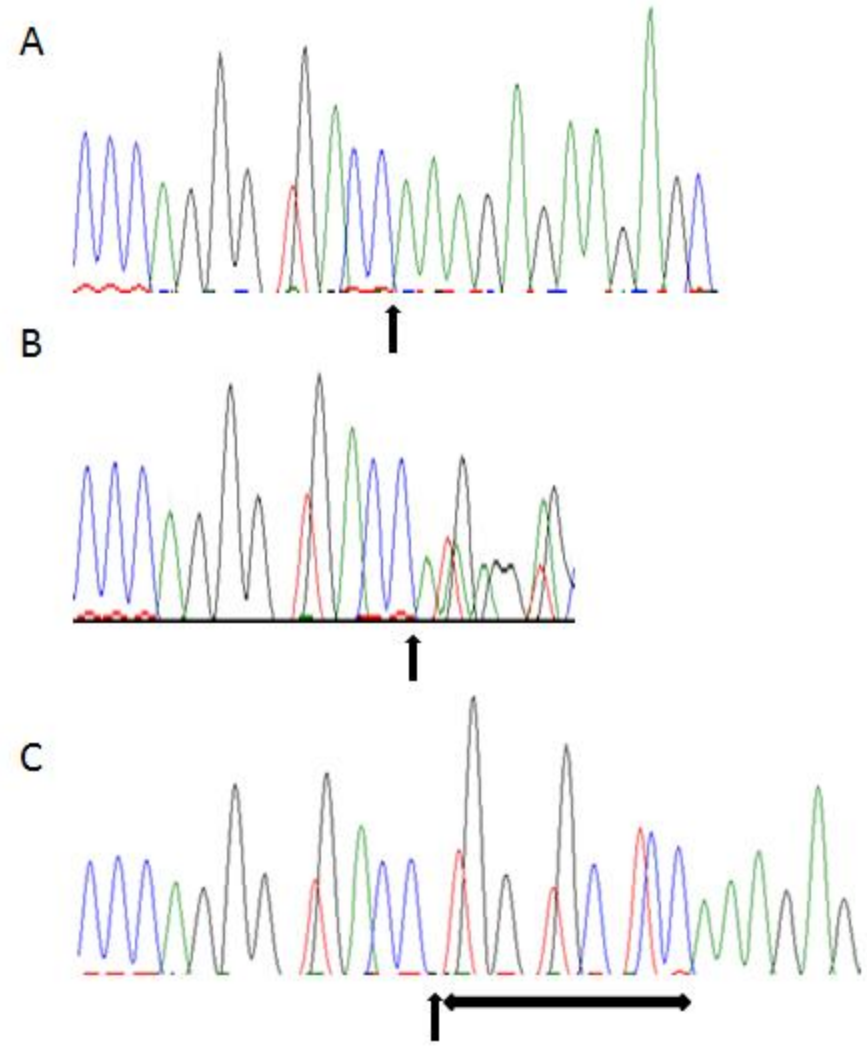


Figure 21. A 9-bp INDEL (TGGTGCTCC) in the coding region of bovine *NK1* gene. (A) Individual homozygous for the deletion. (B) Individual heterozygous for the deletion. (C) Individual homozygous for the presence of this 9-bp fragment. Black arrow (↑) indicates the INDEL site and horizontal black arrow (↔) represents the 9-bp INDEL (TGGTGCTCC).

Gene conversions within the bovine NK-lysin gene family

We employed the method of shared polymorphisms to study the potential gene conversion events and their contributions to the high sequence identity among *NK2A*, *NK2B* and *NK2C* genes. Specific primers for each *NK2* gene were tested in different individuals and the amplicons were compared with the corresponding region from the other two *NK2* genes (Fig. 22). Two gene conversion events were identified within the 893-bp amplicon of *NK2A* gene, and these two events were in linkage disequilibrium. The minimum conversion tract of the first event was 57 bp with *NK2B* being the donor gene while the second one was 154 bp with *NK2C* being the donor (Fig. 22 A, B). A gene conversion event with the minimum tract of 22 bp was also found in the 1479-bp amplicon of *NK2B* gene, and the *NK2C* was the donor. However, we were not able to detect any gene conversion events within the 891-bp amplicon of *NK2C* gene.

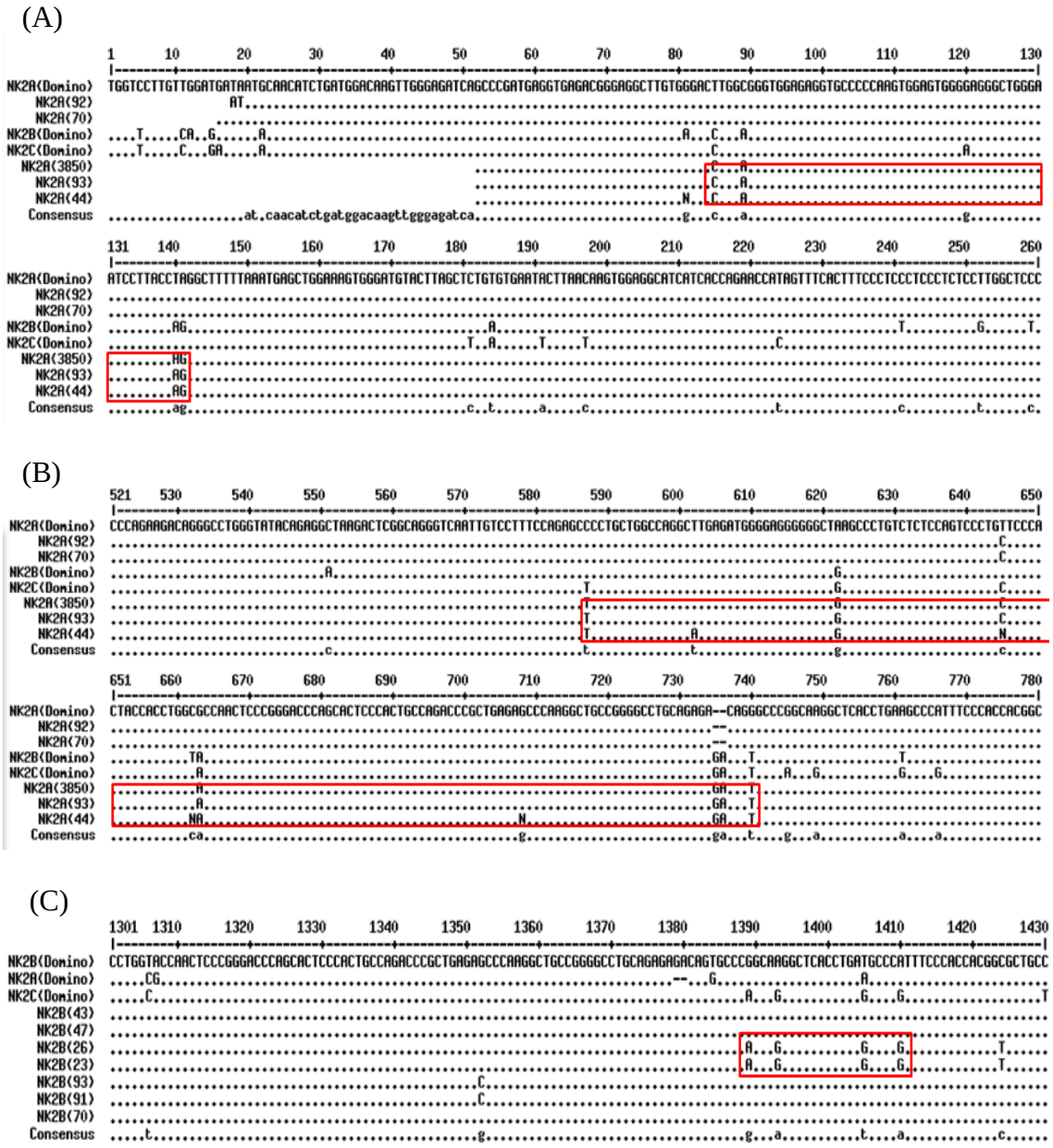


Figure 22. Gene conversion events within the bovine *NK-lysin* gene family. (A) (B) Amplicons of *NK2A* in individuals 92, 70, 3850, 93 and 44 were compared with the corresponding regions of *NK2B* and *NK2C* from L1 Domino 99375. All sequences were aligned to the NK2A (Domino). (C) Amplicons of *NK2B* in individuals 43, 47, 26, 23, 93, 91 and 70 were compared with the corresponding regions of *NK2A* and *NK2C* from L1 Domino 99375. All sequences were aligned to the NK2B (Domino). Red boxes represent the minimum gene conversion tracts.

Copy number variation of NK2B in Holstein cattle

Since *NK2A*, *NK2B* and *NK2C* genes arose by tandem duplications and share high sequence identity to each other, this is a candidate region for unequal crossovers to occur during meiosis resulting in copy number variation of bovine *NK2* genes. To further confirm the authenticity of three *NK2* genes as well as the potential copy number variation of each gene, we designed a pair of primers (Bo-lysin) from the conserved region of three *NK2* genes. To minimize effects of allelic variation in the analysis, we selected four Holstein cattle homozygous for this region based on GWAS genotyping results with the 770K HD SNP array [167]. The SNP array contained 29 SNPs between the two genes flanking the *NK-lysin* region, *ATOH8* [Gene ID: 616225] and *SFTPB* [Gene ID: 507398]. The PLINK program was utilized to identify individuals which were homozygous at all 29 SNP sites, and four cattle (2527, 2796, 2822 and 3850) with different haplotypes were selected for further analysis (Table 8). The number of the sequenced clones and different sequences obtained from each individual are listed in the Table 9. In total, five different sequences (Seq1-5) were recovered from these four individuals. The five sequences formed three clades, corresponding to the *NK2A*, *NK2B* and *NK2C* genes, and were divergent from *NK1* (Fig. 23). Three different arrangements of *NK-lysin* genes were observed in this study. In the *NK2A* cluster, two sequences from individual 2527 were detected. If the individual 2527 was homozygous across the *NK-lysin* region, then at least two copies of *NK2A* were present in this animal. Despite the large number of sequenced clones from both individuals 2822 and 3850, we found no

NK2B-related clones, and we could not obtain *NK2B* amplicons with *NK2B* specific primer (Gs-NK2B), which suggested the absence of the *NK2B* gene in these animals.

Table 8. Haplotypes of Holstein cattle homozygous across the entire *NK-lysin* region. Different colors represent different haplotype structures.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**16**	17	18	19	20	21	22	23	24	25	26	27	28	29
2527	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3069	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3084	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3153	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3255	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3302	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3472	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3489	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3532	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3540	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3611	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3612	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3803	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3856	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
4454	B	B	B	A	B	B	B	A	B	A	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
2796	B	B	B	A	B	B	B	A	B	A	A	A	0	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3170	B	B	B	A	B	B	B	A	B	A	A	A	0	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3443	B	B	B	A	B	B	B	A	B	A	A	A	0	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3022	A	A	A	B	A	B	B	A	A	A	A	A	B	B	A	A	0	B	B	B	A	A	B	A	A	A	B	A	A
3318	A	A	A	B	A	B	B	A	A	A	A	A	B	B	A	A	0	B	B	B	A	A	B	A	A	A	B	A	A
4378	A	A	A	B	A	B	B	A	A	A	A	A	B	B	A	A	0	B	B	B	A	A	B	A	A	A	B	A	A
2822	A	A	A	B	A	B	B	A	A	A	A	B	B	B	A	B	B	B	A	B	A	A	B	B	B	B	A	A	A
4171	A	A	A	B	A	B	B	A	A	A	A	B	B	B	A	B	B	B	A	B	A	A	B	B	B	B	A	A	A
4494	B	B	B	A	B	B	B	A	B	B	A	A	B	B	A	A	B	B	A	B	A	A	B	A	A	B	A	A	A
3850	A	B	B	B	B	A	A	B	B	B	B	B	B	B	A	B	B	B	A	B	A	A	B	B	B	B	A	A	A
4356	A	B	B	B	B	A	A	B	B	B	B	B	B	B	A	B	B	B	A	B	A	A	B	B	B	B	A	A	A

Table 9. Number of sequenced clones and different sequences obtained from each individual in the analysis of homozygous cattle.

Animal ID	No. sequenced clones	No. different clones
2527	18	4
2796	30	3
2822	39	2
3850	30	2

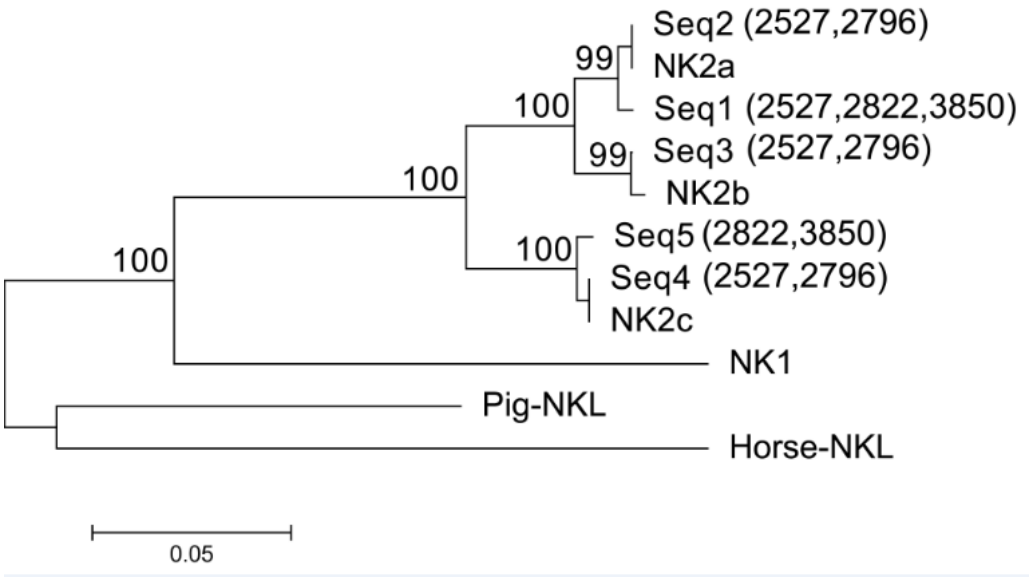


Figure 23. *NK2A, NK2B and NK2C* nucleotide sequence analysis in four homozygous individuals (2527, 2796, 2822 and 3850). Five different clone sequences from four individuals (Seq 1 - 5) were phylogenetically analyzed with four bovine *NK-lysin* reference sequences (*NK1, NK2A, NK2B* and *NK2C*) and corresponding pig (Pig-NKL) and horse (Horse-NKL) orthologs by MEGA 6.0. Bootstrap values are shown at branch points.

Deletion of NK2B in Holstein cattle is in linkage disequilibrium with a SNP from the 770K HD SNP array

A total of 26 Holstein cattle with 6 different haplotypes were identified as homozygous for all the 29 SNPs within the bovine *NK-lysin* region (Table 8). The position of the 16th SNP (BovineHD1100014441, T/C) is Chr. 11: 49093033, which is 2429 bp upstream of the NCBI annotated *LOC104968634* gene (Gene ID: 104968634) corresponding to *NK2B*. Individuals 2527 and 2796 with *NK2B* gene from the homozygous analysis showed genotype A at the 16th SNP while individuals 2822 and 3850 without *NK2B* showed genotype B at this SNP site. To verify the linkage of *NK2B* with this SNP, we genotyped the *NK2B* by PCR with the Gs-NK2B primers in more individuals and the result revealed the complete linkage of *NK2B* deletion with the genotype B at the 16th SNP site among all the tested individuals (Table 7).

Discussion

Although often small in size, the pro-region is an important part of a precursor protein and plays a multifaceted role required for the maturation of an active protein. A 9-bp INDEL in the third exon of the bovine *NK1* gene causing a 3-aa deletion in the pro-region of the protein was identified in this study. Further studies are therefore suggested to investigate the effects of this 3-aa deletion on protein folding, secretion and antimicrobicities. Another genetic variant identified within the bovine *NK-lysin* family was the copy number variation of *NK2B* which is in linkage disequilibrium with a SNP from the bovine 770K HD SNP array. Also, it is clear from the homozygous analysis

that deletion of *NK2B* was not linked with the deletion of the other two *NK2* genes, indicating that three *NK2* genes were not located within the same copy number variation region (CNVR). It will be interesting to test the copy number variations of other bovine *NK-lysin* genes and their linkage with SNPs from the bovine 770K HD SNP array. Since thousands of cattle including both beef and dairy cattle have been genotyped with the 770K HD SNP array and phenotypes scored for BRD symptoms (http://www.brdcomplex.org/), the newly identified genetic variants in this study could be incorporated together with the SNP genotypes to analyze the haplotype structures and investigate the linkage disequilibrium within the bovine *NK-lysin* region. Haplotype – to – phenotype association study could then be performed to identify the haplotypes associated with host resistance to bovine respiratory diseases. On the other hand, haplotype structures could also be studied in different breeds of cattle to gain insights into the evolutionary history of the *NK-lysin* region in breed differentiation.

CHAPTER VI

DISCUSSION AND CONCLUSIONS

Discussion

A single copy of the *NK-lysin* gene is annotated in the genomes of most mammals including humans, but our study identified a family of *NK-lysin* genes in cattle, consistent with the numerical expansion of other immune-related genes in the cattle genome, including *interferons*, *defensins* and *cathelicidins*. These expansions may reflect an evolutionary strategy to deal with the substantial numbers of pathogens and the increased risk of infections in the rumen of cattle. Therefore, a single copy of the *NK-lysin* has also likely evolved into a gene family in other ruminant species. Two *NK-lysin* mRNA sequences were annotated in the NCBI nucleotide database for each of the ruminant species, including the water buffalo, sheep, goat and bison. Phylogenetic analysis revealed that one mRNA sequence from each species was clustered with the bovine *NK1* gene while the other one was clustered with the bovine *NK2* genes (Fig. 24). A pair of primers (Con-NK2 F: AGATTTGATGGACCCGAGCA, R: GTGAAACTATGGTTC TGGTGATGA) was designed within the conserved region of three bovine *NK2* genes and recovered a NK2-related amplicon in several other ruminant species, including elk, white tail deer, swamp buffalo and gaur. Whether all ruminant species contain single copies of *NK1* and *NK2* or a single copy of *NK1* plus three copies of *NK2* genes needs to be determined. Re-sequencing the *NK-lysin* region with PacBio SMRT technology is suggested to investigate the genomic organization of this gene

family in ruminant species to gain insights into the evolutionary history of the gene family expansion.

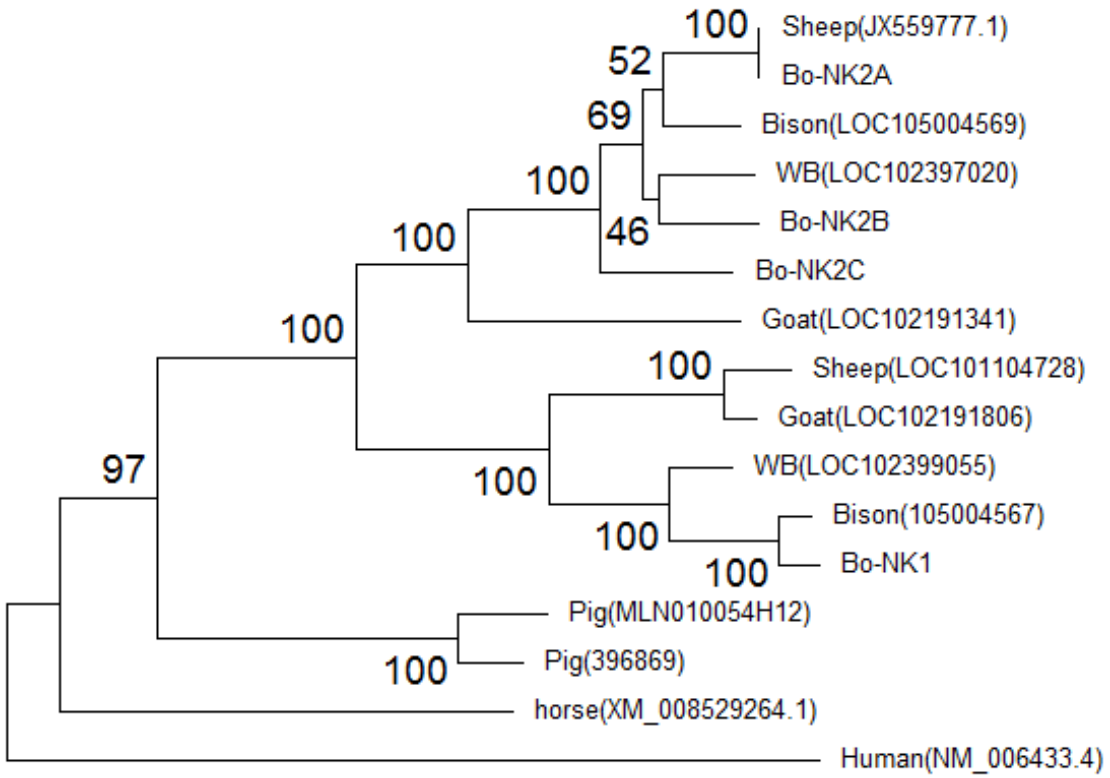


Figure 24. Phylogenetic analysis of the full coding sequences of four bovine *NK-lysins* and *NK-lysin* orthologs in humans, pig, horse, sheep, goat, water buffalo (WB) and bison. The accession number for each sequence in the NCBI nucleotide database is indicated and the bootstrap values are shown at branch points.

During gene family expansion, the ancestor gene usually maintains the original biological function, while the duplicated genes become pseudogenes or are free from purifying selection and may eventually evolve into genes with novel functions. All four bovine *NK-lysin* genes are expressed, albeit with different expression levels and in different tissues. Both expression assays and RNA-seq data revealed high expression of the *NK2C* gene in lung indicating its gain of a novel function in the bovine respiratory system. Further studies could be performed to compare the regulatory sequences of *NK2C* with other *NK-lysin* paralogs to identify the elements responsible for the high expression of *NK2C* in lung. In this study, gene expression was only tested in five tissues. It would be important to determine expression of bovine *NK-lysins* in more tissue types to investigate the potential novel functions of other duplicated genes.

We also confirmed the antimicrobicity of the functional region helices 2 & 3 of each bovine NK-lysin peptide against both Gram-positive and Gram-negative bacteria. Each peptide possesses a specific property, including the charge, hydrophobicity and amphipathicity, which might enable the bovine *NK-lysin* family to be active against a broad spectrum of microorganisms. Although the bacterial killing assays showed relatively weaker antimicrobial activities of the synthetic NK2B peptide against all six tested bacterial strains, it is possible that NK2B has different targeted microorganisms. The activities of four peptides should be tested against more bacterial strains to identify the specific targets of each *NK-lysin* gene.

In Chapter V, our preliminary studies identified several genetic variants within the bovine *NK-lysin* region, including copy number polymorphism for *NK2B* in Holstein cattle and a 9-bp INDEL in the third exon of *NK1* gene causing deletion of 3 amino acids in the pro-region of NK1 protein. Further studies should be performed to test the effects of these two variants on gene function. It will be important to investigate additional *NK-lysin* genetic variants and their associations with host resistance as candidates to improve animal health. However, a big challenge of studying this gene family is the high sequence identity shared by the three *NK2* genes. Extra care must be taken to distinguish allelic variation from paralogous variation.

Conclusion

Seven different *NK-lysin* related mRNA sequences in four phylogenetic clusters were identified in the NCBI bovine nucleotide database, suggesting the existence of multiple copies of *NK-lysin* genes in the cattle genome. This project focused on both the genetic and functional characterization of bovine *NK-lysins*. As described in CHAPTER II, two BAC clones covering the whole bovine *NK-lysin* region were sequenced with PacBio SMRT sequencing technology, and four *NK-lysin* genes were identified, *NK1*, *NK2A*, *NK2B* and *NK2C*. *NK2A*, *NK2B* and *NK2C* are tandemly arrayed as three copies in 30 ~ 35 Kb segments, located 41.8 Kb upstream of *NK1*. Analysis of repeat element revealed that fragments flanking each breakpoint share high homology, and could contribute to the gene family expansion as well as potential unequal crossover during meiosis and structural instability within the bovine *NK-lysin* gene family. CHAPTER III describes

the expression profiles of the four identified bovine *NK-lysin* genes in five different tissue types, and the effects of the synthetic peptides corresponding to the functional helices 2 & 3 of each gene product on both model membranes and bacterial membranes. Different from the other *NK-lysin* paralogs and orthologs, bovine *NK2C* is highly expressed in lung, indicating the evolution of a novel function in the bovine respiratory system. The disruptive effects of bovine NK-lysin peptides on the anionic model membrane and bacterial membrane were also confirmed in this study, consistent with the antimicrobial effects of other positively charged AMP molecules. Since *NK2C* shows a high expression level in lung, we investigated the potential function of bovine *NK-lysins* in host resistance to bovine respiratory pathogens as described in CHAPTER IV. Expression levels of each bovine *NK-lysin* gene in bronchial lymph node and lung were compared between control animals and animals challenged with different BRD-associated pathogens by analyzing RNA-seq data. The expression of several *NK-lysins*, especially *NK2C*, was elevated in challenged relative to control animals, indicating its potential importance in host defense against pathogens involved in bovine respiratory disease. Antimicrobial effects of the bovine NK-lysin peptides on BRD-causing *Pasteurella multocida* and *Mannhemia haemolytica* bacterial strains are also confirmed in this chapter. CHAPTER V describes some identified genetic polymorphisms within the bovine *NK-lysin* region, including a 9-bp INDEL in the third exon of *NK1* causing a 3-aa deletion in the pro-region of the peptide and the deletion of *NK2B* in some Holstein cattle. Evidence for gene conversion events were also found among three *NK2* genes, which might explain the high sequence identity shared by these three genes.

In conclusion, this study is the first to clarify the organization and structure of the *NK-lysin* gene family in the cattle genome and provide a foundation for future studies of *NK-lysin* genes in maintaining the health of cattle as well as other ruminant species.

REFERENCES

1. Hiendleder S, Lewalski H, Janke A. Complete mitochondrial genomes of Bos taurus and Bos indicus provide new insights into intra-species variation, taxonomy and domestication. Cytogenetic and genome research. 2008;120(1-2):150-6. Epub 2008/05/10. doi: 10.1159/000118756. PubMed PMID: 18467841.
2. O'Kelly JC, Spiers WG. Resistance to Boophilus microplus (Canestrini) in genetically different types of calves in early life. The Journal of parasitology. 1976;62(2):312-7. Epub 1976/04/01. PubMed PMID: 1263044.
3. Glass EJ, Preston PM, Springbett A, Craigmile S, Kirvar E, Wilkie G, et al. Bos taurus and Bos indicus (Sahiwal) calves respond differently to infection with Theileria annulata and produce markedly different levels of acute phase proteins. International journal for parasitology. 2005;35(3):337-47. Epub 2005/02/22. doi: 10.1016/j.ijpara.2004.12.006. PubMed PMID: 15722085.
4. Beatty DT, Barnes A, Taylor E, Pethick D, McCarthy M, Maloney SK. Physiological responses of Bos taurus and Bos indicus cattle to prolonged, continuous heat and humidity. Journal of animal science. 2006;84(4):972-85. Epub 2006/03/18. PubMed PMID: 16543576.
5. Diamond J. Evolution, consequences and future of plant and animal domestication. Nature. 2002;418(6898):700-7. Epub 2002/08/09. doi: 10.1038/nature01019. PubMed PMID: 12167878.

6. Vigne JD. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. Comptes rendus biologies. 2011;334(3):171-81. Epub 2011/03/08. doi: 10.1016/j.crvi.2010.12.009. PubMed PMID: 21377611.

7. Mona S, Catalano G, Lari M, Larson G, Boscato P, Casoli A, et al. Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. BMC Evolutionary Biology. 2010;10:83. doi: 10.1186/1471-2148-10-83. PubMed PMID: 20346116; PubMed Central PMCID: PMCPmc2858146.

8. Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P. Evidence for two independent domestications of cattle. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(7):2757-61. Epub 1994/03/29. PubMed PMID: 8146187; PubMed Central PMCID: PMCPmc43449.

9. MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus). Genetics. 1997;146(3):1071-86. Epub 1997/07/01. PubMed PMID: 9215909; PubMed Central PMCID: PMCPmc1208036.

10. Grigson C. An African Origin for African Cattle? Some Archaeological Evidence. The African Archaeological Review. 1991;9:119-44. doi: 10.2307/25130537.

11. Bradley DG, MacHugh DE, Cunningham P, Loftus RT. Mitochondrial diversity and the origins of African and European cattle. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(10):5131-5. Epub 1996/05/14. PubMed PMID: 8643540; PubMed Central PMCID: PMCPmc39419.

12. Rouse JE. The Criollo: Spanish cattle in the Americas. Norman: University of Oklahoma Press.; 1977. xvi + 303 pp. p.

13. Ginja C, Penedo MC, Melucci L, Quiroz J, Martinez Lopez OR, Revidatti MA, et al. Origins and genetic diversity of New World Creole cattle: inferences from mitochondrial and Y chromosome polymorphisms. Animal genetics. 2010;41(2):128-41. Epub 2009/10/13. doi: 10.1111/j.1365-2052.2009.01976.x. PubMed PMID: 19817725.

14. Zasloff M. Antimicrobial peptides of multicellular organisms. Nature. 2002;415(6870):389-95. Epub 2002/01/25. doi: 10.1038/415389a. PubMed PMID: 11807545.

15. Garcia-Olmedo F, Molina A, Alamillo JM, Rodriguez-Palenzuela P. Plant defense peptides. Biopolymers. 1998;47(6):479-91. Epub 1999/05/20. doi: 10.1002/(sici)1097-0282(1998)47:6<479::aid-bip6>3.0.co;2-k. PubMed PMID: 10333739.

16. Bowdish DM, Davidson DJ, Hancock RE. A re-evaluation of the role of host defence peptides in mammalian immunity. Current protein & peptide science. 2005;6(1):35-51. Epub 2005/01/11. PubMed PMID: 15638767.

17. Diamond G, Beckloff N, Weinberg A, Kisich KO. The Roles of Antimicrobial Peptides in Innate Host Defense. Current pharmaceutical design. 2009;15(21):2377-92. PubMed PMID: 19601838; PubMed Central PMCID: PMCPmc2750833.

18. Schittek B, Hipfel R, Sauer B, Bauer J, Kalbacher H, Stevanovic S, et al. Dermcidin: a novel human antibiotic peptide secreted by sweat glands. Nature immunology. 2001;2(12):1133-7. Epub 2001/11/06. doi: 10.1038/ni732. PubMed PMID: 11694882.

19. Steiner H, Hultmark D, Engstrom A, Bennich H, Boman HG. Sequence and specificity of two antibacterial proteins involved in insect immunity. Nature. 1981;292(5820):246-8. Epub 1981/07/16. PubMed PMID: 7019715.

20. Zasloff M. Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. Proceedings of the National Academy of Sciences of the United States of America. 1987;84(15):5449-53. Epub 1987/08/01. PubMed PMID: 3299384; PubMed Central PMCID: PMCPmc298875.

21. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol. 2005;3(3):238-50. Epub 2005/02/11. doi: 10.1038/nrmicro1098. PubMed PMID: 15703760.

22. Zanetti M, Gennaro R, Romeo D. Cathelicidins: a novel protein family with a common proregion and a variable C-terminal antimicrobial domain. FEBS letters. 1995;374(1):1-5. Epub 1995/10/23. PubMed PMID: 7589491.

23. Selsted ME, Harwig SS, Ganz T, Schilling JW, Lehrer RI. Primary structures of three human neutrophil defensins. The Journal of clinical investigation. 1985;76(4):1436-9. Epub 1985/10/01. doi: 10.1172/jci112121. PubMed PMID: 4056036; PubMed Central PMCID: PMCPmc424095.

24. Lehrer RI, Ganz T. Defensins of vertebrate animals. Current opinion in immunology. 2002;14(1):96-102. Epub 2002/01/16. PubMed PMID: 11790538.

25. O'Brien JS, Kishimoto Y. Saposin proteins: structure, function, and role in human lysosomal storage disorders. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 1991;5(3):301-8. Epub 1991/03/01. PubMed PMID: 2001789.

26. Hawgood S, Benson BJ, Schilling J, Damm D, Clements JA, White RT. Nucleotide and amino acid sequences of pulmonary surfactant protein SP 18 and evidence for cooperation between SP 18 and SP 28-36 in surfactant lipid adsorption. Proceedings of the National Academy of Sciences of the United States of America. 1987;84(1):66-70. Epub 1987/01/01. PubMed PMID: 3467361; PubMed Central PMCID: PMCPmc304142.

27. Ponting CP. Acid sphingomyelinase possesses a domain homologous to its activator proteins: saposins B and D. Protein science : a publication of the Protein Society. 1994;3(2):359-61. Epub 1994/02/01. doi: 10.1002/pro.5560030219. PubMed PMID: 8003971; PubMed Central PMCID: PMCPmc2142785.

28. Staab JF, Ginkel DL, Rosenberg GB, Munford RS. A saposin-like domain influences the intracellular localization, stability, and catalytic activity of human acyloxyacyl hydrolase. The Journal of biological chemistry. 1994;269(38):23736-42. Epub 1994/09/23. PubMed PMID: 8089145.

29. Kervinen J, Tobin GJ, Costa J, Waugh DS, Wlodawer A, Zdanov A. Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. The EMBO journal. 1999;18(14):3947-55. Epub 1999/07/16. doi: 10.1093/emboj/18.14.3947. PubMed PMID: 10406799; PubMed Central PMCID: PMCPmc1171470.

30. Keller F, Hanke W, Trissl D, Bakker-Grunwald T. Pore-forming protein from Entamoeba histolytica forms voltage- and pH-controlled multi-state channels with properties similar to those of the barrel-stave aggregates. Biochimica et biophysica acta. 1989;982(1):89-93. Epub 1989/06/26. PubMed PMID: 2472838.

31. Zhai Y, Saier MH, Jr. The amoebapore superfamily. Biochimica et biophysica acta. 2000;1469(2):87-99. Epub 2000/09/22. PubMed PMID: 10998571.

32. Pena SV, Hanson DA, Carr BA, Goralski TJ, Krensky AM. Processing, subcellular localization, and function of 519 (granulysin), a human late T cell activation molecule with homology to small, lytic, granule proteins. Journal of immunology (Baltimore, Md : 1950). 1997;158(6):2680-8. Epub 1997/03/15. PubMed PMID: 9058801.

33. Andersson M, Gunne H, Agerberth B, Boman A, Bergman T, Sillard R, et al. NK-lysin, a novel effector peptide of cytotoxic T and NK cells. Structure and

cDNA cloning of the porcine form, induction by interleukin 2, antibacterial and antitumour activity. The EMBO journal. 1995;14(8):1615-25. Epub 1995/04/18. PubMed PMID: 7737114; PubMed Central PMCID: PMCPmc398254.

34. Dubos R é J. Studies on a bactericidal agent extracted from a Soil Bacillus: I. preparation of the agent. Its activity in vitro. The Journal of Experimental Medicine. 1939;70(1):1-10. PubMed PMID: 19870884; PubMed Central PMCID: PMCPmc2133784.

35. Hotchkiss RD, Dubos RJ. Fractionation of the bactericidal agent from cultures of a Soil Bacillus. Journal of Biological Chemistry. 1940;132(2):791-2.

36. Hirsch JG. Phagocytin: a bactericidal substance from polymorphonuclear leucocytes. The Journal of Experimental Medicine. 1956;103(5):589-611. PubMed PMID: 13319580; PubMed Central PMCID: PMCPmc2136631.

37. Aerts AM, Francois IE, Cammue BP, Thevissen K. The mode of antifungal action of plant, insect and human defensins. Cellular and molecular life sciences : CMLS. 2008;65(13):2069-79. Epub 2008/03/25. doi: 10.1007/s00018-008-8035-0. PubMed PMID: 18360739.

38. Wilson SS, Wiens ME, Smith JG. Antiviral mechanisms of human defensins. Journal of molecular biology. 2013;425(24):4965-80. Epub 2013/10/08. doi: 10.1016/j.jmb.2013.09.038. PubMed PMID: 24095897; PubMed Central PMCID: PMCPmc3842434.

39. Nascimento VV, Mello EO, Carvalho LP, de Melo EJ, Carvalho AO, Fernandes KV, et al. PvD1 defensin, a plant antimicrobial peptide with inhibitory activity

against Leishmania amazonensis. Bioscience reports. 2015. Epub 2015/08/20. doi: 10.1042/bsr20150060. PubMed PMID: 26285803.

40. Guzman-Rodriguez JJ, Ochoa-Zarzosa A, Lopez-Gomez R, Lopez-Meza JE. Plant antimicrobial peptides as potential anticancer agents. BioMed research international. 2015;2015:735087. Epub 2015/03/31. doi: 10.1155/2015/735087. PubMed PMID: 25815333; PubMed Central PMCID: PMCPmc4359852.

41. Territo MC, Ganz T, Selsted ME, Lehrer R. Monocyte-chemotactic activity of defensins from human neutrophils. The Journal of clinical investigation. 1989;84(6):2017-20. Epub 1989/12/01. doi: 10.1172/jci114394. PubMed PMID: 2592571; PubMed Central PMCID: PMCPmc304087.

42. Yang D, Chen Q, Chertov O, Oppenheim JJ. Human neutrophil defensins selectively chemoattract naive T and immature dendritic cells. Journal of leukocyte biology. 2000;68(1):9-14. Epub 2000/07/29. PubMed PMID: 10914484.

43. Lillard JW, Jr., Boyaka PN, Chertov O, Oppenheim JJ, McGhee JR. Mechanisms for induction of acquired host immunity by neutrophil peptide defensins. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(2):651-6. Epub 1999/01/20. PubMed PMID: 9892688; PubMed Central PMCID: PMCPmc15191.

44. Scott MG, Dullaghan E, Mookherjee N, Glavas N, Waldbrook M, Thompson A, et al. An anti-infective peptide that selectively modulates the innate immune

response. Nature biotechnology. 2007;25(4):465-72. Epub 2007/03/27. doi: 10.1038/nbt1288. PubMed PMID: 17384586.

45. Heilborn JD, Nilsson MF, Kratz G, Weber G, Sorensen O, Borregaard N, et al. The cathelicidin anti-microbial peptide LL-37 is involved in re-epithelialization of human skin wounds and is lacking in chronic ulcer epithelium. The Journal of investigative dermatology. 2003;120(3):379-89. Epub 2003/02/27. doi: 10.1046/j.1523-1747.2003.12069.x. PubMed PMID: 12603850.

46. Utsugi T, Schroit AJ, Connor J, Bucana CD, Fidler IJ. Elevated expression of phosphatidylserine in the outer membrane leaflet of human tumor cells and recognition by activated human blood monocytes. Cancer research. 1991;51(11):3062-6. Epub 1991/06/01. PubMed PMID: 2032247.

47. Yoon WH, Park HD, Lim K, Hwang BD. Effect of O-glycosylated mucin on invasion and metastasis of HM7 human colon cancer cells. Biochemical and biophysical research communications. 1996;222(3):694-9. Epub 1996/05/24. doi: 10.1006/bbrc.1996.0806. PubMed PMID: 8651907.

48. Zachowski A. Phospholipids in animal eukaryotic membranes: transverse asymmetry and movement. The Biochemical journal. 1993;294 (Pt 1):1-14. Epub 1993/08/15. PubMed PMID: 8363559; PubMed Central PMCID: PMCPmc1134557.

49. Dathe M, Wieprecht T. Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells.

Biochimica et biophysica acta. 1999;1462(1-2):71-87. Epub 1999/12/11. PubMed PMID: 10590303.

50. Yount NY, Yeaman MR. Peptide antimicrobials: cell wall as a bacterial target. Ann N Y Acad Sci. 2013;1277:127-38. Epub 2013/01/11. doi: 10.1111/nyas.12005. PubMed PMID: 23302022.

51. Cho JH, Sung BH, Kim SC. Buforins: histone H2A-derived antimicrobial peptides from toad stomach. Biochim Biophys Acta. 2009;1788(8):1564-9. Epub 2008/12/02. doi: 10.1016/j.bbamem.2008.10.025. PubMed PMID: 19041293.

52. Nan YH, Park KH, Park Y, Jeon YJ, Kim Y, Park IS, et al. Investigating the effects of positive charge and hydrophobicity on the cell selectivity, mechanism of action and anti-inflammatory activity of a Trp-rich antimicrobial peptide indolicidin. FEMS Microbiol Lett. 2009;292(1):134-40. Epub 2009/02/05. doi: 10.1111/j.1574-6968.2008.01484.x. PubMed PMID: 19191872.

53. Haney EF, Petersen AP, Lau CK, Jing W, Storey DG, Vogel HJ. Mechanism of action of puroindoline derived tryptophan-rich antimicrobial peptides. Biochim Biophys Acta. 2013;1828(8):1802-13. Epub 2013/04/09. doi: 10.1016/j.bbamem.2013.03.023. PubMed PMID: 23562406.

54. Bals R, Wilson JM. Cathelicidins--a family of multifunctional antimicrobial peptides. Cell Mol Life Sci. 2003;60(4):711-20. Epub 2003/06/06. PubMed PMID: 12785718.

55. Cruz-Chamorro L, Puertollano MA, Puertollano E, de Cienfuegos GA, de Pablo MA. In vitro biological activities of magainin alone or in combination with nisin.

Peptides. 2006;27(6):1201-9. Epub 2005/12/17. doi: 10.1016/j.peptides.2005.11.008. PubMed PMID: 16356589.

56. Zhang HT, Wu J, Zhang HF, Zhu QF. Efflux of potassium ion is an important reason of HL-60 cells apoptosis induced by tachyplesin. Acta Pharmacol Sin. 2006;27(10):1367-74. Epub 2006/09/30. doi: 10.1111/j.1745-7254.2006.00377.x. PubMed PMID: 17007745.

57. Norrby SR, Nord CE, Finch R. Lack of development of new antimicrobial drugs: a potential serious threat to public health. Lancet Infect Dis. 2005;5(2):115-9. Epub 2005/02/01. doi: 10.1016/s1473-3099(05)01283-1. PubMed PMID: 15680781.

58. Fjell CD, Hiss JA, Hancock RE, Schneider G. Designing antimicrobial peptides: form follows function. Nature reviews Drug discovery. 2012;11(1):37-51. Epub 2011/12/17. doi: 10.1038/nrd3591. PubMed PMID: 22173434.

59. Lipsky BA, Holroyd KJ, Zasloff M. Topical versus systemic antimicrobial therapy for treating mildly infected diabetic foot ulcers: a randomized, controlled, double-blinded, multicenter trial of pexiganan cream. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2008;47(12):1537-45. Epub 2008/11/08. doi: 10.1086/593185. PubMed PMID: 18990064.

60. Hancock RE, Sahl HG. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nature biotechnology. 2006;24(12):1551-7. Epub 2006/12/13. doi: 10.1038/nbt1267. PubMed PMID: 17160061.

61. Hilpert K, Volkmer-Engert R, Walter T, Hancock REW. High-throughput generation of small antibacterial peptides with improved activity. Nat Biotech. 2005;23(8):1008-12.

62. Desai TR, Wong JP, Hancock RE, Finlay WH. A novel approach to the pulmonary delivery of liposomes in dry powder form to eliminate the deleterious effects of milling. Journal of pharmaceutical sciences. 2002;91(2):482-91. Epub 2002/02/09. PubMed PMID: 11835207.

63. Andersson M, Gunne H, Agerberth B, Boman A, Bergman T, Sillard R, et al. NK-lysin, a novel effector peptide of cytotoxic T and NK cells. Structure and cDNA cloning of the porcine form, induction by interleukin 2, antibacterial and antitumour activity. The EMBO journal. 1995;14(8):1615-25. PubMed PMID: 7737114; PubMed Central PMCID: PMCPmc398254.

64. Jongstra J, Schall TJ, Dyer BJ, Clayberger C, Jorgensen J, Davis MM, et al. The isolation and sequence of a novel gene from a human functional T cell line. J Exp Med. 1987;165(3):601-14. Epub 1987/03/01. PubMed PMID: 2434598; PubMed Central PMCID: PMCPMC2188281.

65. Manning WC, O'Farrell S, Goralski TJ, Krensky AM. Genomic structure and alternative splicing of 519, a gene expressed late after T cell activation. Journal of immunology (Baltimore, Md : 1950). 1992;148(12):4036-42. Epub 1992/06/15. PubMed PMID: 1318339.

66. Yabe T, McSherry C, Bach FH, Houchins JP. A cDNA clone expressed in natural killer and T cells that likely encodes a secreted protein. The Journal of

Experimental Medicine. 1990;172(4):1159-63. Epub 1990/10/01. PubMed PMID: 2212946; PubMed Central PMCID: PMCPmc2188624.

67. Liepinsh E, Andersson M, Ruysschaert JM, Otting G. Saposin fold revealed by the NMR structure of NK-lysin. Nature structural biology. 1997;4(10):793-5. Epub 1997/10/23. PubMed PMID: 9334742.

68. Andra J, Leippe M. Candidacidal activity of shortened synthetic analogs of amoebapores and NK-lysin. Medical microbiology and immunology. 1999;188(3):117-24. Epub 2000/04/25. PubMed PMID: 10776841.

69. Jacobs T, Bruhn H, Gaworski I, Fleischer B, Leippe M. NK-lysin and its shortened analog NK-2 exhibit potent activities against Trypanosoma cruzi. Antimicrobial agents and chemotherapy. 2003;47(2):607-13. Epub 2003/01/25. PubMed PMID: 12543667; PubMed Central PMCID: PMCPmc151766.

70. Hata A, Zerboni L, Sommer M, Kaspar AA, Clayberger C, Krensky AM, et al. Granulysin blocks replication of varicella-zoster virus and triggers apoptosis of infected cells. Viral Immunol. 2001;14(2):125-33. Epub 2001/06/12. doi: 10.1089/088282401750234501. PubMed PMID: 11398808.

71. Schroder-Borm H, Bakalova R, Andra J. The NK-lysin derived peptide NK-2 preferentially kills cancer cells with increased surface levels of negatively charged phosphatidylserine. FEBS letters. 2005;579(27):6128-34. Epub 2005/11/05. doi: 10.1016/j.febslet.2005.09.084. PubMed PMID: 16269280.

72. Krensky AM. Granulysin: a novel antimicrobial peptide of cytolytic T lymphocytes and natural killer cells. Biochem Pharmacol. 2000;59(4):317-20. Epub 2000/01/22. PubMed PMID: 10644038.

73. Andreu D, Carreno C, Linde C, Boman HG, Andersson M. Identification of an anti-mycobacterial domain in NK-lysin and granulysin. Biochem J. 1999;344 Pt 3:845-9. Epub 1999/12/10. PubMed PMID: 10585872; PubMed Central PMCID: PMCPMC1220707.

74. Stenger S, Hanson DA, Teitelbaum R, Dewan P, Niazi KR, Froelich CJ, et al. An antimicrobial activity of cytolytic T cells mediated by granulysin. Science (New York, NY). 1998;282(5386):121-5. Epub 1998/10/02. PubMed PMID: 9756476.

75. Zaitsev SV, Andersson M, Efanov AM, Efanova IB, Ostenson CG, Juntti-Berggren L, et al. An endogenous peptide isolated from the gut, NK-lysin, stimulates insulin secretion without changes in cytosolic free Ca2+ concentration. FEBS letters. 1998;439(3):267-70. Epub 1998/12/09. PubMed PMID: 9845335.

76. Andersson M, Girard R, Cazenave P. Interaction of NK lysin, a peptide produced by cytolytic lymphocytes, with endotoxin. Infection and immunity. 1999;67(1):201-5. Epub 1998/12/24. PubMed PMID: 9864216; PubMed Central PMCID: PMCPmc96297.

77. Ran S, Downes A, Thorpe PE. Increased exposure of anionic phospholipids on the surface of tumor blood vessels. Cancer research. 2002;62(21):6132-40. Epub 2002/11/05. PubMed PMID: 12414638.

78. Clayberger C, Krensky AM. Granulysin. Current opinion in immunology. 2003;15(5):560-5. Epub 2003/09/23. PubMed PMID: 14499265.

79. Davis EG, Sang Y, Rush B, Zhang G, Blecha F. Molecular cloning and characterization of equine NK-lysin. Veterinary immunology and immunopathology. 2005;105(1-2):163-9. Epub 2005/03/31. doi: 10.1016/j.vetimm.2004.12.007. PubMed PMID: 15797485.

80. Hong YH, Lillehoj HS, Dalloul RA, Min W, Miska KB, Tuo W, et al. Molecular cloning and characterization of chicken NK-lysin. Veterinary immunology and immunopathology. 2006;110(3-4):339-47. Epub 2006/01/03. doi: 10.1016/j.vetimm.2005.11.002. PubMed PMID: 16387367.

81. Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nat Biotech. 2013;31(2):135-41. doi: 10.1038/nbt.2478

82. Kandasamy S, Mitra A. Characterization and expression profile of complete functional domain of granulysin/NK-lysin homologue (buffalo-lysin) gene of water buffalo (Bubalus bubalis). Veterinary immunology and immunopathology. 2009;128(4):413-7. Epub 2009/01/13. doi: 10.1016/j.vetimm.2008.11.029. PubMed PMID: 19136155.

83. Endsley JJ, Furrer JL, Endsley MA, McIntosh MA, Maue AC, Waters WR, et al. Characterization of bovine homologues of granulysin and NK-lysin. Journal of immunology (Baltimore, Md : 1950). 2004;173(4):2607-14. Epub 2004/08/06. PubMed PMID: 15294977.

84. Jorde LB, Wooding SP. Genetic variation, classification and 'race'. Nat Genet. 2004.

85. Tishkoff SA, Kidd KK. Implications of biogeography of human populations for 'race' and medicine. Nat Genet. 2004;36(11 Suppl):S21-7. Epub 2004/10/28. doi: 10.1038/ng1438. PubMed PMID: 15507999.

86. Styrkarsdottir U, Thorleifsson G, Sulem P, Gudbjartsson DF, Sigurdsson A, Jonasdottir A, et al. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. Nature. 2013;497(7450):517-20. doi: 10.1038/nature12124

87. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature. 2001;411(6837):603-6. Epub 2001/06/01. doi: 10.1038/35079114. PubMed PMID: 11385577.

88. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54. doi: 10.1038/nature05329. PubMed PMID: 17122850; PubMed Central PMCID: PMCPmc2669898.

89. Korbel JO, Kim PM, Chen X, Urban AE, Weissman S, Snyder M, et al. The current excitement about copy-number variation: how it relates to gene duplications and protein families. Current opinion in structural biology. 2008;18(3):366-74. Epub 2008/05/31. doi: 10.1016/j.sbi.2008.02.005. PubMed PMID: 18511261; PubMed Central PMCID: PMCPmc2577873.

90. Behe MJ, Snoke DW. Simulating evolution by gene duplication of protein features that require multiple amino acid residues. Protein science : a publication of the Protein Society. 2004;13(10):2651-64. Epub 2004/09/02. doi: 10.1110/ps.04802904. PubMed PMID: 15340163; PubMed Central PMCID: PMCPmc2286568.

91. Ohta T. Role of gene duplication in evolution. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada. 1989;31(1):304-10. Epub 1989/01/01. PubMed PMID: 2687099.

92. Cook EH, Jr., Scherer SW. Copy-number variations associated with neuropsychiatric conditions. Nature. 2008;455(7215):919-23. Epub 2008/10/17. doi: 10.1038/nature07458. PubMed PMID: 18923514.

93. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet. 2006;7(7):552-64.

94. Lakich D, Kazazian HH, Jr., Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat Genet. 1993;5(3):236-41. Epub 1993/11/01. doi: 10.1038/ng1193-236. PubMed PMID: 8275087.

95. Petersen MB, Adelsberger PA, Schinzel AA, Binkert F, Hinkel GK, Antonarakis SE. Down syndrome due to de novo Robertsonian translocation t(14q;21q): DNA polymorphism analysis suggests that the origin of the extra 21q is maternal. American Journal of Human Genetics. 1991;49(3):529-36. PubMed PMID: 1831959; PubMed Central PMCID: PMCPmc1683126.

96. Bailey JA, Liu G, Eichler EE. An Alu transposition model for the origin and expansion of human segmental duplications. American journal of human genetics. 2003;73(4):823-34. Epub 2003/09/25. doi: 10.1086/378594. PubMed PMID: 14505274; PubMed Central PMCID: PMCPMC1180605.

97. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al. Copy number variation: new insights in genome diversity. Genome research. 2006;16(8):949-61. Epub 2006/07/01. doi: 10.1101/gr.3677206. PubMed PMID: 16809666.

98. Sharp AJ, Cheng Z, Eichler EE. Structural variation of the human genome. Annu Rev Genomics Hum Genet. 2006;7:407-42. Epub 2006/06/20. doi: 10.1146/annurev.genom.7.080505.115618. PubMed PMID: 16780417.

99. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. Trends Genet. 2002;18(2):74-82. Epub 2002/01/31. PubMed PMID: 11818139.

100. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics. 2008;1(1):4. Epub 2008/11/19. doi: 10.1186/1755-8417-1-4. PubMed PMID: 19014668; PubMed Central PMCID: PMCPMC2583991.

101. Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Mol Cell Biol. 1996;16(5):2164-73. PubMed PMID: 8628283.

102. Boulton SJ, Jackson SP. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. EMBO J. 1996;15(18):5093-103. PubMed PMID: 8890183.

103. Haviv-Chesner A, Kobayashi Y, Gabriel A, Kupiec M. Capture of linear fragments at a double-strand break in yeast. Nucleic Acids Res. 2007;35(15):5192-202. Epub 2007/08/03. doi: 10.1093/nar/gkm521. PubMed PMID: 17670800; PubMed Central PMCID: PMCPMC1976456.

104. Moore JK, Haber JE. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. Nature. 1996;383(6601):644-6. Epub 1996/10/17. doi: 10.1038/383644a0. PubMed PMID: 8857544.

105. Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature. 2005;437(7055):94-100. Epub 2005/09/02. doi: 10.1038/nature04029. PubMed PMID: 16136133; PubMed Central PMCID: PMCPMC1368961.

106. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nature reviews Genetics. 2009;10(8):551-64. Epub 2009/07/15. doi: 10.1038/nrg2593. PubMed PMID: 19597530; PubMed Central PMCID: PMCPMC2864001.

107. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell.

2007;131(7):1235-47. Epub 2007/12/28. doi: 10.1016/j.cell.2007.11.037. PubMed PMID: 18160035.

108. Slack A, Thornton PC, Magner DB, Rosenberg SM, Hastings PJ. On the mechanism of gene amplification induced under stress in Escherichia coli. PLoS genetics. 2006;2(4):e48. Epub 2006/04/11. doi: 10.1371/journal.pgen.0020048. PubMed PMID: 16604155; PubMed Central PMCID: PMCPMC1428787.

109. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science (New York, NY). 2004;305(5683):525-8. Epub 2004/07/27. doi: 10.1126/science.1098918. PubMed PMID: 15273396.

110. Dupuis MC, Zhang Z, Durkin K, Charlier C, Lekeux P, Georges M. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. Animal genetics. 2013;44(2):206-8. Epub 2012/05/16. doi: 10.1111/j.1365-2052.2012.02373.x. PubMed PMID: 22582820.

111. Wang Y, Tang Z, Sun Y, Wang H, Wang C, Yu S, et al. Analysis of genome-wide copy number variations in Chinese indigenous and western pig breeds by 60 K SNP genotyping arrays. PloS one. 2014;9(9):e106780. Epub 2014/09/10. doi: 10.1371/journal.pone.0106780. PubMed PMID: 25198154; PubMed Central PMCID: PMCPmc4157799.

112. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, et al. Analysis of copy number variations among diverse cattle breeds. Genome research.

2010;20(5):693-703. Epub 2010/03/10. doi: 10.1101/gr.105403.110. PubMed PMID: 20212021; PubMed Central PMCID: PMCPmc2860171.

113. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. Nature genetics. 2007;39(7 Suppl):S37-42. Epub 2007/09/05. doi: 10.1038/ng2080. PubMed PMID: 17597780.

114. Pankratz N, Dumitriu A, Hetrick KN, Sun M, Latourelle JC, Wilk JB, et al. Copy number variation in familial Parkinson disease. PloS one. 2011;6(8):e20988. Epub 2011/08/11. doi: 10.1371/journal.pone.0020988. PubMed PMID: 21829596; PubMed Central PMCID: PMCPmc3149037.

115. Clop A, Vidal O, Amills M. Copy number variation in the genomes of domestic animals. Animal genetics. 2012;43(5):503-17. Epub 2012/04/14. doi: 10.1111/j.1365-2052.2012.02317.x. PubMed PMID: 22497594.

116. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, et al. Copy number variation of individual cattle genomes using next-generation sequencing. Genome research. 2012;22(4):778-90. Epub 2012/02/04. doi: 10.1101/gr.133967.111. PubMed PMID: 22300768; PubMed Central PMCID: PMCPmc3317159.

117. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. Nat Meth. 2011;8(1):61-5. doi: 10.1038/nmeth.1527

118. Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature. 2015;517(7536):608-11. doi: 10.1038/nature13907

119. Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, et al. Reconstructing complex regions of genomes using long-read sequencing technology. Genome research. 2014;24(4):688-96. Epub 2014/01/15. doi: 10.1101/gr.168450.113. PubMed PMID: 24418700; PubMed Central PMCID: PMCPmc3975067.

120. Parsons JD. Miropeats: graphical DNA sequence comparisons. Computer applications in the biosciences : CABIOS. 1995;11(6):615-9. Epub 1995/12/01. PubMed PMID: 8808577.

121. Golosova O, Henderson R, Vaskin Y, Gabrielian A, Grekhov G, Nagarajan V, et al. Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. PeerJ. 2014;2:e644. Epub 2014/11/14. doi: 10.7717/peerj.644. PubMed PMID: 25392756; PubMed Central PMCID: PMCPmc4226638.

122. Okonechnikov K, Golosova O, Fursov M. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics (Oxford, England). 2012;28(8):1166-7. Epub 2012/03/01. doi: 10.1093/bioinformatics/bts091. PubMed PMID: 22368248.

123. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015;6:11. doi: 10.1186/s13100-015-0041-9. PubMed PMID: 26045719; PubMed Central PMCID: PMC4455052.

124. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 2006;7:474. Epub 2006/10/27. doi: 10.1186/1471-2105-7-474. PubMed PMID: 17064419; PubMed Central PMCID: PMCPMC1634758.

125. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS computational biology. 2013;9(8):e1003118. doi: 10.1371/journal.pcbi.1003118. PubMed PMID: 23950696; PubMed Central PMCID: PMC3738458.

126. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004;5(10):R80. doi: 10.1186/gb-2004-5-10-r80. PubMed PMID: 15461798; PubMed Central PMCID: PMC545600.

127. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

128. McVean G. What drives recombination hotspots to repeat DNA in humans? Philos Trans R Soc Lond B Biol Sci. 2010;365(1544):1213-8. Epub 2010/03/24. doi: 10.1098/rstb.2009.0299. PubMed PMID: 20308096; PubMed Central PMCID: PMCPMC2871820.

129. Stoppa-Lyonnet D, Duponchel C, Meo T, Laurent J, Carter PE, Arala-Chaves M, et al. Recombinational biases in the rearranged C1-inhibitor genes of hereditary angioedema patients. American journal of human genetics. 1991;49(5):1055-62.

Epub 1991/11/01. PubMed PMID: 1656734; PubMed Central PMCID: PMCPMC1683256.

130. Lehrman MA, Schneider WJ, Sudhof TC, Brown MS, Goldstein JL, Russell DW. Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. Science (New York, NY). 1985;227(4683):140-6. Epub 1985/01/11. PubMed PMID: 3155573; PubMed Central PMCID: PMCPMC4449727.

131. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, et al. Recent segmental duplications in the human genome. Science (New York, NY). 2002;297(5583):1003-7. Epub 2002/08/10. doi: 10.1126/science.1072047. PubMed PMID: 12169732.

132. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, et al. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature. 2005;437(7055):88-93. Epub 2005/09/02. doi: 10.1038/nature04000. PubMed PMID: 16136132.

133. Tuzun E, Bailey JA, Eichler EE. Recent segmental duplications in the working draft assembly of the brown Norway rat. Genome research. 2004;14(4):493-506. Epub 2004/04/03. doi: 10.1101/gr.1907504. PubMed PMID: 15059990; PubMed Central PMCID: PMCPmc383293.

134. She X, Cheng Z, Zollner S, Church DM, Eichler EE. Mouse segmental duplication and copy number variation. Nature genetics. 2008;40(7):909-14.

Epub 2008/05/27. doi: 10.1038/ng.172. PubMed PMID: 18500340; PubMed Central PMCID: PMCPmc2574762.

135. Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM. The genomic architecture of segmental duplications and associated copy number variants in dogs. Genome research. 2009;19(3):491-9. Epub 2009/01/09. doi: 10.1101/gr.084715.108. PubMed PMID: 19129542; PubMed Central PMCID: PMCPmc2661811.

136. Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, et al. Analysis of recent segmental duplications in the bovine genome. BMC genomics. 2009;10:571. Epub 2009/12/03. doi: 10.1186/1471-2164-10-571. PubMed PMID: 19951423; PubMed Central PMCID: PMCPmc2796684.

137. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental duplications and copy-number variation in the human genome. American journal of human genetics. 2005;77(1):78-88. Epub 2005/05/27. doi: 10.1086/431652. PubMed PMID: 15918152; PubMed Central PMCID: PMCPmc1226196.

138. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, et al. A high-resolution map of segmental DNA copy number variation in the mouse genome. PLoS genetics. 2007;3(1):e3. Epub 2007/01/09. doi: 10.1371/journal.pgen.0030003. PubMed PMID: 17206864; PubMed Central PMCID: PMCPmc1761046.

139. Meade KG, Cormican P, Narciandi F, Lloyd A, O'Farrelly C. Bovine beta-defensin gene family: opportunities to improve animal health? Physiological genomics. 2014;46(1):17-28. Epub 2013/11/14. doi: 10.1152/physiolgenomics.00085.2013. PubMed PMID: 24220329.

140. Scocchi M, Wang S, Zanetti M. Structural organization of the bovine cathelicidin gene family and identification of a novel member. FEBS letters. 1997;417(3):311-5. Epub 1997/12/31. PubMed PMID: 9409740.

141. Zanetti M. Cathelicidins, multifunctional peptides of the innate immunity. Journal of leukocyte biology. 2004;75(1):39-48. Epub 2003/09/10. doi: 10.1189/jlb.0403147. PubMed PMID: 12960280.

142. Walker AM, Roberts RM. Characterization of the bovine type I IFN locus: rearrangements, expansions, and novel subfamilies. BMC genomics. 2009;10:187. Epub 2009/04/28. doi: 10.1186/1471-2164-10-187. PubMed PMID: 19393062; PubMed Central PMCID: PMCPmc2680415.

143. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science (New York, NY). 2009;324(5926):522-8. Epub 2009/04/25. doi: 10.1126/science.1169588. PubMed PMID: 19390049; PubMed Central PMCID: PMCPmc2943200.

144. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annual review of genetics. 2005;39:121-52. Epub 2005/11/16. doi:

10.1146/annurev.genet.39.073003.112240. PubMed PMID: 16285855; PubMed Central PMCID: PMCPmc1464479.

145. Mak P, Szewczyk A, Mickowska B, Kicinska A, Dubin A. Effect of antimicrobial apomyoglobin 56-131 peptide on liposomes and planar lipid bilayer membrane. International journal of antimicrobial agents. 2001;17(2):137-42. Epub 2001/02/13. PubMed PMID: 11165118.

146. Huang H, Schroeder F, Estes MK, McPherson T, Ball JM. Interaction(s) of rotavirus non-structural protein 4 (NSP4) C-terminal peptides with model membranes. The Biochemical journal. 2004;380(Pt 3):723-33. Epub 2004/03/12. doi: 10.1042/bj20031789. PubMed PMID: 15012630; PubMed Central PMCID: PMCPmc1224213.

147. Chen PS, Toribara TY, Warner H. Microdetermination of Phosphorus. Analytical Chemistry. 1956;28(11):1756-8. doi: 10.1021/ac60119a033.

148. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature methods. 2012;9(7):671-5. Epub 2012/08/30. PubMed PMID: 22930834.

149. Yeaman MR, Yount NY. Mechanisms of antimicrobial peptide action and resistance. Pharmacological reviews. 2003;55(1):27-55. Epub 2003/03/05. doi: 10.1124/pr.55.1.2. PubMed PMID: 12615953.

150. Lee MO, Kim EH, Jang HJ, Park MN, Woo HJ, Han JY, et al. Effects of a single nucleotide polymorphism in the chicken NK-lysin gene on antimicrobial activity and cytotoxicity of cancer cells. Proceedings of the National Academy of

Sciences of the United States of America. 2012;109(30):12087-92. Epub 2012/07/12. doi: 10.1073/pnas.1209161109. PubMed PMID: 22783018; PubMed Central PMCID: PMCPMC3409721.

151. Chen Y, Guarnieri MT, Vasil AI, Vasil ML, Mant CT, Hodges RS. Role of peptide hydrophobicity in the mechanism of action of alpha-helical antimicrobial peptides. Antimicrob Agents Chemother. 2007;51(4):1398-406. Epub 2006/12/13. doi: 10.1128/aac.00925-06. PubMed PMID: 17158938; PubMed Central PMCID: PMCPMC1855469.

152. Griffin D. Economic impact associated with respiratory disease in beef cattle. Vet Clin North Am Food Anim Pract. 1997;13(3):367-77. Epub 1997/11/22. PubMed PMID: 9368983.

153. Snowder GD, Van Vleck LD, Cundiff LV, Bennett GL, Koohmaraie M, Dikeman ME. Bovine respiratory disease in feedlot cattle: phenotypic, environmental, and genetic correlations with growth, carcass, and longissimus muscle palatability traits. J Anim Sci. 2007;85(8):1885-92. Epub 2007/05/17. doi: 10.2527/jas.2007-0008. PubMed PMID: 17504959.

154. Garcia MD, Thallman RM, Wheeler TL, Shackelford SD, Casas E. Effect of bovine respiratory disease and overall pathogenic disease incidence on carcass traits. J Anim Sci. 2010;88(2):491-6. Epub 2009/11/10. doi: 10.2527/jas.2009-1874. PubMed PMID: 19897630.

155. Taylor JD, Fulton RW, Lehenbauer TW, Step DL, Confer AW. The epidemiology of bovine respiratory disease: What is the evidence for

predisposing factors? Can Vet J. 2010;51(10):1095-102. Epub 2011/01/05. PubMed PMID: 21197200; PubMed Central PMCID: PMCPMC2942046.

156. Muggli-Cockett NE, Cundiff LV, Gregory KE. Genetic analysis of bovine respiratory disease in beef calves during the first year of life. J Anim Sci. 1992;70(7):2013-9. Epub 1992/07/01. PubMed PMID: 1644673.

157. Neibergs H, Zanella R, Casas E, Snowder GD, Wenz J, Neibergs JS, et al. Loci on Bos taurus chromosome 2 and Bos taurus chromosome 26 are linked with bovine respiratory disease and associated with persistent infection of bovine viral diarrhea virus. J Anim Sci. 2011;89(4):907-15. Epub 2010/12/15. doi: 10.2527/jas.2010-3330. PubMed PMID: 21148784.

158. Cernicchiaro N, Renter DG, White BJ, Babcock AH, Fox JT. Associations between weather conditions during the first 45 days after feedlot arrival and daily respiratory disease risks in autumn-placed feeder cattle in the United States. J Anim Sci. 2012;90(4):1328-37. Epub 2011/12/08. doi: 10.2527/jas.2011-4657. PubMed PMID: 22147486.

159. Cernicchiaro N, White BJ, Renter DG, Babcock AH, Kelly L, Slattery R. Associations between the distance traveled from sale barns to commercial feedlots in the United States and overall performance, risk of respiratory disease, and cumulative mortality in feeder cattle during 1997 to 2009. J Anim Sci. 2012;90(6):1929-39. Epub 2012/01/17. doi: 10.2527/jas.2011-4599. PubMed PMID: 22247119.

160. Snowder GD, Van Vleck LD, Cundiff LV, Bennett GL. Bovine respiratory disease in feedlot cattle: environmental, genetic, and economic factors. J Anim Sci. 2006;84(8):1999-2008. Epub 2006/07/26. doi: 10.2527/jas.2006-046. PubMed PMID: 16864858.

161. Welsh RD, Dye LB, Payton ME, Confer AW. Isolation and antimicrobial susceptibilities of bacterial pathogens from bovine pneumonia: 1994--2002. J Vet Diagn Invest. 2004;16(5):426-31. Epub 2004/10/06. PubMed PMID: 15460326.

162. Rice JA, Carrasco-Medina L, Hodgins DC, Shewen PE. Mannheimia haemolytica and bovine respiratory disease. Anim Health Res Rev. 2007;8(2):117-28. Epub 2008/01/26. doi: 10.1017/s1466252307001375. PubMed PMID: 18218156.

163. Shahriar FM, Clark EG, Janzen E, West K, Wobeser G. Coinfection with bovine viral diarrhea virus and Mycoplasma bovis in feedlot cattle with chronic pneumonia. Can Vet J. 2002;43(11):863-8. Epub 2002/12/25. PubMed PMID: 12497963; PubMed Central PMCID: PMCPMC339759.

164. Klima CL, Zaheer R, Cook SR, Booker CW, Hendrick S, Alexander TW, et al. Pathogens of bovine respiratory disease in North American feedlots conferring multidrug resistance via integrative conjugative elements. J Clin Microbiol. 2014;52(2):438-48. Epub 2014/01/31. doi: 10.1128/jcm.02485-13. PubMed PMID: 24478472; PubMed Central PMCID: PMCPMC3911356.

165. Gagea MI, Bateman KG, van Dreumel T, McEwen BJ, Carman S, Archambault M, et al. Diseases and pathogens associated with mortality in Ontario beef

feedlots. J Vet Diagn Invest. 2006;18(1):18-28. Epub 2006/03/29. PubMed PMID: 16566254.

166. Ellis JA. Bovine parainfluenza-3 virus. Vet Clin North Am Food Anim Pract. 2010;26(3):575-93. Epub 2010/11/09. doi: 10.1016/j.cvfa.2010.08.002. PubMed PMID: 21056802.

167. Neibergs HL, Seabury CM, Wojtowicz AJ, Wang Z, Scraggs E, Kiser J, et al. Susceptibility loci revealed for bovine respiratory disease complex in pre-weaned holstein calves. BMC genomics. 2014;15(1):1164. Epub 2014/12/24. doi: 10.1186/1471-2164-15-1164. PubMed PMID: 25534905.

168. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 2009;10(1):57-63. PubMed PMID: 19015660.

169. Tizioto PC, Kim J, Seabury CM, Schnabel RD, Gershwin LJ, Van Eenennaam AL, et al. Immunological Response to Single Pathogen Challenge with Agents of the Bovine Respiratory Disease Complex: An RNA-Sequence Analysis of the Bronchial Lymph Node Transcriptome. PloS one. 2015;10(6):e0131459. Epub 2015/06/30. doi: 10.1371/journal.pone.0131459. PubMed PMID: 26121276; PubMed Central PMCID: PMCPMC4484807.

170. Gershwin LJ, Van Eenennaam AL, Anderson ML, McEligot HA, Shao MX, Toaff-Rosenstein R, et al. Single Pathogen Challenge with Agents of the Bovine Respiratory Disease Complex. PloS one. 2015;10(11):e0142479. Epub

2015/11/17. doi: 10.1371/journal.pone.0142479. PubMed PMID: 26571015; PubMed Central PMCID: PMCPMC4646450.

171. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 9(4):357-9. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286; PubMed Central PMCID: PMCPmc3322381.

172. Glass EJ, Baxter R, Leach RJ, Jann OC. Genes controlling vaccine responses and disease resistance to respiratory viral pathogens in cattle. Veterinary immunology and immunopathology. 2012;148(1-2):90-9. Epub 2011/05/31. doi: 10.1016/j.vetimm.2011.05.009. PubMed PMID: 21621277; PubMed Central PMCID: PMCPMC3413884.

173. Casas E, Hessman BE, Keele JW, Ridpath JF. A genome-wide association study for the incidence of persistent bovine viral diarrhea virus infection in cattle. Animal genetics. 2015;46(1):8-15. Epub 2014/11/14. doi: 10.1111/age.12239. PubMed PMID: 25394207.

174. Bermingham ML, Bishop SC, Woolliams JA, Pong-Wong R, Allen AR, McBride SH, et al. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. Heredity (Edinb). 2014;112(5):543-51. Epub 2014/02/06. doi: 10.1038/hdy.2013.137. PubMed PMID: 24496092; PubMed Central PMCID: PMCPMC3998787.

175. Silen JL, McGrath CN, Smith KR, Agard DA. Molecular analysis of the gene encoding alpha-lytic protease: evidence for a preproenzyme. Gene. 1988;69(2):237-44. Epub 1988/09/30. PubMed PMID: 3234766.

176. Ikemura H, Takagi H, Inouye M. Requirement of pro-sequence for the production of active subtilisin E in Escherichia coli. J Biol Chem. 1987;262(16):7859-64. Epub 1987/06/05. PubMed PMID: 3108260.

177. Winther JR, Sorensen P. Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(20):9330-4. Epub 1991/10/15. PubMed PMID: 1924396; PubMed Central PMCID: PMCPMC52708.

178. Fabre E, Nicaud JM, Lopez MC, Gaillardin C. Role of the proregion in the production and secretion of the Yarrowia lipolytica alkaline extracellular protease. J Biol Chem. 1991;266(6):3782-90. Epub 1991/02/25. PubMed PMID: 1995632.

179. Valls LA, Winther JR, Stevens TH. Yeast carboxypeptidase Y vacuolar targeting signal is defined by four propeptide amino acids. J Cell Biol. 1990;111(2):361-8. Epub 1990/08/01. PubMed PMID: 2199455; PubMed Central PMCID: PMCPMC2116205.

180. Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. Nature reviews Genetics. 2007;8(10):762-75. Epub 2007/09/12. doi: 10.1038/nrg2193. PubMed PMID: 17846636.

181. Teshima KM, Innan H. The effect of gene conversion on the divergence between duplicated genes. Genetics. 2004;166(3):1553-60. Epub 2004/04/15. PubMed PMID: 15082568; PubMed Central PMCID: PMCPMC1470786.

182. Wen Y, Irwin DM. Mosaic evolution of ruminant stomach lysozyme genes. Mol Phylogenet Evol. 1999;13(3):474-82. Epub 2000/01/06. doi: 10.1006/mpev.1999.0651. PubMed PMID: 10620405.

183. Koonin EV. Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet. 2005;39:309-38. Epub 2005/11/16. doi: 10.1146/annurev.genet.39.073003.114725. PubMed PMID: 16285863.

184. Mansai SP, Innan H. The power of the methods for detecting interlocus gene conversion. Genetics. 2010;184(2):517-27. Epub 2009/12/02. doi: 10.1534/genetics.109.111161. PubMed PMID: 19948889; PubMed Central PMCID: PMCPMC2828729.

185. Stephens JC. Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. Mol Biol Evol. 1985;2(6):539-56. Epub 1985/11/01. PubMed PMID: 3870876.

186. Betran E, Rozas J, Navarro A, Barbadilla A. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. Genetics. 1997;146(1):89-99. Epub 1997/05/01. PubMed PMID: 9136003; PubMed Central PMCID: PMCPMC1207963.

187. Innan H. A method for estimating the mutation, gene conversion and recombination parameters in small multigene families. Genetics.

2002;161(2):865-72. Epub 2002/06/20. PubMed PMID: 12072480; PubMed Central PMCID: PMCPMC1462133.

188. Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ. Complex SNP-related sequence variation in segmental genome duplications. Nature genetics. 2004;36(8):861-6. Epub 2004/07/13. doi: 10.1038/ng1401. PubMed PMID: 15247918.

189. Hallast P, Nagirnaja L, Margus T, Laan M. Segmental duplications and gene conversion: Human luteinizing hormone/chorionic gonadotropin beta gene cluster. Genome research. 2005;15(11):1535-46. Epub 2005/10/28. doi: 10.1101/gr.4270505. PubMed PMID: 16251463; PubMed Central PMCID: PMCPMC1310641.

190. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9. Epub 2013/10/18. doi: 10.1093/molbev/mst197. PubMed PMID: 24132122; PubMed Central PMCID: PMCPMC3840312.

APPENDIX

PHOSPHORUS ASSAY PROTOCOL

Reagents:

Sulfuric acid (H_2SO_4) (SIGMA-ALDAICH, cat # 339741-100ML)

Ammonium molybdate tetrahydrate (SIGMA-ALDAICH, cat #09878-25G)

L-Ascorbic acid (SIGMA-ALDAICH, cat # A5960-25G)

0.65 mM Phosphorus standard solution (SIGMA-ALDAICH, cat # P3869-25ML)

Hydrogen peroxide (SIGMA-ALDAICH, cat # 216763-100ML)

Procedure:

Prepare the solutions

1, 8.9 N H_2SO_4 solution: 247 ml of H_2SO_4 is slowly mixed with 753 ml of deionized water. This solution can be stored at room temperature.

2, 10% Ascorbic acid solution. 10 g of ascorbic acid is dissolved into 100 ml of deionized water. This solution should be stored in an amber screw-cap bottle at 4℃ before use.

3, 2.5% Ammonium molybdate tetrahydrate solution. 2.5 g of ammonium molybdate tetrahydrate is dissolved into 100 ml of deionized water. This solution should be stored in an amber screw-cap bottle at 4℃ before use.

Prepare the sample tubes

Sample containing about 0.1 μmoles phosphorus is placed into the bottom of a tube, and the solvent is gently removed from the tubes with N2. Three technical replications are performed.

Prepare the standard tubes

Six standard tubes containing known quantities of phosphorus are prepared by placing the phosphorus standard solution into six separate tubes: 0 μmoles (0 μl) blank, 0.0325 μmoles (50 μl), 0.065 μmoles (100 μl), 0.114 μmoles (175 μl), 0.163 μmoles (250 μl), and 0.228 μmoles (350 μl).

Digestion of organic sample to inorganic phosphate

1, An aliquot of 0.45 ml 8.9 N H2SO4 is added to each of the standard tubes and sample tubes.

2, All the tubes are heated in an aluminum block in the hood at 200-215 ℃ for 25 minutes.

3, Tubes are removed from the block and cooled at room temperature for 5 minutes before added with an aliquot of 150 μl H2O2 to the bottom of each tube.

4, Tubes are heated for an additional 30 minutes in the same aluminum block in the hood at 200-215 ℃. The samples should be colorless at this point. If any brown color persists, another aliquot of 50 μl H2O2 can be added to all the cooled tubes and the tubes continue to be heated for another 15 minutes.

5, An aliquot of 3.9 ml deionized water is added to each cooled tube.

6, An aliquot of 0.5 ml ammonium molybdate tetrahydrate solution is added to each tube, and tube is vortex for 5 times.

7, An aliquot of 0.5 ml 10% ascorbic acid solution is added to each tube, and tube is vortex for 5 times.

8, All tubes are heated at 100 ℃ for 7 minutes in an aluminum block in the hood.

9, Tubes are cooled to the ambient temperature.

Spectrophotometric analysis of samples

1, Spectrophotometer is standardized using the 0 μmoles standard solution.

2, The absorbance of each of the five standards is measured at 820 nm.

3, The absorbance of each of the samples is measured at 820 nm.

4, The calibration curve is determined by using the standards, and concentration of phosphorus in the samples are calculated based on the calibration curve.