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

Copyright 2016 Junfeng Chen
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

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMP</td>
<td>Antimicrobial Peptide</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy Number Variation</td>
</tr>
<tr>
<td>SD</td>
<td>Segmental Duplication</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
</tr>
<tr>
<td>CD</td>
<td>Circular Dichroism</td>
</tr>
<tr>
<td>INDEL</td>
<td>Insertion/Deletion</td>
</tr>
<tr>
<td>BRD</td>
<td>Bovine Respiratory Disease</td>
</tr>
<tr>
<td>BVDV</td>
<td>Bovine Viral Diarrhea Virus</td>
</tr>
<tr>
<td>BRSV</td>
<td>Bovine Respiratory Syncytial Virus</td>
</tr>
<tr>
<td>IBR</td>
<td>Infectious Bovine Rhinotracheitis</td>
</tr>
</tbody>
</table>
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>NOMENCLATURE</td>
<td>v</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
</tbody>
</table>

### CHAPTER I INTRODUCTION

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

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

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
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-lysins 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.

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 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.

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).

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.

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.

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.
LIST OF TABLES

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
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 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. Peptidoglycan 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, peptidoglycan is not found in eukaryotic cells. Therefore, compounds that can inhibit the synthesis of peptidoglycan 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-lysins**

Human granulysin and pig NK-lysins 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-lysins 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-lysins* 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-lysins 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 those 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 6th 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 \textit{RAD50}, \textit{XRS2} and \textit{MRE11} can significantly increase the frequency of extensive deletions at the broken ends in \textit{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.
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.

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Accession</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA-1</td>
<td>XM_005192449</td>
<td>NK2A</td>
</tr>
<tr>
<td>mRNA-2</td>
<td>XM_005192450</td>
<td>NK2A</td>
</tr>
<tr>
<td>mRNA-3</td>
<td>BC114176</td>
<td>NK2B</td>
</tr>
<tr>
<td>mRNA-4</td>
<td>AY245798</td>
<td>NK2A</td>
</tr>
<tr>
<td>mRNA-5</td>
<td>BC114178</td>
<td>NK2C</td>
</tr>
<tr>
<td>mRNA-6</td>
<td>AY245799</td>
<td>NK2C</td>
</tr>
<tr>
<td>mRNA-7</td>
<td>NM_001046578</td>
<td>NK1</td>
</tr>
</tbody>
</table>
Materials and methods

Confirmation of NK-lys in inclusion in the BAC clones

Two overlapping BAC clones covering the bovine NK-lys in 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-lys in inclusion in the BAC clones was conducted with the Bo-lysin primers (Table 2) which were designed within the conserved region of NK-lys in 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-lys in primer, 0.8 mM dNTPs, 1µL (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 °C, 15 s at 62 °C, 20 s at 68 °C, and a final extension of 5 min at 68 °C. 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

<table>
<thead>
<tr>
<th>Primers</th>
<th>Forward (5' to 3')</th>
<th>Reverse (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bo-lysin</td>
<td>ACCCAGCACTCCCACTG</td>
<td>ACATACCTGGCTTGCTTTTG</td>
</tr>
<tr>
<td>JP-1</td>
<td>CTAAGTGCCGGAATGGTTGT</td>
<td>CAGGGTCTTCTCCTCGACG</td>
</tr>
<tr>
<td>JP-2</td>
<td>GAAATGCTCTCACAGCAACA</td>
<td>AATAGCAATGAAATGATGATGGT</td>
</tr>
<tr>
<td>JP-3</td>
<td>AAAATGCTCTCACAGCAATGAA</td>
<td>AATAGCAATGAAATGATGATGCTG</td>
</tr>
<tr>
<td>JP-4</td>
<td>GATATGCTCTCCACAACCAGTCAG</td>
<td>GAATTGCTGAGGCTGAGAAGG</td>
</tr>
<tr>
<td>BP-1</td>
<td>GCCTGCCTCATGGAGTTTA</td>
<td>TGGCACAGGTAATGGGATAA</td>
</tr>
</tbody>
</table>

**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.
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 \textit{NK-lysin} gene, and \textit{NK1} was 41.8 Kb downstream from the \textit{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 \textit{NK-lysin} region, and demonstrated that four \textit{NK-lysin} genes are located in this region on cattle chromosome 11. Complete genomic sequences of four \textit{NK-lysin} genes were compared to determine genetic organization and structure (Fig. 6). All four bovine \textit{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 \textit{NK1} were larger than introns from the other genes, accounting for a larger genomic size of \textit{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-lysins 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).

**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 \textit{NK-lysins} and \textit{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.

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

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

<table>
<thead>
<tr>
<th>Class</th>
<th>Superfamily</th>
<th>Family</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLT1F</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MLT1F1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>MLT1J</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>MLT1J1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MLT1J2</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>MLT1M</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>LTR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gypsy</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>LTR88b</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>DNA Transposon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA transposon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X25_DNA</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>hAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Charlie13a</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>MER91B</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>UCON52</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Mariner/Tc1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MER47B</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>TIGGER5_B</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TIGGER5A</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 8. Repeat element analysis within bovine \textit{NK-lys} 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 predominated 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,
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.

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 (Invitorgen). 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 ºC for 2 min, 95 ºC for 10 min, followed by 95 ºC for 15 s and 60 ºC 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

<table>
<thead>
<tr>
<th>Primer</th>
<th>Forward (5' to 3')</th>
<th>Reverse (5' to 3')</th>
<th>Probe (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex-NK1</td>
<td>CCAGCAAGAATGTC</td>
<td>GTCCTTAGAGATGCAG</td>
<td>CTTTGCAACCAG</td>
</tr>
<tr>
<td>NK1</td>
<td>ATCATCC</td>
<td>TGAGATAC</td>
<td>ATGA</td>
</tr>
<tr>
<td>Ex-NK2A</td>
<td>AGGAGAAGAGCTG</td>
<td>GCTGATCTCCTCAACTTG</td>
<td>TCCTTGTGGGATG</td>
</tr>
<tr>
<td>NK2A</td>
<td>GGCCTAC</td>
<td>TCC</td>
<td>ATAATG</td>
</tr>
<tr>
<td>Ex-NK2B</td>
<td>GAGAATACCGTCAT</td>
<td>TTGCACAGACCTTTTCAG</td>
<td>TCCAAGGTGTGC</td>
</tr>
<tr>
<td>NK2B</td>
<td>CGAGGC</td>
<td>CG</td>
<td>GGC</td>
</tr>
<tr>
<td>Ex-NK2C</td>
<td>AATTTCTCCGTACC</td>
<td>ATGAAACCTACTGGCTT</td>
<td>AGGACATCGTAG</td>
</tr>
<tr>
<td>NK2C</td>
<td>ATCGCT</td>
<td>GCTT</td>
<td>CTGG</td>
</tr>
</tbody>
</table>

### 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 °C 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

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
<th>Length (aa)</th>
<th>Charge</th>
<th>Net Charge (pH 7)</th>
<th>Hydrophobicity (pH 6.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NK1</td>
<td>VIIHVTSTKCSVCSKMLWISILC</td>
<td>30</td>
<td>+6</td>
<td>4.93</td>
<td>37.1</td>
</tr>
<tr>
<td>NK2A</td>
<td>TVIEVASKMCSDKRLLKGL</td>
<td>30</td>
<td>+8</td>
<td>7.82</td>
<td>31.43</td>
</tr>
<tr>
<td>NK2B</td>
<td>TVIEAASKVCGKMGLKL</td>
<td>30</td>
<td>+7</td>
<td>6.82</td>
<td>26.1</td>
</tr>
<tr>
<td>NK2C</td>
<td>TVIEEASKVCSDKRLLKGLC</td>
<td>30</td>
<td>+6</td>
<td>5.82</td>
<td>30.57</td>
</tr>
</tbody>
</table>

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 °C 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 N₂ 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 °C.
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°C with aeration were sub-cultured to fresh LB at a ratio of 1:50 and grown at 37°C 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 $3 \times 10^6$ 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 °C for 2 h, and then plated onto LB agar plates. Colonies of the surviving bacteria were manually counted after overnight incubation at 37 °C. 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) ($3 \times 10^8$ CFU/mL) were incubated with 10 µL buffer or buffer plus 5 µM NK1 peptide at 37 °C 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 °C, 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-lysins 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 NKI 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-lysins 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-lysins 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 °C 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 °C 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 °C. 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 °C. 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-lysins_ 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.
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
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-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/](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/](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 \( NK\text{-}lysin \) 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 \( NK2A\text{-}C \) by MEGA 6.0 [190], pig and horse \( NK\text{-}lysin \) 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.
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**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Forward (5' to 3')</th>
<th>Reverse (5' to 3')</th>
<th>Amplicon size</th>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>NK1</td>
<td>ATGACCTCTTG</td>
<td>CAGATGGCCTG</td>
<td>395 bp</td>
<td>(NK1) coding region amplification</td>
</tr>
<tr>
<td>CDS</td>
<td>GGCTGTTCCT</td>
<td>AGGTGTCTTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NK1</td>
<td>TGACTTCTCCTC</td>
<td>AGGGAATTCTC</td>
<td>873 bp</td>
<td>Identification of genomic region of 9-bp INDEL</td>
</tr>
<tr>
<td>-9bp</td>
<td>TCCCTCTGTG</td>
<td>CAAAGTCCATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bollysin</td>
<td>ACCCAGCACTC</td>
<td>ACATACTGGGC</td>
<td>293 bp</td>
<td>Homozygotes analysis</td>
</tr>
<tr>
<td></td>
<td>CCACTG</td>
<td>TTGCTTTTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gs-NK2</td>
<td>TGGTCCTTTGGT</td>
<td>TTTCCAGCTTG</td>
<td>893 bp</td>
<td>Gene conversion test</td>
</tr>
<tr>
<td>A</td>
<td>GATGATAATG</td>
<td>ATGTCTGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gs-NK2</td>
<td>CTGTTCATGCT</td>
<td>TTGCCAGAGACC</td>
<td>1479 bp</td>
<td>(NK2B) deletion test</td>
</tr>
<tr>
<td>B</td>
<td>GTTTCTTCCAT</td>
<td>TTTCAGCG</td>
<td></td>
<td>Gene conversion test</td>
</tr>
<tr>
<td>Gs-NK2</td>
<td>CTACGCTGTGG</td>
<td>GTTTTTCCAGCT</td>
<td>891 bp</td>
<td>Gene conversion test</td>
</tr>
<tr>
<td>C</td>
<td>TTCTTGTCC</td>
<td>ACGATGTCTCT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7. Individuals used in the genetic study of bovine *NK-lysin* gene family

<table>
<thead>
<tr>
<th>Individual</th>
<th>Breed</th>
<th>Utilization</th>
<th>9-bp INDEL</th>
<th>SNP (BovineHD1100014441)</th>
<th>NK2B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domino</td>
<td>Hereford</td>
<td></td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Brahman</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Brahman</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Charolais</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Charolais</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Gelbvieh</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2822</td>
<td>Holstein</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3850</td>
<td>Holstein</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4171</td>
<td>Holstein</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4356</td>
<td>Holstein</td>
<td>9-bp INDEL test</td>
<td>+/-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2527</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3069</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3084</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3153</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3489</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3532</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3540</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3611</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3612</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3803</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3856</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>4454</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>2796</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3443</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>3022</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
Table 7. Continued

<table>
<thead>
<tr>
<th>Individual</th>
<th>Breed</th>
<th>Utilization</th>
<th>9-bp INDEL</th>
<th>SNP (BovineHD1100014441)</th>
<th>NK2B</th>
</tr>
</thead>
<tbody>
<tr>
<td>3318</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>4378</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>2822</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>C</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>4171</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>C</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>4494</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>T</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>3850</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>C</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>4356</td>
<td>Holstein</td>
<td>NK2B test</td>
<td>C</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>Brahman</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Brahman</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Charolais</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Charolais</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>Charolais</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Gelbvieh</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>Red Angus</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>Red Angus</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>Red Angus</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3850</td>
<td>Holstein</td>
<td>Gene conversion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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-lys in 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.
Table 9. Number of sequenced clones and different sequences obtained from each individual in the analysis of homozygous cattle.

<table>
<thead>
<tr>
<th>Animal ID</th>
<th>No. sequenced clones</th>
<th>No. different clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>2527</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>2796</td>
<td>30</td>
<td>3</td>
</tr>
<tr>
<td>2822</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td>3850</td>
<td>30</td>
<td>2</td>
</tr>
</tbody>
</table>

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 16\textsuperscript{th} 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 16\textsuperscript{th} 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 16\textsuperscript{th} 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 antimicrobialities. 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: GTGAAAACTATGGTTC 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 antimicrobiality 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-lysin*. 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-lysin 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-lysin, 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


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


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


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.


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


    genomic architecture of segmental duplications and associated copy number
    10.1101/gr.084715.108. PubMed PMID: 19129542; PubMed Central PMCID:
    PMCPmc2661811.

    PubMed Central PMCID: PMCPmc2796684.

    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.

    high-resolution map of segmental DNA copy number variation in the mouse
    10.1371/journal.pgen.0030003. PubMed PMID: 17206864; PubMed Central
    PMCID: PMCPmc1761046.


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


135


APPENDIX

PHOSPHORUS ASSAY PROTOCOL

Reagents:

Sulfuric acid (H2SO4) (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 H2SO4 solution: 247 ml of H2SO4 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°C 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°C before use.

Prepare the sample tubes
Sample containing about 0.1 µ moles 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 µ moles (0µ l) blank, 0.0325 µ moles (50 µ l), 0.065 µ moles (100 µ l), 0.114 µ moles (175 µ l), 0.163 µ moles (250 µ l), and 0.228 µ moles (350 µ l).

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°C 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 µ l 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°C. The samples should be colorless at this point. If any brown color persists, another aliquot of 50 µ l 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°C 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.