GENETIC ANALYSES OF BOVINE CARD15, A PUTATIVE DISEASE RESISTANCE GENE

A Dissertation

by

KRISTEN HAWKINS TAYLOR

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

DOCTOR OF PHILOSOPHY

May 2004

Major Subject: Genetics

GENETIC ANALYSES OF BOVINE CARD15, A PUTATIVE DISEASE RESISTANCE GENE

A Dissertation

by

KRISTEN HAWKINS TAYLOR

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

James E. Womack (Chair of Committee)	L. Garry Adams (Member)
James N. Derr (Member)	Allen J. Roussel (Member)
Ann B. Kier (Head of Department)	Geoffrey M. Kapler (Chair of Genetics Faculty)

May 2004

Major Subject: Genetics

ABSTRACT

Genetic Analyses of Bovine CARD15, a Putative Disease
Resistance Gene. (May 2004)
Kristen Hawkins Taylor, B.S., Lamar University
Chair of Advisory Committee: Dr. James E. Womack

Through a binding partner the CARD15 gene activates NF-kB, a molecule with a role in the initiation of the inflammatory immune response. The gene is highly conserved in both structure and function in human and mouse and has recently been implicated as a disease resistance gene in Crohn's disease and Blau Syndrome in human. The gene's relationship to disease and its conservation between species suggests that it may also have a conserved role in bovine disease resistance. To elucidate the potential role of bovine CARD15 in disease resistance, the gene was characterized in cattle. Bovine CARD15 is located 4.2 cR_{5000} telomeric to ADCY7 on chromosome 18. It spans ~30 kb and is comprised of 12 exons, 11 of which are coding. Bovine CARD15 is expressed in many tissues, but is most abundant in peripheral blood leukocytes. An extensive comparative analysis between the bovine, mouse and human CARD15 genes revealed high levels of inter-species conservation in sequence, genomic structure and protein domains. Conserved putative regulatory motifs were identified in the three species comparison of the 5'UTR, 3'UTR and the intronic sequences flanking exons. Additionally, diverse regulatory motifs were identified in each of the species indicating an evolutionary divergence in the mechanisms of regulation of gene expression. To assess the extent of genetic diversity within bovine CARD15, 41 individuals from nine breeds representing two subspecies were sequenced and screened for polymorphisms. Thirty-six single nucleotide polymorphisms (SNPs) were identified including 26 within the gene transcript. Haplotypes were estimated for each individual and parsimonious SNP sets were identified with which the multi-locus *Bos taurus* and *Bos indicus* haplotypes may be reconstructed. There was a significantly higher rate of substitutions within *Bos indicus* than in *Bos taurus*. A significantly higher rate of nonsynonymous to synonymous substitutions was found in *Bos taurus* indicating that positive Darwinian selection is acting on the gene within this subspecies. Association analyses were performed between these SNP loci and haplotypes with Johne's disease. No overwhelming evidence for a simple causal relationship was detected. Assays are provided to screen populations of cattle for variation in the CARD15 gene.

DEDICATION

This work is dedicated with love and admiration to the memory of my grandfather, Jack “Papa”Anderson. He was a devoted enthusiast throughout this journey and the journey of my life as a whole. I am extremely blessed to have had him in my life for 37 years and his loss has been a terrible misfortune. It is with great pride and joy that his memory will live on through me.

ACKNOWLEDGMENTS

First and foremost, I want to acknowledge the guidance and support of my advisor, Dr. Jim Womack. When I was without a home, he took me into his lab and provided me with the resources, guidance and support that ultimately led to the completion of this study. For their assistance in procuring DNA samples and in providing me with instruction into the nuances of Johne's disease, I gratefully acknowledge the support of Drs. Allen Roussel and Garry Adams. For his support, assistance in providing me with breed panel DNA samples and in assisting me with BAC work, I am grateful to Dr. Jim Derr. To all of the members of my Graduate Committee, I am very appreciative of your counsel, guidance and support.

I also appreciate the support and assistance of several faculty members at the University of Missouri-Columbia. Dr. Jerry Taylor allowed me to make his lab my home for the last year of my program and provided the tissue samples for the Q-PCR experiments. Unfortunately, he personally made me kill the cows from which the tissue samples were derived! Dr. Bob Schnabel was always available for technical guidance and advice. I have appreciated his fresh perspective and friendship. Dr. Ed Rucker helped me to develop the primers used in the Q-PCR experiments and Dr. Shahnawaz Khan helped me to develop the assay for the ABI 7700.

During my year at the University of Missouri, I could not have survived without the assistance of Jan Elliott in Dr. Womack's lab who helped solve each administrative crisis as it arose at Texas A&M University. Jan helped me out innumerable times and I am indebted to her.

I want to thank Dr. Clare Gill and Collette Abbey for providing me with bovine BACs harboring the CARD15 gene. I also appreciate Dr. Tim Smith from the USDA MARC for expeditiously sending me bovine CARD15 EST clones at short notice.

I am deeply grateful for the opportunity I had to work with a number of gifted and talented fellow graduate students while at Texas A&M University. In particular, Susan Tanksley, Dannialle Clayton, Stephen White, Ashley Gustafson and Chris Seabury assisted me with a number of technical aspects of this study; I learned much from their guidance. My friends and fellow graduate students at Texas A&M University: Shanna Moore, Greg and Jen Peters, Lori and Brian Adams, Archana Dhasarathy and Avni Santani helped me to survive when it seemed that the whole world had fallen apart.

My closest friends and best supporters are Jenn and Larry Whitfield, who let me stay at their home for 6 months after I joined Dr. Womack's lab. Not only were they my only connection to the world outside of academia, but they helped me to keep it all together for the 18 months that my husband and I were forced to live apart.

Finally none of this would have been possible without the support of my husband and family. My family helped me financially, emotionally and spiritually when I decided to return to graduate school and have been unfailing during this phase of my life. My husband fully supported my goal of returning to graduate school even though this meant that we would be forced to live apart. After 18 months of living apart, he then suffered a year of our living together in which I had to complete my lab work and write this dissertation.

Thanks to you all! I hope that I have made you proud of me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Infectious diseases can have devastating effects on the livestock sector in terms of production and economic loss. The cost of treatment and veterinary care of infected animals can be an onerous financial burden. There is considerable evidence that resistance to infectious disease in animals has a genetic basis and that additive genetic variation exists among animals in their response to various infectious challenges (Bishop and MacKenzie 2003). Therefore, breeding for enhanced disease resistance appears to be an alternative to the exorbitant cost of treatment and veterinary care and has the desirable consequence of enhancing animal welfare and well-being while simultaneously enhancing productivity and enterprise profitability. Natural selection may not have a dramatic impact on increasing the resistance of animals to diseases such as "shipping fever" or Johne's disease since the relative fitness of susceptible and resistant animals may be similar. Shipping fever is a transient infection and Johne's disease may not be expressed until late in an animal's life allowing similar reproductive rates and progeny survival among susceptible and resistant animals.

The efficiency of breeding animals possessing innate disease resistance might be greatly enhanced if the genes and mutations responsible for variation in disease resistance could be identified. The identification of these genes and mutations underlying resistance and susceptibility will also provide new tools to facilitate research into the mechanisms of infection possibly leading to additional pharmacological and management approaches for the control of disease transmission. Once the causal mutations have been identified, simple and relatively inexpensive DNA tests can be developed to screen populations of animals for presence of the desirable alleles. Consequently, considerable research effort has been directed towards identifying disease resistance genes in livestock. Genes that have previously been identified in human and mouse to be associated with disease resistance phenotypes and that possess conserved sequence homology and conserved function in both species are excellent candidates for

---

This dissertation follows the style of Mammalian Genome.

investigation in cattle.

Toll-like receptors (TLRs) are a relatively new family of proteins identified in mammals that have unequivocally been shown to participate in the recognition of microbial components and the activation of innate immunity, leading to the development of antigen-specific immune responses (Aderem and Ulevitch 2000, Zhang and Ghosh 2001). Members of this family are composed of three domains: a cytoplasmic signal transduction toll/interleukin-1 receptor domain (TIR), a short transmembrane sequence and extracellular leucine rich repeats (LRR) (Rock et al. 1998). They recognize, via their extracellular LRR, pathogen associated molecular patterns and initiate immune responses via their cytoplasmic domain (Janeway 1989). Each of these membrane-associated TLRs appears to recognize a specific fraction of microbial components (Akira et al. 2001). Two TLRs, TLR2 and TLR4, have been extensively studied in human and mouse and have been shown to play a role in resistance to certain pathogens (O'Brien et al. 1980; Takeuchi et al. 1999). Since these genes are conserved in both structure and function between human and mouse, they were recently examined as candidate disease resistance genes in cattle (White et al. 2003).

A second family of proteins that have been implicated in the recognition of pathogen components share a nucleotide oligomerization domain and have thus been termed the "Nods." Unlike the TLRs, these proteins appear to function within the cell (Inohara et al. 2001). Furthermore, these proteins display a striking similarity to a class of disease resistance (R) gene protein products found in plants which are known to aid in the plant's defense against multiple pathogens (Dixon et al. 2000). In particular, the product of the caspase recruitment domain 15 gene (CARD15), like the plant R gene products, contains an amino-terminal effector domain. In the case of CARD15, this domain is comprised of two caspase recruitment domains (CARDs) and it is the only gene identified to date to possess two such domains (Ogura et al. 2001b). CARD15 and the plant R proteins also have centrally located nucleotide binding/oligomerization domains (NODs) which mediate both nucleotide binding and oligomerization (Hu et al. 1998; Yang et al. 1998). In the case of CARD15, the NOD is believed to comprise a "self-oligomerization" motif that promotes activation of effector molecules through self-association and the induced proximity of binding partners (Inohara and Nunez 2001). Finally, both CARD15 and plant R proteins have C-terminal domains containing leucine rich repeats (Bertin et al. 1999; Inohara et al. 1999; Ogura et al. 2001b). In plants, genetic variation in the LRR determines pathogen specificity (Parniske et al. 1997), while variation in the LRR can result in unresponsiveness to particular pathogens (Parniske et al. 1997; Ellis et al. 1999; Dixon et al. 2000). Likewise, the LRR of CARD15 are believed to serve as a sensor for intracellular ligands (Ogura et al. 2001a) such as pathogenic bacteria.

The human CARD15 mRNA has two in-frame translation initiation sites separated by 81 nucleotides resulting in the encoding of proteins of 1040 and 1013 amino acids (Ogura et al. 2001b), although the 1013 amino acid protein seems dominant (Lesage et al. 2002). The full-length mouse CARD15 transcript also contains 2 potential translation initiation sites and encodes proteins of 1020 and 1013 amino acids (Iwanaga et al. 2003). In both species, CARD15 comprises 12 coding exons and domain analyses have revealed that human and mouse CARD15 both contain 2 N-terminal CARD domains, one centrally located nucleotide binding domain and 10 tandem LRR in the C-terminus (Ogura et al. 2001b; Iwanaga et al. 2003; Ogura et al. 2003). Alignment of human and mouse CARD15 reveals that 78% of the amino acids are identical (Iwanaga et al. 2003).

Over-expression of both human and mouse CARD15 in HEK293T cells induces NF-kB activation (Ogura et al. 2001b; Iwanaga et al. 2003; Ogura et al. 2003) and the CARD15 gene has been shown to confer responsiveness to lipopolysaccharide (LPS) and peptidoglycan (PGN) in both human and mouse (Ogura et al. 2001b ; Ogura et al. 2003). More specifically, it has recently been demonstrated that the muramyl dipeptide moieties from PGN are the key components selected by the LRR of the CARD15 gene (Girardin et al. 2003; Inohara et al. 2003). CARD15 functions through the CARD-containing serine/threonine kinase Rip2 (also known as RICK, CARDIAK, CCK and Ripk2) via a homophilic CARD-CARD interaction to activate a NF-kB signaling pathway (Ogura et al. 2001b). Thus, not only is the CARD15 gene product highly conserved between human and mouse, it also has similar activity in activating NF-kB which is known to be involved in the triggering of cytokines, enzymes and adhesion molecules in response to chronic inflammatory diseases.

In human, CARD15 has been mapped to chromosome 16q12 which harbors the IBD1 locus associated with susceptibility to Crohn's disease (CD), a chronic inflammatory disorder of the gastrointestinal tract in human (Cavanaugh et al. 1998). The disease is characterized by granulomas, most frequently involves the ileum and results in malnutrition and wasting (Crohn et al. 1932). Onset is most common between the ages of 15 and 24, although cases have been reported in every age group. The prevalence of CD has been estimated to be about 1 per 1000 in the western world (Hugot et al. 1996). Recently a frameshift and two single nucleotide polymorphisms in the CARD15 gene have been found to be associated with susceptibility to CD (Hugot et al. 2001; Ogura et al. 2001a). These mutations have been tested in several independent populations and the association between the mutations and CD has been validated by several groups (Hampe et al. 2001; Lesage et al. 2002, Murillo et al. 2002, Vermeire et al. 2002, Croucher et al. 2003). The frameshift mutation results in a truncated CARD15 protein which is deficient in inducing LPS-mediated NF-kB activation (Ogura et al. 2001b). Mouse CARD15 mutants corresponding to the frameshift mutation and the G908R point mutation in human were also deficient in responding to LPS and PGN (Ogura et al. 2003) suggesting that the mutations associated with Crohn's disease share a signaling defect in response to bacterial components. Additional single nucleotide polymorphisms located in the NOD of the CARD15 gene have been shown to be associated with Blau Syndrome (Miceli-Richard et al. 2001). Blau syndrome is an autosomal dominant disorder which results in arthritis, uveitis and skin rash. Both disorders are characterized by granulomatous inflammation but have no further similarity.

The CARD15 gene appears to be an excellent candidate for disease resistance in cattle. The seminal work performed in human and mouse provides insight to the function of the gene and its possible role in immune response to pathogens. More specifically, the bovine CARD15 gene could play an important role in protection against granulomatous disorders. One such disorder, Johne's disease, is believed by some to be related to Crohn's disease. Although there is no conclusive evidence that Crohn's and Johne's diseases are related, the parallels between the diseases are striking. For example, both diseases have similar clinical manifestations including chronic diarrhea and weight loss, both species respond to the disease by developing granulomas in the ileum and both species suffer periods of remission and relapse.

Like Crohn's disease, Johne's disease (JD) is a chronic digestive disorder. Although the etiology of CD is unknown, JD is caused by the intracellular bacterium, *Mycobacterium avium* subspecies *paratuberculosis* (*M. ptb.*) (Chiodini et al. 1984a). The disease is characterized by a granulomatous enteritis and clinical indications usually include progressive weight loss, persistent diarrhea and eventual death. The infection can be transmitted by either direct or indirect contact between infected and susceptible animals and most often occurs via the fecal-oral route (Sweeney 1996). Resistance to the development of a paratuberculosis infection appears to be age related, since susceptibility to infection appears to be much greater during an animal's first few months of life than in adults, who are quite resistant (Sweeney 1996). Although the initial infection frequently occurs during calfhood, the manifestation of clinical disease does not occur until adulthood. Animals infected with *M. ptb.* may become clinically diseased, subclinically infected, or may be asymptomatic carriers of infection. Both diseased and clinically normal animals may shed bacilli in feces, and carrier animals readily spread infection in the herd. The degree of fecal shedding of bacilli tends to increase as the disease progresses and whenever animals are subjected to stress, such as during calving or transport.

Johne's disease is considered, by some, to be one of the most serious diseases that affects cattle, costing the U.S. industry as much as $1.5 billion annually as a result of reduced productivity and premature culling (Stabel 1998). Some breeds of cattle, including Jersey (Jakobsen et al. 2000) and Brahman (Roussel, unpublished data) possess an increased risk of being seropositive against paratuberculosis. The heritability of susceptibility to *M. ptb.* in cattle has been estimated to be 0.09 (Koets et al. 2000), which is consistent with estimates of heritability of susceptibility to other disease traits (Philipsson et al. 1980; Simianer et al. 1991). The disease is widespread, affecting approximately 22% of dairy herds and 8% of beef cattle herds in the U.S. (United States Department of Agriculture 1997). There is no cure for the disease and no effective treatment for the elimination or reduction of bacteria in feces, however, there currently are available treatments to improve the condition and well-being of infected animals.

M. ptb. is an obligate pathogen and though it is unable to multiply outside the bodies of host animals (Chiodini et al. 1984a; McClure et al. 1987; Chiodini 1989) it can survive for many years in water or soil (Chiodini et al. 1984a). Signs observed in infected animals are caused by the animal's immune system mounting an attack on its own tissues. This phenomenon is common in mycobacterial infections. The bacterium can be transmitted across the uterine and placental barriers before birth as well as through the ingestion of colostrum, milk or feces from an infected individual (Sweeney 1996). The disease may be rapidly propagated through contact in young calves; whereas, animals exposed after weaning are much less likely to develop the disease (Whitlock and Buergelt 1996). Infected animals can harbor the bacterium without showing signs of disease or reacting to serological tests, which makes diagnosis and detection of the disease particularly difficult. The organism can be found in large quantities in the intestine of the diseased animal and to a lesser extent in lymph nodes, liver and spleen. Animals with clinical or subclinical disease shed the bacteria in feces and to a lesser extent in their milk, making the disease difficult to manage and control.

Previous research suggests that CD may be caused by infection with *M. ptb.* and that the autoimmune symptoms which characterize the disease are caused by the body mounting an immune response against *M. ptb.* infection as occurs with JD (Cuvelier et al. 1994; Dell'Isola et al. 1994). There is considerable evidence to support an association between *M. ptb.* infection and CD. Several researchers have found associations based on microbiological data, multiple host studies, genetic probe data and antimicrobial drug studies. However, there are also researchers who do not acknowledge the relationship. Their disbelief is based on the lack of a pathologic hallmark, low cultivation success, and variability in the PCR data. Each of these concerns, however, can be explained. For instance, the lack of a pathogenic hallmark can be explained by the fact that the bacterium thought to be responsible for CD was found in spheroplast form (which does not have a cell wall) and therefore does not stain using the acid fast staining technique used for the identification of the presence of the bacterium. Low cultivation success is most likely due to the fact that the bacterium is extremely slow growing and therefore is difficult to grow in culture, as well as to the fact that it is easy to kill the spheroplast form of the bacterium with normal purification techniques. The presence of conflicting PCR data (i.e., some studies show the presence of the mycobacterium in controls) can be explained if the bacteria are present in low amounts in the gut, but require susceptible individuals to contract the disease and develop symptoms.

There is considerable evidence supporting an association between *M. ptb.* and CD. Many groups have been successful in isolating *M. ptb.* from patients with CD (Chiodini et al. 1984b; Chiodini 1989; Cocito et al. 1994; Thompson 1994) suggesting a significant association between infection with the bacteria and expression of the disease. Furthermore, *M. ptb.* strains isolated from humans have produced intestinal disease in other hosts such as goats and chickens (Van Kruiningen et al. 1986; Van Kruiningen et al. 1991) which indicates that the bacterium is capable of causing similar signs of disease in several species. Additional evidence has been provided from the spontaneous infection of non-human primates with *M. ptb.* One such case involves a colony of 38 stump-tail macaques, of which, 29 had paratuberculosis (McClure et al. 1987). A second example is the isolation of *M. ptb.* from cotton-top tamarins, which are now considered to be one of the best animal models for Inflammatory Bowel disease (Warren and Watkins 1994). A third case involves a Mandrill baboon in a midwestern zoo (Zwick et al. 2002). The cases described here indicate that *M. ptb.* is versatile and is undeniably linked to intestinal disease in a variety of species. Evidence is further provided in several studies that have used genetic probes and shown the presence of *M. ptb.* in a significantly higher percentage of CD patients than among people without the disease (Moss et al. 1992; McFadden et al. 1992; Sanderson et al. 1992; Wal et al. 1993; Dell'Isola et al. 1994; Fidler et al. 1994; Lisby et al. 1994; Yokoyama et al. 1994). One study detected *M. ptb.* RNA in 100% of Crohn's patients and in none of the controls (Mishina et al. 1996). Finally, support for the connection between *M. ptb.* and CD has been generated through the use of antimicrobial drugs. For example, Gui et al. (1997) found that more than 90% of the Crohn's patients treated with antimicrobial drugs active against *M. ptb.* experienced clinical remission.

Due to the conservation of sequence and function between species and the detected association with CD, CARD15 is strongly suggested as a candidate for disease resistance in cattle. CARD15 activates NF-kB in response to LPS which is found in gram negative bacteria and also to PGN which is found in gram negative bacteria, gram positive bacteria and mycobacteria. For these reasons, a series of analyses of bovine CARD15 was performed in order to provide the fundamental tools that will be required for future studies involving CARD15 as a candidate gene for disease resistance in cattle.

The first study reports the sequence, genomic structure and tissue expression profiles for the bovine CARD15 gene. In this study, a somatic cell hybrid panel and a radiation hybrid panel (Womack and Moll 1986; Womack et al. 1997) were utilized to unequivocally determine the map location of bovine CARD15 and to confirm homology to human CARD15 based on chromosomal location. This information will prove useful in comparative studies of the gene involving other species, or in QTL or candidate gene studies aimed at identifying regions of the bovine genome harboring genes implicated in disease resistance.

The second study provides a sequence comparison of the CARD15 gene in human, mouse and bovine. These data provide information concerning the homology of the CARD15 gene among species and reveal putative regulatory regions included in the 3'UTR, 5'UTR and the introns. Finally, in this study consensus sequences for the CARD15 gene in *Bos indicus* and *Bison bison* are reported.

The third study provides a survey of single nucleotide polymorphisms and the haplotypes found in the CARD15 gene among selected cattle breeds. This study identifies variant forms of the CARD15 gene which may be used in future studies to identify mutations associated with a particular disease. This study also includes tests of association between the identified polymorphisms or haplotypes and Johne's disease.

SEQUENCE, GENOMIC STRUCTURE, CHARACTERIZATION AND EXPRESSION OF BOVINE CARD15

Introduction

CARD15 is a gene that has been identified in human with structural homology to plant disease resistance genes (R). In plants, these R genes are known to aid in the recognition and defense of various intracellular pathogens (Dixon et al. 2000). In human, CARD15 expression has been shown to be enhanced by proinflammatory cytokines and bacterial components including lipopolysaccharide (LPS) and peptidoglycan (PGN) via NF-kB (Gutierrez et al. 2002). NF-kB is a molecule that plays a role in almost every aspect of cell regulation: immune cell activation, proliferation, apoptosis, stress responses, differentiation and oncogenic transformation. NF-kB can be activated by a variety of pathogenic stimuli, including bacterial products. Activation occurs within minutes of stimulation and mediates the expression of a diverse set of inflammatory and immune response mediators. It is considered to be a central regulator of cellular responses and plays a pivotal role both at the stage of initiation and perpetuation of chronic inflammation. The behavior of CARD15 when stimulated with bacterial components and its association with the NF-kB pathway, suggests that CARD15 may function as an intracellular regulator of pathogens in mammals.

Human CARD15 consists of two N-terminal caspase recruitment domains (CARD), a centrally located nucleotide binding oligomerization domain (NOD) and 10 leucine rich repeat domains (LRR) at the C-terminus (Ogura et al. 2001a). Associations between mutations in CARD15 and two granulomatous disorders, Crohn’s disease (Hugot et al. 2001; Ogura et al. 2001a) and Blau syndrome (Miceli-Richard et al. 2001), have recently been reported. Little is known about the mechanism(s) by which these mutations confer susceptibility to disease. However, initial experiments in human have shown that the 3020insC mutation associated with susceptibility to Crohn’s disease decreases the responsiveness of CARD15 to LPS and to muramyl dipeptide (MDP) derived from PGN (Ogura et al. 2001b; Gutierrez et al. 2002; Inohara and Nunez 2003;

Girardin et al. 2003). Additional experiments involving mouse Card15 have shown that mutations equivalent to the human G908R and 3020insC mutations associated with susceptibility to Crohn's disease result in a loss of NF-kB activation in response to LPS and PGN (Ogura et al. 2003). These experiments suggest that mutations in the LRR of CARD15 inhibit recognition or binding ability to bacterial components resulting in a perturbation of NF-kB activation which is coupled with the formation of granulomas. While the exact mechanism(s) by which the mutations contribute to the development of Crohn's disease and Blau syndrome is unknown, it is clear that mutations in CARD15 are associated with susceptibility to both diseases.

Granulomatous disorders have detrimental consequences to animal health, which result in reduced longevity and productivity and cost the livestock industry millions of dollars each year. Furthermore, these disorders are often highly contagious and can rapidly spread, devastating entire operations. A far more significant consequence of these disorders is the contamination of food products for human consumption and ensuing zoonoses. In particular, cattle are affected by granulomatous disorders such as Johne's disease, brucellosis and tuberculosis; all of which are known, or are believed by some, to be zoonotic. Of these, some researchers believe that Johne's disease has a similar etiology to Crohn's disease (Chiodini 1989; Cocito et al. 1994; Thompson 1994). Consequently, one might expect genes involved in susceptibility to Crohn's disease to also be involved in susceptibility to Johne's disease. Consequently, characterization of the bovine CARD15 homolog should prove useful in downstream tests of association with granulomatous disorders in cattle such as Johne's disease and this information may also have utility in other livestock species. Here, the genomic location, sequence, structure and tissue-specific mRNA expression patterns for the bovine CARD15 gene are reported.

Materials and Methods

Probe development. Nucleotide sequences with homology to human CARD15 (GenBank accession AF178930) were identified in the GenBank non-human, nonmouse, expressed sequence tag database (EST_other). One of these sequences, *Bos taurus* (GenBank accession BF605150), was used for primer design using Primer3 available at http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi. The primer pair (5'-GCACAACCTCCAGATCACAG-3', 5'-GACACCGCTGGACACAATC-3') was optimized for annealing temperature and PCR was performed on a Perkin Elmer 9700 thermocycler. The PCR reactions were carried out in a 50 µl reaction volume: 15 mM Tris, 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP (dATP, dCTP, dGTP and dTTP), 1 µM each primer, 50 ng DNA, 0.5 units of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, Calif.). The general PCR reaction included an initial hot start of 10 min at 95°C followed by 35 cycles for 30 s at 94°C, 30 s at 58°C and 45 s at 72°C, followed by a final elongation step at 72°C for 7 min. The PCR products were analyzed by electrophoresis on 2% agarose gels in 1.0× TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide. The PCR product was purified using QIAquick® (Qiagen, Valencia, Calif.) spin columns. Five volumes of Buffer PB (binding buffer) was added to each sample and was well mixed. The sample was then applied to a QIAquick spin column and centrifuged at maximum speed for 1 min. The flow through was discarded and 750 µl of Buffer PE (wash buffer) was added. The column was twice centrifuged at maximum speed for 1 min with the flow through being discarded after each spin. The DNA was then eluted by adding 30 µl of Buffer EB to the center of the QIAquick membrane and incubated at room temperature for 1 min. The assembly was centrifuged for 1 min at maximum speed to release the purified DNA. Cycle sequence reactions were performed in a 10 µl reaction volume: 1 µl Big Dye v3.0 (Applied Biosystems), 10 ng amplified DNA, 0.5 µl of 5% dimethylsulfoxide (DMSO), 0.5 µl of 20 µM primer. The general cycle sequence reaction included an initial denaturation step of 2 min at 96°C followed by 35 cycles for 15 s at 96°C, 20 s at 50°C and 4 min at 60°C, followed by a final elongation step at 60°C for 15 min. Unincorporated nucleotides and primers were removed from the reactions using a BioMax® Spin-50 mini-sephadex column (Millipore Corp., Bedford, Mass.). The columns were initially centrifuged at 1000 × g for 3 min to remove excess water, 10 μl of double distilled water was then added to each sequencing reaction and a total volume of 20 μl was added to each column. The assembly was centrifuged a second time at 1000 × g for 3 min and the tube containing the sample was dried in a SpeedVac for 30 min. All sequencing samples were stored at -20°C until ready for loading. Sequencing reactions were run on a 3100 DNA analyzer (Applied Biosystems). The resulting sequence was analyzed by BLAST search and sequence alignment against the GenBank database to confirm homology to human CARD15.

Genomic localization of bovine CARD15. The appropriate annealing temperature for the confirmed primer pair was established to yield specific amplification of a bovine PCR product in a murine/hamster background. The PCR products were analyzed by electrophoresis on a 2% agarose gel in 1.0× TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide. PCR was performed in 31 somatic and 90 selected radiation hybrid cell lines from a cattle-hamster somatic hybrid cell (SHC) panel and a 5000 rad whole-genome radiation hybrid ($WGRH_{5000}$) panel, respectively (Womack and Moll 1986; Womack et al. 1997). PCR products were scored as: '+' for present, '–' for absent, or '?' for ambiguous. Syntenic assignment was made from correlations of marker retention (Chevalet and Corpet 1986). All $WGRH_{5000}$ typing experiments were performed twice and were independently scored to increase the accuracy of the results. Only data concordant in both experiments were used for RH mapping. To assign bovine CARD15 to a chromosomal region, two-point linkage analysis was performed using the RHMAPPER software (Slonim et al. 1997).

cDNA clone sequencing. Three cDNA clones developed at the United States Department of Agriculture (USDA) Meat Animal Research Center (MARC), MARC 3 BOV plate 54 row N column 10 (GenBank accession BF605150), MARC 2 BOV plate 134 row A column 15 (GenBank accession BM032079) and MARC 3 BOV plate 41 row J column 13 (GenBank accession BF601658) previously identified as harboring ESTs with homology to human CARD15 were purchased from the Children's Hospital Oakland Research Institute (CHORI) for complete sequence analysis. The clones were received as bacterial LB agar stab cultures. On receipt, the cultures were streaked on LB agar plates containing 12.5 μg/ml chloramphenicol. A single colony was inoculated in 3 ml LB culture with 12.5 μg/ml chloramphenicol and incubated at 37°C for 16 hr with agitation at ~300 rpm. The overnight culture was split and transferred into two 1.5 ml Eppendorf tubes and centrifuged for 3 min at 6500 rpm. The supernatant was removed and the pellet was resuspended in 300 μl of Buffer P1. To lyse the cells, 300 μl of Buffer P2 was added and incubated at room temperature for no more than 5 min. To precipitate the DNA 300 μl of chilled Buffer P3 was added, mixed by inverting the tubes 4-6 times and incubated on ice for 10 min. The sample was centrifuged at maximum speed for 10 min and the cleared lysate removed for purification. A QIAGEN-tip 20 (Qiagen) was equilibrated by applying 1 ml Buffer QBT and allowing the column to empty by gravity flow. The supernatant was then applied to the column and allowed to enter the resin by gravity flow. The tip was then washed 4 times with 1 ml Buffer QC heated to 65°C and the DNA eluted with 800 μl Buffer QF. The eluate was precipitated with 0.7 volumes of room temperature isopropanol and centrifuged at maximum speed for 30 min. The supernatant was carefully removed with a pipette and the pellet was washed with 1 ml of 70% ethanol. The sample was centrifuged at maximum speed for 10 min and the ethanol completely removed. The pellet was allowed to air dry for 15 min before resuspending in TE buffer. Approximately 300 ng of the purified plasmid DNA was used as a template in a 10 μl reaction volume: 2 μl Big Dye v3.0, 2 μl *half*BD (Genetix USA Inc., Boston, Mass.), 0.5 μl MasterAmp, 0.5 μl of 20 μM primer (Universal M13 primers: 5'-TGTAAAACGACGGCCAGT-3' or 5'-CACACAGGAAACAGCTATGA-3'). All sequencing reactions and subsequent clean-ups were performed as previously described.

BAC library screen. The primer pair previously described was used to screen a bovine BAC library (Cai et al. 1995) for clones harboring the bovine CARD15 gene. The PCR systematic screening strategy of Green and Olson (1990) was used. First, the DNA super pools which each consist of 480 BAC clones were screened. The positive pools identified in this screen point to a single 96-well microtiter plate. By screening 8 row and 12 column DNA pools prepared from this plate, positive clones were identified by locating the intersection of the positive row and column pools. The PCR and visualization protocols were as previously described.

BAC DNA extractions. DNA from each BAC clone was prepared using a DNA midi-prep procedure from Qiagen. A single colony from an LB/chloramphenicol agar plate streaked with BAC culture was inoculated into 100 ml 2× YT with chloramphenicol at a final concentration of 12.5 μg/ml into a 1 L flask. The culture was incubated at 37°C for 20 hr with agitation at ~300 rpm. The bacterial culture was split into two 50 ml conical Falcon tubes and centrifuged for 20 min at 4°C at 5,600 rpm in a Beckman centrifuge with a JA-12 rotor. The supernatants were poured off and tubes were inverted for about 1 min to remove all traces of the supernatant without disturbing the pellet. Each pellet was resuspended in 10 ml of P1 solution (Qiagen) with 10 μl of RnaseA (10 mg/ml). Ten ml of P2 solution (Qiagen) was added, each tube was inverted four to six times, and the lysing reaction was allowed to continue for 5 min at room temperature or until the solution was clear. Ten ml of solution P3 (Qiagen) was added to neutralize the reaction, tubes were inverted four to six times and incubated on ice for 10 min, then centrifuged for 30 min at 4°C at 8000 rpm to separate the cell debris from the lysate. The clear supernatants were transferred by pipette into a single clean 50 ml conical tube and again centrifuged at 8000 rpm for 10 min to ensure the removal of all cell debris that might interfere with column purification. The supernatant was again transferred into a clean 50 ml tube. A Qiagen column was placed in a 50 ml Falcon tube and 4 ml of Buffer QBT added as an equilibration wash. The sample was added to the column and all flow through liquid was discarded. Two volumes of 10 ml of Buffer QC were added to the column as a wash, and the BAC DNA was eluted using five 1 ml volumes of Buffer QF, preheated to 65°C. DNA was precipitated by adding 3.5 ml of isopropyl alcohol and mixing. After at least 5 min, samples were centrifuged at 8000 rpm for 30 min at 4°C to precipitate the DNA and the supernatant was carefully poured off. The pellets were resuspended in 1.5 ml of 70% ethanol, transferred to a 1.5 ml

Eppendorf tube and then centrifuged for 10 min at room temperature. The ethanol was carefully removed using a 200 µl pipette and samples were inverted and allowed to dry for 20-30 min. BAC DNA was dissolved in 40 µl of elution buffer (Buffer EB) and stored at room temperature overnight before use.

BAC fingerprinting. BACs harboring the bovine CARD15 gene were digested using *Bam*HI and *Sau*3AI restriction enzymes. Each reaction used 1 µg of DNA, 2.5 µl of appropriate buffer and 1 unit of enzyme in a 25 µl volume. The digests were incubated at 37°C for 1 hr and heat inactivated at 65°C for 15 min. Analysis was by electrophoresis on a 0.65% agarose gel in 1.0× filtered TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide. The fragments were size analyzed based on a *Hind*III digest of lambda DNA and a 1 kb DNA ladder.

BAC sequencing. Approximately 1 µg of DNA was used as a template in a 10 µl reaction volume: 2 µl Big Dye v3.0, 2 µl *half*BD (Genetix), .5 µl MasterAmp, .5 µl of 20 µM primer. All sequencing reactions were run under the following conditions: an initial denaturation step at 96°C for 2 min, 8 cycles of 96°C for 30 s followed by 58°C for 30 s with a 1 degree per cycle touchdown followed by 65°C for 4 min, 60 cycles of 96°C for 30 s followed by 50°C for 25 s followed by 65°C for 4 min, and a final round of 96°C for 1 min followed by 50°C for 1 min followed by 65°C for15 min. Unincorporated nucleotides and primers were removed from the reactions using a BioMax Spin-50 mini-sephadex column (Millipore). The columns were initially centrifuged at 1000 × g for 3 min to remove excess water, 15 µl of double distilled water was then added to each sequencing reaction and a total volume of 25 µl was added to each column. The assembly was centrifuged a second time at 1000 × g for 3 min and the tube containing the sample was dried in a SpeedVac for 30 min. All sequencing samples were stored at -20°C until ready for loading. Sequencing reactions were run on a 3100 DNA analyzer (Applied Biosystems). This strategy was used with both gene specific primers and primers with homology to BAC ends developed by Cai et al. (1995) and given the names LEIPCR and REPCR due to homology to the left end and the right end of the BAC respectively. A description of primers used to derive sequence from BAC clones is presented in Table 1. All resulting sequence was analyzed by BLAST search against the GenBank database for validation.

RNA processing (Blood). Whole blood was collected in heparin tubes, stored on ice and centrifuged at 2000 × g for 15 min at room temperature within 8 hr of collection. The plasma layer was removed using transfer pipettes and the white blood cell (WBC) layer was collected and combined with 1.2 ml of RNAlater (Ambion, Austin, Tex.). The sample was mixed thoroughly by vortexing. The WBC mixture was then centrifuged for 1 min at maximum speed in a microcentrifuge. The supernatant was completely removed by aspiration. To lyse the cells and solubilize the pellet, 800 μl of lysis solution and 100 μl of sodium acetate solution were added and the tube was vortexed vigorously. The RNA was then extracted by adding 500 μl of acid phenol chloroform, vortexing for 30 s and incubating at room temperature for 3 min. To separate the aqueous and organic phases, the sample was centrifuged for 1 min at maximum speed. The aqueous phase was transferred to a new 2 ml tube and 600 μl of 100% ethanol was added. The sample was mixed thoroughly and then applied 700 μl at a time to the cartridge assembly. With each addition of sample the cartridge was centrifuged briefly to pass the liquid through the filter and the flow through was discarded. The sample was then washed by adding 700 μl of Wash Solution 1, centrifuging and discarding flow through. An additional wash with Wash Solution 2/3 was repeated 2 times with a final centrifugation for 1 min to remove all residual fluid from the filter. The RNA was eluted by applying 100 μl of Elution Solution preheated to 75°C. The RNA was recovered by centrifugation for 30 s at maximum speed. The eluted RNA was then treated with DNase I to remove contaminating genomic DNA.

Table 1. Primers used to amplify regions of bovine CARD15 from BACs

Primer	*Sequence 5' to 3'*
LEIPCR[a]	CTAGAGTCGCCTGCAGG
REPCR[a]	GCGGATAACAATTTCACACAGG
E2-5'.1[b]	CAGAATACTCTCAAAGCCCTCCAG
E2-3'.1[b]	GACACCATCTGGAATAAGGGTACT
E3con5'[c]	TGTAGAAGGAAGGCAGCCAAT
E3con3'[c]	GGTTGGCTGCCTTTCTTCTAC
E4 5'[d]	CCATGGGTGGCTAAGGGTAG

[a] BAC end-sequencing primers.
[b] Primers to amplify in the 5' and 3' directions from exon 2.
[c] Primers to amplify in the 5' and 3' directions from exon 3.
[d] Primers to amplify in the 5' directions from exon 4.

RNA processing (Tissue). Tissue was collected at slaughter from yearling Angus steers sourced from the Circle A Ranch, Iberia, Missouri. Samples were obtained from: pancreas, liver, *longissimus dorsi* (ribeye) muscle, lower intestine, heart, lung, kidney, thyroid, spleen, hypothalamus, adrenal, anterior pituitary and bone marrow. The tissue was dissected into approximately 500 mg pieces and placed immediately into 5 ml of RNAlater (Ambion) and stored on ice before finally being transferred to a -20°C freezer.

RNA extractions were performed using TRIzol reagent. The tissue samples were removed from RNAlater and approximately 100 mg of tissue was chopped into small pieces using a clean razor blade. The tissues were then homogenized using a Dispergierstation T8.10 (IKA Works Inc., Wilmington, N. Carol.) in 1 ml of TRIzol reagent. After homogenization, the samples were incubated at room temperature for at least 5 min to allow complete dissociation of nucleoprotein complexes. In order to separate the phases, 0.2 mL of chloroform was added and each tube was shaken by hand for 15 s and then incubated at room temperature for 2 min. The samples were centrifuged at 12,000 × g for 15 min and the aqueous phase was transferred to a clean tube. To precipitate the RNA, 0.5 ml of isopropyl alcohol was added and incubated at room temperature for 10 min. The tubes were then centrifuged at 12,000 × g for 10 min. The supernatant was removed by pouring and the pellet was washed in 1 ml of 75% ethanol. The samples were mixed by vortexing and then centrifuged at 7500 × g for 5 min. The supernatant was removed by pouring and the pellet allowed to dry for 15 min. The pellet was then dissolved in 20 µl of DEPC treated water and incubated for 5 min at 42°C. RNA purity was determined from calculations of 260/280 ratios. Integrity of the RNA was checked via gel electrophoresis.

RT-PCR. 1 µg of total RNA was primed with random decamers using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen Life Technologies, Carlsbad, Calif.). RNA/primer mixtures were set up in 10 µl reaction volumes: 1 µg RNA, 1 µl of 10 mM dNTP mix, 1 µl of 20 µM gene specific primer. The reaction was incubated at 65°C for 5 min and then placed on ice for at least 1 min. The RT reaction mixture was also set up in a 10 μl reaction volume: 2 μl 10× RT buffer, 4 μl of 25 mM $MgCl_2$, 0.1 M DTT, 1 μl RNaseOUT. To each RNA/primer mixture, 9 μl of RT reaction mix was added and incubated at 42°C for 2 min. Each tube then received 50 units of SuperScript II RT and was incubated for an additional 50 min at 42°C. The reactions were terminated at 70°C for 15 min and then chilled on ice. RNase H (1 μl) was added to each tube and incubated for an additional 20 min at 37°C. All RT products were tested for the presence of CARD15 and beta-actin transcripts before being used as a template in quantitative PCR.

5' and 3' RACE. Gene specific nested primers were designed for both the 5' and 3' ends of bovine CARD15. Using the FirstChoice RLM-RACE kit (Ambion), 10 μg of total RNA was treated with calf intestinal phosphatase to remove the 5' phosphate from degraded mRNA, rRNA, tRNA and DNA and incubated at 37°C for 1 hr. The reaction was terminated and a phenol:chloroform extraction performed. The sample was then subjected to a Tobacco Acid Pyrophosphatase treatment to remove the cap structure from full-length mRNA leaving a 5'-monophosphate. Next, a 45 base RNA adapter was ligated to the RNA population using T4 ligase. Reverse transcription was then performed using random decamers supplied with the kit at 42°C for 1 hr.

Nested PCR using outer gene specific primer and outer 5'RACE primer in a 50 μl reaction volume: 1 μl RT product from the previous step, 5 μl 10× PCR buffer, 4 μl dNTP mix, 2 μl 10 μM gene specific outer primer, 2 μl 5' RACE outer primer, 1.25 units AmpliTaq Gold DNA Polymerase (Applied Biosystems). The PCR reaction included an initial denaturation step of 10 min at 94°C followed by 35 cycles for 30 s at 94°C, 30 s at 60°C, 30 s at 72°C with a final elongation of 7 min at 72°C. The product from the 5'RACE outer PCR was used as a template for a second round of PCR using a nested gene specific primer and the inner 5'RACE primer supplied with the kit. The same PCR conditions were used as for the outer RACE PCR. The PCR products were analyzed by electrophoresis on 2% agarose gels in 1.0× TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide.

The FirstChoice RLM-RACE kit (Ambion) was also used for 3'RACE. Total RNA (1 μg) was used for reverse transcription in a 20 μl volume: 4 μl dNTP mix, 2 μl 3'RACE adapter, 2 μl 10× RT buffer, 1 μl RNase Inhibitor, and 1 μl MMLV Reverse Transcriptase. The reaction was incubated at 42°C for 1 hr. The RT product was then used for an outer 3'RACE PCR reaction in a 50 μl volume: 1 μl RT reaction, 5 μl 10× PCR buffer, 4 μl dNTP mix, 2 μl 10 μM gene specific primer, 2 μl 3'RACE outer primer, 1.25 units AmpliTaq Gold DNA Polymerase (Applied Biosystems). The PCR reaction included an initial denaturation step of 10 min at 94°C followed by 35 cycles for 30 s at 94°C, 30 s at 60°C, 4 min at 72°C with a final elongation of 15 min at 72°C. The product from the 3'RACE outer PCR was used as a template for a second round of PCR using a nested gene specific primer and the inner 3'RACE primer supplied with the kit. The same PCR conditions were used as described for the outer RACE PCR. The PCR products were analyzed by electrophoresis on a 1% agarose gel in 1.0× TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide.

cDNA sequencing. Total RNA was primed using a gene specific primer from the 3' end of the bovine CARD15 gene (5'-CTAGGGAGCTGATGTGGTTGTTAG-3') using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen) as previously described.

Overlapping PCR primers were amplified in a 50 μl volume: 15 mM Tris, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM of each dNTP (dATP, dCTP, dGTP, and dTTP), 1 μM each primer, 2 μl of template from the RT reaction, 0.75 units of AmpliTaq Gold DNA Polymerase (Applied Biosystems). The general PCR reaction included an initial hot start of 10 min at 95°C followed by 35 cycles for 30 s at 94°C, 30 s at 58°C and 2 min at 72°C, followed by a final elongation step at 72°C for 15 min. Primers used to amplify the three overlapping cDNA amplicons were:

A1F: 5’-ACCAGCCATTGTCAGGAGAC-3’
A1R: 5’-GCAGAAGGTTGAAGAGCAGACTC-3’ (Amplicon 1);
A2F: 5’-CTTTGCCTGGAAGAAGTATATACAGAG-3’
A2R: 5’-CATGCTCTTGGCCTCACC-3’ (Amplicon 2); and
A3F: 5’-GCACAACCTCCAGATCACAG-3’
A3R: 5’-CTAGGGAGCTGATGTGGTTGTTAG-3’ (Amplicon 3).

The PCR products were analyzed by electrophoresis on 2% agarose gels in 1.0× TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide. The PCR product was purified using QIAquick spin columns.

Cycle sequencing reactions were performed in a 10 µl reaction volume: 1.5 µl Big Dye v3.0, 1.5 µl *half*BD (Genetix), 0.5 µl MasterAmp, 0.5 µl of 20 µM primer. The general cycle sequence reaction included an initial denaturation step of 2 min at 96°C followed by 35 cycles for 15 s at 96°C, 20 s at 50°C and 4 min at 60°C, followed by a final elongation step at 60°C for 15 min. Unincorporated nucleotides and primers were removed from the reactions using a BioMax Spin-50 mini-sephadex column. Sequencing reactions were run on a 3100 DNA analyzer (Applied Biosystems). The resulting sequence was entered into ContigExpress and aligned with the genomic sequence of bovine CARD15.

Intron sizing. Primers were developed using sequence data from each exon of bovine CARD15. Two primers were developed from each exon, one extending in the 5’ direction and the other extending in the 3’ direction. The primers were then paired in order to amplify introns. For example, the primer extending in the 3’ direction in exon 1 was paired with the primer extending in the 5’ direction in exon 2 to amplify intron 1. The resulting amplicon contains partial sequence from exon 1, partial sequence from exon 2 and the complete sequence from intron 1. Primer sequences used to amplify introns are presented in Table 2. All intron amplifications were run in a 50 µl reaction volume: 15 mM Tris, 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP (dATP, dCTP, dGTP, and dTTP), 1µM each primer, 5 µl Master Amp, 100 ng bovine genomic

Table 2. Primers used to amplify introns from bovine CARD15

Intron	*Sense 5' to 3'*	*Antisense 5' to 3'*
1	GGAGCTCTGTGAGATCGCTTCC	CAGAATACTCTCAAAGCCCTCCAG
2	ACCAGCCATTGTCAGGAGAC	TGCTGATGTTTTTGCTCTTCA
3	GGCCTTTTATTGTGGTGGAA	GAAGAGCAGACTCTGGACTGACG
4	GCAGCTGGACCACAACTCTG	CAGCTGCTCACAGCGAAGAG
5	GTCTCTGTGGGGGTTTTGTC	GTCCTTGTTTTCAGCGAGGT
6	ACAAGTTGACCGATGGCTGT	AACTGCAAGGAGTTGTTAGTTCTGAG
7	GCTCAGAACTAACAACTCCTTGC	CCCCTCGTCACCCACCTG
8	AACCAGGTGGGTGACGAG	AGTTCTTCCAGGGCCATATTCT
9	CTGGTGGGGAACAACATTGG	CTTGAGTTTCTTGCAAGTCCTTTG
10	AGAACCACGTCCAGGATGAAG	CTAGGGAGCTGATGTGGTTGTTAG
11	CTAACAACCACATCAGCTCCCTA	GAAACATCAGAGCAAGAGTCTGGTATC

DNA and 0.75 units of High Fidelity Taq (Applied Biosystems). The general PCR reaction included an initial 2 min hot start at 94°C followed by 10 cycles of 94°C for 15 s, 61°C for 30 s with a 0.5°C touchdown each cycle, 68°C for 6 min, followed by an additional 20 cycles of 94°C for 15 s, 56°C for 30 s, 68°C for 6 min with an additional 5 s added each cycle, followed by a final elongation step at 68°C for 7 min. The resulting amplicons were then partially sequenced from both directions to confirm identity and to obtain sequence flanking each exon.

Real Time PCR/TaqMan. Oligonucleotide primers and TaqMan probes were designed using Primer Express, version 1.0 (Applied Biosystems). Primers used to amplify CARD15 were:

C15TMF: 5'-CAGCCAGTACGAAACTGATGAAA-3'
C15TMR: 5'-CACTGTGGCGAGATCAAGGA-3'.

Primers used to amplify beta-actin (ACTB) were:

ACTBTMF: 5'- TGTGCTCTCCCTCTATGCTTCTG-3'
ACTBTMR: 5'- GGTGACACCATCCCCTGAAT-3'.

The CARD15 dual labeled probe (C15TMP: 5'-CAGGCGGCCCATCTTCACTTCATC-3') was designed to have a 5' reporter dye (6-FAM) and a 3' quenching dye (TAMRA). The beta-actin dual labeled probe (ACTBTMP: 5'-CGCACAACGGGCATCGTCCTG-3') was designed to have a 5'reporter dye (VIC) and a 3' quenching dye (TAMRA).

Relative quantitative real-time PCR: All semi-quantitative PCR was performed using an ABI PRISM 7700 Sequence Detection System (Applied Biosystems). The PCR reaction mixture (50 μl) contained 25 μl of 2× TaqMan Universal PCR Master Mix (Applied Biosystems), 300 nM primers, 200 nM TaqMan probe, 1 μl of cDNA sample and water. The thermal cycling conditions comprised the initial steps at 50°C for 2 min and at 95°C for 10 min, followed by 50 cycles at 95°C for 15 s and at 60°C for 1 min.

Following RT-PCR, quantitation of gene amplification was made by setting the threshold cycle (C_T) in the geometric region of the plot after examining the semi-log view of the amplification plot. Relative quantification of gene expression was evaluated using the comparative C_T method. The ΔC_T value is determined by subtracting the gene

C_T of each sample from the corresponding sample beta-actin C_T value. Calculation of $\Delta\Delta C_T$ involves using the mean ΔC_T from liver as an arbitrary constant to subtract from all other ΔC_T mean values. Fold changes in gene expression were then determined by exponentiating the mean difference, $2^{-\Delta\Delta Ct}$.

Results

Genomic localization of bovine CARD15. To identify the bovine ortholog of CARD15, the EST-others database in GenBank was searched for sequences encoding peptides with homology to human CARD15 (GenBank accession AF178930). Four EST clones were identified with significant E values for homology to approximately 1100 bases of the 4485 bp human CARD15 gene (Table 3). Of these, the first is a clone derived from *Sus scrofa* and the remaining 3 are *Bos taurus* clones. Sequence from the *Bos taurus* cDNA clone with the highest E value (GenBank accession BF605150) was chosen to develop a set of primers to be used as a probe for integrating bovine CARD15 into the $WGRH_{5000}$ RH map. The 141 bp PCR product produced using these primers was sequenced and homology to the bovine EST from which it was derived was confirmed. The sequence was then subjected to BLAST searches which revealed significant homology exclusively to human CARD15 and mouse Card15. This confirmed the use of the primer pair as a detector of bovine sequence homologous to CARD15. This confirmed primer pair was then used to screen a cattle-hamster somatic hybrid cell panel and localized bovine CARD15 to BTA18 with 97% concordancy. The same primer pair was used to screen a bovine-hamster RH panel and 14 of the 90 hybrid cell lines retained fragments harboring CARD15 yielding a retention frequency of 0.156 (data presented in Table 4). Integration of CARD15 into the RH map of BTA18 revealed two alternatives for gene order with high statistical support. The most likely order had a log-likelihood of -97.072355 and placed CARD15 centromeric to ADCY7, while the second best estimate of order had a log-likelihood of -97.477136 and placed CARD15 telomeric to ADCY7 (Fig. 1).

Table 3. EST sequences in GenBank with homology to human CARD15

Accession Number	*Organism*	*Length (bp)*	*Identities*	*Human homology (bp)*	*E value*
BI339864	*Sus scrofa*	511	286/334	2895-3228	2.00E-72
BF605150[a]	***Bos taurus***	**326**	**219/261**	**2075-2335**	**4.00E-43**
BM032079	*Bos taurus*	545	211/254	1708-1961	2.00E-36
			88/104	1573-1676	
			40/43	2075-2117	
BF601658	*Bos taurus*	489	101/123	3836-3958	4.00E-06

[a] Boldface entry indicates EST selected for primer development and used as a probe for somatic cell hybrid and radiation hybrid mapping.

Table 4. Results from screening the bovine-hamster RH panel

Cell Line	*Score*[a]	*Cell Line*	*Score*	*Cell Line*	*Score*	*Cell Line*	*Score*	*Cell Line*	*Score*	*Cell Line*	*Score*
1	-	16	+	31	-	46	+	61	+	76	-
2	-	17	-	32	-	47	-	62	-	77	-
3	+	18	-	33	-	48	-	63	-	78	-
4	-	19	-	34	+	49	-	64	-	79	-
5	-	20	-	35	+	50	-	65	-	80	-
6	-	21	-	36	-	51	-	66	-	81	-
7	-	22	-	37	-	52	-	67	-	82	-
8	-	23	-	38	+	53	-	68	-	83	+
9	-	24	-	39	-	54	-	69	-	84	-
10	-	25	-	40	-	55	-	70	-	85	-
11	-	26	-	41	+	56	-	71	-	86	-
12	+	27	-	42	-	57	-	72	-	87	-
13	-	28	-	43	-	58	-	73	+	88	-
14	-	29	-	44	-	59	+	74	-	89	-
15	-	30	-	45	+	60	+	75	-	90	-

[a] PCR products were scored as: '+' for present, '–' for absent, or '?' for ambiguous.

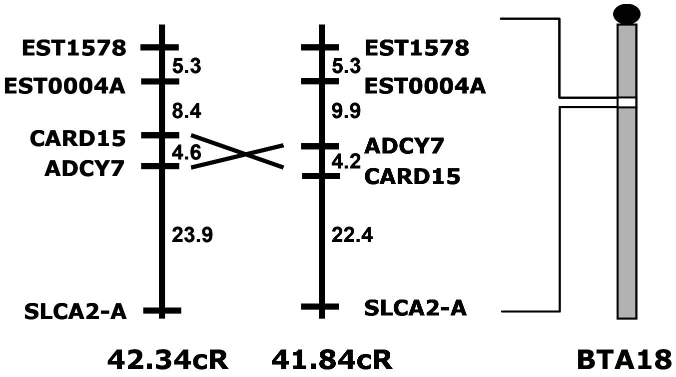


Figure 1. RH map localization of CARD15 to BTA18. Figure to the left represents the most likely marker order, the center figure represents the second most likely marker order, while the figure to the right shows the approximate location of the region of BTA18 harboring CARD15.

Sequence of bovine CARD15. EST clones: To obtain complete sequence data on the bovine CARD15 gene, three partial cDNA clones developed at the USDA MARC were purchased which contained ESTs with homology to human CARD15. These clones were derived from polyA RNA digested with *Not*I and ligated into the cloning vector pCMVSPORT6 (Smith et al. 2001). Universal M13 primers were used to sequence the clones and sequence data were obtained for regions with homology to exons 4-12 of the human CARD15 gene (Fig. 2).

BAC library screen: The Texas A&M University bovine BAC library was screened with the confirmed primer pair developed as previously described and 2 BAC clones harboring sequence with homology to CARD15 were identified (BBAC114R1C4; BBAC7138-84R4C4). Sequence analysis of PCR products confirmed the inclusion of bovine CARD15 in both clones. End sequencing of each of the BACs revealed that clone BBAC114R1C4 contains sequence with homology to CYLD, while end sequences for BBAC7138-84R4C4 produced no homologies to the human genome. The BAC clones were digested with restriction enzymes in order to estimate insert size and to confirm that the clones did not contain the same insert. These digests confirmed that the two clones contain different inserts with estimated sizes of 138 kb for BBAC114R1C4 and 93 kb for BBAC7138-84R4C4.

To obtain coding sequence data not present in the MARC cDNA clones, human and mouse sequence was aligned in Vector NTI (Informax, Frederick, Mar.) and primers were developed from areas of high conservation to be used to sequence directly from the BAC clones identified as harboring bovine CARD15. This strategy allowed for the capture of missing sequence data from exons 2, 3 and 4 (Fig. 2). However, due to the low levels of sequence homology between mouse and human for exon 1, no suitable primers were developed to amplify bovine exon 1 by this process.

Exon order was confirmed in the coding sequence by generating a population of cDNAs using a bovine CARD15 gene specific primer and used this population as a template to produce 3 overlapping amplicons. This approach allowed us to verify both

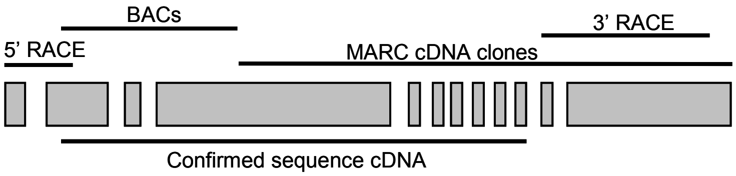


Figure 2. Strategy for obtaining genomic sequence for bovine CARD15. Each boxed region represents an exon and each labeled line represents the region for which sequence was obtained by the indicated approach.

the sequence and exon order from exon 2 through exon 10 of the bovine CARD15 gene (Fig. 2).

5’ and 3’ RACE: To acquire the remaining sequence data from the 5’ and 3’ portions of the coding sequence, rapid amplification of cDNA ends (RACE) was used. A 14 bp mononucleotide A repeat located in the 3’UTR rendered the 3’RACE experiments unsuccessful in gathering the sequence for the entire 3’UTR, as the method required attaching an adapter via the polyA tail to the 3’ end of the transcript. This adapter preferentially ligated to the internal mononucleotide A repeat and resulted in the production of a single amplicon which lacked the most 3’ sequence of the bovine CARD15 gene. Sequence from an additional EST derived from a cDNA clone developed at MARC with homology to human CARD15 was recently deposited to the GenBank database (GenBank accession CB427378). This published EST contained sequence that included the internal polyA site in addition to some downstream 3’ sequence. The clone harboring this EST was obtained from MARC (Tim Smith, pers. comm.) and was sequenced using the primer 5’-CAGCGTTGTGTTAATTTCTATACAGC-3’ located downstream of the A_{14} sequence. This primer was also used to sequence directly from the two BACs harboring bovine CARD15. All of the produced sequences were aligned leading to the identification of a putative polyA signal and the remnants of a polyA tail, thus elucidating the full 3’UTR sequence. The cDNA and genomic BAC sequences had 100% sequence identity up to the polyA tail and subsequent vector sequence of the cDNA clone (Fig. 3). The resulting combination of sequence data demonstrates that the bovine CARD15 messenger RNA is 5105 bp and the protein it encodes is comprised of 1013 amino acids (Fig. 4). The complete coding sequence has been submitted to GenBank as accession AY518737.

Genomic structure of bovine CARD15. To obtain intronic flanking sequence and to estimate intron size, primers were developed to amplify individual introns. Of these, all but two were successful in amplifying the individual introns. Problems were encountered in developing suitable primers from exon 3, therefore, BACs clones were

cDNA TTATGCTTTTTCACAGGATAAACTACTCATAGTTCCTTGTGCTATACATTAGGACTTTTTTATGTATCCA
BAC TTATGCTTTTTCACAGGANAAACTACTCATAGTTCCTTGTGCTATACATTAGGACTTTTTTATGTATCCA

cDNA TTC**AATATA**T**AATACA**TTTGCTAGCCTCAAAAAAA*GAATTCCACTCCCACTGTC*
BAC TTC**AATATA**T**AATACA**TTTGCTAACCTCATAATACATTTGCTAGCCTCAATCTC
⬆

Figure 3. Alignment of cDNA to genomic BAC sequence for determination of the 3' transcript end of bovine CARD15. Potential alternative polyadenylation signals are indicated in bold face. Arrow indicates cleavage site, shaded region represents remnant of polyA tail and vector sequence is italicized.

```
   1    CTCTGGACTC CGTGACCATC ACAGGAGGGA ACCTGCTGAG AGATTTCCTC TGCTGCGTGG

  61    GGACCCTGCC AGGGCTTGGA GCTCTGTGAG ATCGCTTCCC ACGGACTCCC AGGACCCAGA
                                       M   C  A  Q   D  A  F   Q  T  Q  R
 121    GTCTGAGGGC TGAGCCCAGG ATTGTGAAAT GTGCGCACAA GATGCTTTTC AGACACAGAG
        · S  Q  L   V  E  L   L  V  S  G   S  L  E   G  F  E   S  I  L  D
 181    AAGCCAACTG GTGGAGTTGC TGGTCTCGGG GTCCCTGGAG GGCTTTGAGA GTATTCTGGA
        · R  L  L   S  R  E   V  L  S  W   E  D  Y   E  G  L   S  L  V  G
 241    CCGGCTGCTT TCCCGGGAAG TCCTCTCCTG GGAGGACTAT GAGGGGCTTA GCCTCGTGGG
        · Q  P  I   S  H  L   A  R  R  L   L  D  T   I  W  N   K  G  T  W
 301    GCAGCCCATC TCCCACTTGG CCAGGCGCCT CCTGGACACC ATCTGGAATA AGGGTACTTG
        · G  C  E   Q  L  T   A  A  V  R   E  A  Q   A  D  S   Q  P  P  E
 361    GGGCTGTGAA CAACTGACTG CAGCTGTGCG GGAGGCCCAG GCCGACAGCC AGCCCCCCGA
        · L  P  S   S  W  D   P  H  S  P   H  P  A   R  D  L   Q  S  H  R
 421    GCTTCCCAGC TCCTGGGACC CCCACTCACC CCACCCAGCC CGTGACCTGC AGAGTCACCG
        · P  A  I   V  R  R   L  Y  G  H   V  E  G   V  L  D   L  T  Q  Q
 481    ACCAGCCATT GTCAGGAGAC TCTACGGCCA CGTGGAGGGT GTGCTGGACC TGACACAGCA
        · R  G  F   I  S  Q   Y  E  T  D   E  I  R   R  P  I   F  T  S  S
 541    GCGGGGTTTC ATCAGCCAGT ACGAAACTGA TGAAATCAGG CGGCCCATCT TCACTTCATC
        · Q  R  A   R  R  L   L  D  L  A   T  V  K   A  N  G   L  A  A  F
 601    CCAGCGGGCA AGAAGGCTCC TTGATCTCGC CACAGTGAAG GCGAATGGGT TGGCTGCCTT
        · L  L  Q   C  I  Q   E  L  P  V   P  L  A   L  P  F   E  D  A  A
 661    TCTTCTACAG TGTATTCAGG AATTACCGGT CCCATTGGCC CTGCCTTTTG AAGATGCTGC
        · C  K  K   Y  V  S   K  L  R  T   V  I  S   A  Q  S   R  F  L  S
 721    CTGTAAGAAG TACGTGTCCA AGCTGAGGAC CGTTATATCA GCTCAGTCTC GTTTCCTGAG
        · T  Y  D   G  A  E   N  L  C  L   E  E  V   Y  T  E   N  V  L  E
 781    CACCTACGAT GGAGCAGAGA ATCTTTGCCT GGAAGAAGTA TATACAGAGA ATGTTCTGGA
        · I  Q  M   E  V  G   M  A  G  P   S  Q  Q   S  P  T   T  L  G  L
 841    AATCCAGATG GAGGTGGGCA TGGCTGGACC TTCGCAGCAG AGCCCTACCA CCCTAGGCCT
        · E  E  L   F  S  T   R  D  H  F   N  K  E   A  D  T   V  L  V  V
 901    GGAGGAGCTC TTCAGCACCC GTGACCATTT CAACAAAGAG GCAGACACTG TGCTGGTGGT
        · G  E  A   G  S  G   K  S  T  L   L  Q  Q   L  H  L   L  W  A  S
 961    GGGCGAGGCG GGCAGCGGCA AGAGCACGCT CTTGCAGCAG CTGCACCTGC TCTGGGCTTC
        · G  R  A   F  Q  E   F  L  F  V   F  P  F   S  C  R   Q  L  Q  C
1021    CGGGCGGGCC TTCCAGGAAT TTCTCTTCGT CTTCCCATTT AGCTGCCGGC AGCTGCAGTG
        · L  V  K   P  L  S   M  R  T  L   L  F  E   H  C  C   W  P  D  L
1081    CCTGGTGAAA CCGCTGTCCA TGCGGACGCT GCTCTTCGAA CACTGCTGTT GGCCCGACCT
        · G  P  Q   D  V  F   Q  V  L  L   D  H  P   E  R  I   L  L  T  F
1141    TGGCCCCCAG GACGTCTTCC AGGTCCTCCT TGACCACCCT GAGCGCATCC TCTTAACCTT
        · D  G  F   D  E  F   R  F  R  F   T  D  Q   E  R  H   C  C  P  T
1201    TGATGGCTTT GATGAGTTCA GGTTCAGGTT CACGGATCAG GAGCGTCACT GCTGTCCGAC
        · A  P  T   S  V  Q   S  L  L  F   N  L  L   Q  G  N   L  L  K  N
1261    CGCCCCCACG TCAGTCCAGA GTCTGCTCTT CAACCTTCTG CAGGGCAACC TGCTAAAGAA
        · A  R  K   V  L  T   S  R  P  S   A  V  S   A  S  L   R  K  H  V
1321    TGCCCGCAAG GTGTTGACCA GCCGCCCCAG CGCGGTATCG GCGAGCCTCC GAAAGCACGT
        · R  T  E   L  S  L   K  G  F  S   E  E  G   I  E  L   Y  L  R  K
1381    GCGCACGGAA CTCAGCCTCA AGGGCTTCTC GGAAGAGGGC ATCGAACTGT ACCTGAGGAA
        · R  H  R   E  P  G   V  A  D  R   L  L  C   L  L  R   A  T  S  A
1441    GCGGCATCGC GAGCCTGGGG TGGCCGACCG CCTCCTCTGC CTGCTCAGAG CCACCTCGGC
        · L  H  G   L  C  H   L  P  V  F   S  W  M   V  S  K   C  H  E  E
1501    CCTGCACGGT CTGTGCCACC TGCCTGTCTT CTCCTGGATG GTGTCCAAGT GCCACGAGGA
        · L  L  L   Q  G  R   G  S  P  K   T  T  T   D  M  Y   L  L  I  L
1561    GCTGTTGCTG CAGGGCCGGG GGTCCCCAAA GACCACCACG GATATGTACC TGCTGATCCT
        · R  H  F   L  L  H   A  S  P  L   P  L  A   T  H  G   L  G  P  S
1621    GCGGCACTTT CTGCTGCACG CCTCCCCGCT ACCCTTAGCC ACCCATGGCC TGGGACCCAG
```

Figure 4. Nucleotide and amino acid sequence of bovine CARD15

```
        ·L  I  Q   G  R  L  P   T  L  L   H  L  G   R  L  A  L   W  G  L
1681    CCTGATTCAG GGCAGGCTCC CCACACTCCT GCATCTTGGC CGCCTGGCTC TCTGGGGCCT
        ·G  T  C   C  Y  V  F   S  A  K   Q  L  Q   A  A  H  V   D  S  E
1741    GGGCACATGC TGCTACGTGT TCTCAGCCAA ACAGCTGCAG GCGGCACATG TCGACAGTGA
        ·D  L  S   L  G  F  L   V  L  A   K  R  V   V  P  G  S   T  A  P
1801    GGACCTTTCT CTTGGCTTCC TGGTGCTTGC CAAGAGGGTT GTACCTGGGA GTACAGCCCC
        ·L  E  F   L  H  I  T   F  Q  C   F  F  A   A  F  Y  L   A  L  S
1861    CCTGGAATTT TTGCATATCA CTTTTCAGTG CTTCTTTGCT GCATTCTACC TCGCCCTCAG
        ·A  D  T   P  P  S  S   L  R  H   L  F  Q   D  H  R  P   E  S  S
1921    TGCCGACACC CCGCCATCCT CGCTCAGACA TCTCTTCCAA GATCACAGGC CTGAAAGCTC
        ·P  L  A   R  V  L  P   K  L  F   L  R  G   S  R  C  R   E  G  S
1981    GCCACTGGCC AGGGTGCTGC CCAAATTGTT CCTGCGGGGC TCCCGATGCA GAGAGGGCAG
        ·V  A  A   L  L  Q  G   A  E  P   H  N  L   Q  I  T  G   A  F  L
2041    CGTGGCTGCT TTGCTGCAGG GGGCCGAGCC GCACAACCTC CAGATCACAG GGGCCTTCCT
        ·A  G  L   L  S  Q  E   H  R  S   L  L  A   E  C  Q  A   S  E  T
2101    GGCGGGGCTG TTGTCACAGG AGCACCGGAG CTTGCTGGCG GAGTGCCAGG CCTCTGAGAC
        ·A  L  L   R  R  W  D   C  V  R   R  C  L   T  R  S  L   R  E  H
2161    GGCCCTGCTC CGGCGCTGGG ATTGTGTCCG GCGGTGTCTG ACCCGCAGCC TCCGCGAGCA
        ·F  R  S   I  P  P  A   L  P  G   E  A  K   S  M  H  A   L  P  G
2221    TTTCCGCTCC ATCCCACCCG CCTTGCCGGG TGAGGCCAAG AGCATGCACG CCCTGCCTGG
        ·F  L  W   L  I  R  S   L  Y  E   M  Q  E   E  R  L  A   R  E  A
2281    CTTCCTCTGG CTCATCCGGA GCCTGTATGA GATGCAGGAG GAGCGACTGG CGCGGGAGGC
        ·V  C  R   L  N  V  G   H  L  K   L  T  F   C  G  V  G   P  A  E
2341    CGTTTGCAGG CTGAACGTTG GGCACCTCAA GCTGACCTTC TGCGGTGTGG GCCCGGCCGA
        ·C  A  A   L  A  F  V   L  R  H   L  R  R   P  V  A  L   Q  L  D
2401    GTGTGCTGCC CTGGCCTTCG TGCTGCGCCA CCTCCGGCGG CCTGTGGCCC TGCAGCTGGA
        ·H  N  S   V  G  D  I   G  V  E   Q  L  L   P  C  L  G   V  C  K
2461    CCACAACTCT GTGGGCGACA TCGGCGTGGA GCAGCTGCTG CCTTGCCTGG GCGTCTGCAA
        ·A  L  Y   L  R  D  N   N  I  S   D  R  G   I  C  K  L   V  E  H
2521    AGCTCTTTAC TTGCGAGATA ACAACATCTC AGACCGAGGC ATCTGCAAGT TGGTTGAACA
        ·A  L  R   C  E  Q  L   Q  K  L   A  L  F   N  N  K  L   T  D  G
2581    TGCTCTTCGC TGTGAGCAGC TGCAGAAGTT AGCTCTTTTC AACAACAAGT TGACCGATGG
        ·C  A  H   S  M  A  R   L  L  A   C  K  Q   N  F  L  A   L  R  L
2641    CTGTGCACAC TCCATGGCCA GGCTCCTTGC GTGCAAGCAG AACTTCTTGG CTTTGAGGCT
        ·G  N  N   H  I  T  A   A  G  A   E  V  L   A  Q  G  L   R  T  N
2701    AGGAAACAAC CACATCACGG CTGCGGGAGC CGAGGTGCTT GCCCAGGGGC TCAGAACTAA
        ·N  S  L   Q  F  L  G   F  W  G   N  Q  V   G  D  E  G   A  Q  A
2761    CAACTCCTTG CAGTTTTTGG GGTTCTGGGG CAACCAGGTG GGTGACGAGG GGGCCCAGGC
        ·L  A  A   A  L  G  D   H  Q  S   L  R  W   L  S  L  V   G  N  N
2821    CTTGGCTGCA GCCTTGGGTG ATCACCAGAG CTTGAGGTGG CTCAGCCTGG TGGGGAACAA
        ·I  G  S   V  G  A  Q   A  L  A   L  M  L   E  K  N  M   A  L  E
2881    CATTGGCAGC GTGGGTGCTC AAGCCTTAGC ATTGATGTTG GAAAAGAATA TGGCCCTGGA
        ·E  L  C   L  E  E  N   H  V  Q   D  E  G   V  C  F  L   A  K  G
2941    AGAACTCTGC CTGGAGGAGA ACCACGTCCA GGATGAAGGT GTGTGTTTCC TCGCCAAAGG
        ·L  A  R   N  S  S  L   K  V  L   K  L  S   N  N  H  I   S  S  L
3001    ACTTGCAAGA AACTCAAGTC TGAAAGTCCT GAAGCTGTCT AACAACCACA TCAGCTCCCT
        ·G  A  E   A  L  L  R   A  L  E   K  N  D   T  I  L  E   V  W  L
3061    AGGGGCAGAG GCCCTCCTGC GGGCCCTTGA AAAGAATGAC ACCATTCTGG AAGTCTGGCT
        ·R  G  N   T  F  S  P   E  E  I   E  K  L   S  H  Q  D   T  R  L
3121    CCGAGGAAAC ACTTTCTCTC CGGAGGAAAT TGAGAAACTC AGCCACCAGG ATACCAGACT
        ·L  L  *
3181    CTTGCTCTGA TGTTTCCAGG CCAGTGTTCA GCTCAGTGTG TTTGGGAGGA GGCCATTGGT
3241    TTGGATCCCA GGATGGGACG ACATCTGAGC ACAGCCCACT CAGATGGAAC CTGGATCTGC
3301    CCAGGGCCAA CCCAATAGGT CACCTTTGTT CTGGCACAGG AAAGCACATC AGTGCCCTGT
3361    GGAGTAGACT TCACTGAATC CCAACTTTGC CATCAACTTC TTGCCAAGAT TCAATCCTGG
3421    GATGTTGAAG AGGGGCAGCC TGCCTGTACA GGATGGGGCT GGTCTCAAGT CAAGCTGACA
3481    TGCGTCAGGG AGGCCCATGG ATGCCACTGA GTATTTATGG GTGTGGAGAG CTCCCCACGA
```

Figure 4. Continued.

```
3541    GGAGGGATGC TCGGGAAGTA ACTGTTTGCT TTGTCTTAGC TCATGGTCAT CCATCAGGTT
3601    GAGTGGTTCG TCCACTCATC CTGGACCCTA CACATGGCAC TTCCTCTGTG GTTGAGATTC
3661    GGAATGTAGG CATTCTCACC TTTGATCTTT CCCTTGTACC CTGGCCCTGC TCCCACCCTC
3721    CCATCCGCCT CAACCCTCCC CCCATCCAGG GTGGGAGGGG CTACAACTCA CCCTGCTCTC
3781    CTTCTGGTAC TTAGGACAGT ATTGAAAGGG GACAGGTGAC ATACATGTGT TCCTCAAGAC
3841    ATTCTAGAGT TTCAAGAAAA ATATGACTGC CCAGCAACTG GACTTTTATT TCCAGTGAAA
3901    TCAATTACTC TTCAGTTAAA CCTTTGGGAA CAACTCTGTA TCCAAATGCA ACTTTTAAAA
3961    CTAACCTAGG CCAGAATTTT GAACAGCCCC ACCAGGTCTC TGAAGCCTGT GAACTGAACT
4021    CTGGCAGCAG ACTTCCAAAA TATATTCATA AGAGATGGTT TGGTTTTGTT TGTGCCAGGC
4081    CACTTTAGGA TATAAAGTTA TAGATCAAAA GTTTACAGGG CAAAATCAAA GGCCCTTCCT
4141    TACAAAACAA ATGTTTTTCT CTGAATTTTT CAGAAGCTTC TGTAAACTGT CAGGTACTGT
4201    GCAAGTGTTA TTATTTCAAC ACTGTTATTT GTGAAAAACT GGTTAATGTT TATAAACCAC
4261    TTTGTTTTAT TCTCCCTAGT TCATGATTTT ATAAAAAAAA AAAAAATGAC CATGAATGTT
4321    ATGCTGTAAA TAATCACAGA AGATAAAACT ATTGAGTCAC CAGAACTATC TTCATTGTGA
4381    CCAAACACAA TGAAGTATTT AAAAATACTC TGAACATTAT CACATATTAA AGCACAATAT
4441    TCTCCTTGAA GGGAGGAGAC ATGATGTTTC AACCAGATAA TTGATTGCTT AAGGCACAAG
4501    CAGTGTTTAG AAATAGCCTC GCAATCAAAA CACATTTGGC TTCAGTTTAG AGAAGTCTAG
4561    CCCAGCGTGG AGTTGTAAGT TATAGAGGAA CCTCAGTGTC CCGGCAGAAA CACAGATGAG
4621    AGAGACGCAA GCAGGCCCCT GGGCCTCCCT CCATTCTCTC CAAGTGTCTC CAGGGGAGAA
4681    GGATGGAGAA GACTGGGGAA CAGTTCTCCT CTGCAAGCAG CCTCGTGGGT AGGCCTTGGT
4741    GAAATAATTC TTAGCTGAAT TTAATTAGCA AGGACTCAGG TGGCTGCTCA TCAAGGTAGA
4801    ATCGGCTTCC TTGAATGGTT TCCTGTGTGT CTGGTTGGTT AAATACTGTG GCATCTCCCT
4861    GGGCGCCTCC CCAGTAAGGG CATGTGTGTG GGTTCTCTTC ATTTGTTTGA ATTTATTTTA
4921    TTGAGGTATG GTTGATTTAC AGCGTTGTGT TAATTTCTAT ACAGCTGAAT GATTCTGTTA
4981    TACATATGTG TACATACTCT TTTTCCATTA TGCTTTTTCA CAGGATAAAC TACTCATAGT
5041    TCCTTGTGCT ATACATTAGG ACTTTTTTAT GTATCCATTC AATATATAAT ACATTTGCTA
5101    ACCTC
```

Figure 4. Continued.

used in direct sequencing reactions to obtain flanking sequence for exon 3. Primer pairs were then designed to amplify intron 2 and intron 3 by using the flanking sequence. PCR was performed and intron size was estimated as the size of the fragment based on molecular weight (Fig. 5) reduced by the known size of the coding sequence included in each amplicon. This method was used for all introns except intron 1 and intron 5 for which the exact intron size of 1184 bp and 210 bp, respectively, is reported based on sequence data. The estimates for the remaining introns are: intron 2, 2400 bp; intron 3, 3150 bp; intron 4, 5350 bp; intron 6, 2950 bp; intron 7, 1950 bp; intron 8, 1025 bp; intron 9, 1750 bp; intron 10, 3600 bp and intron 11, 1500 bp. The PCR products containing each intron were partially sequenced from both directions to obtain intron/exon boundary sequence information (Table 5). The exon and flanking sequence data have been submitted to GenBank as accession numbers AY518738-AY518748. The combined sequencing results demonstrate that the bovine CARD15 gene is comprised of 12 exons, 11 of which are translated, and that the gene spans approximately 30 kB (Fig. 6).

The expression of bovine CARD15 is greatest in peripheral blood leukocytes. Bovine CARD15 mRNA was detected in all tissues examined with the exception of small intestine. The expression profile shows a 4.75 fold difference in expression between peripheral blood leukocytes and liver, the latter being arbitrarily chosen as a calibrator for the analysis (Table 6; Fig. 7). Thyroid was the only other analyzed tissue in which the expression of CARD15 mRNA exceeded that of liver. These results are consistent with the results of Ogura et al. (2001b) in human and in mouse (2003) which found that CARD15 is expressed most abundantly in peripheral blood leukocytes.

Discussion

The bovine homolog of the human CARD15 gene was sequenced. The gene transcript is 5105 bp and is translated into a protein of 1013 amino acids. By somatic hybrid cell and radiation hybrid mapping, bovine CARD15 was unequivocally localized

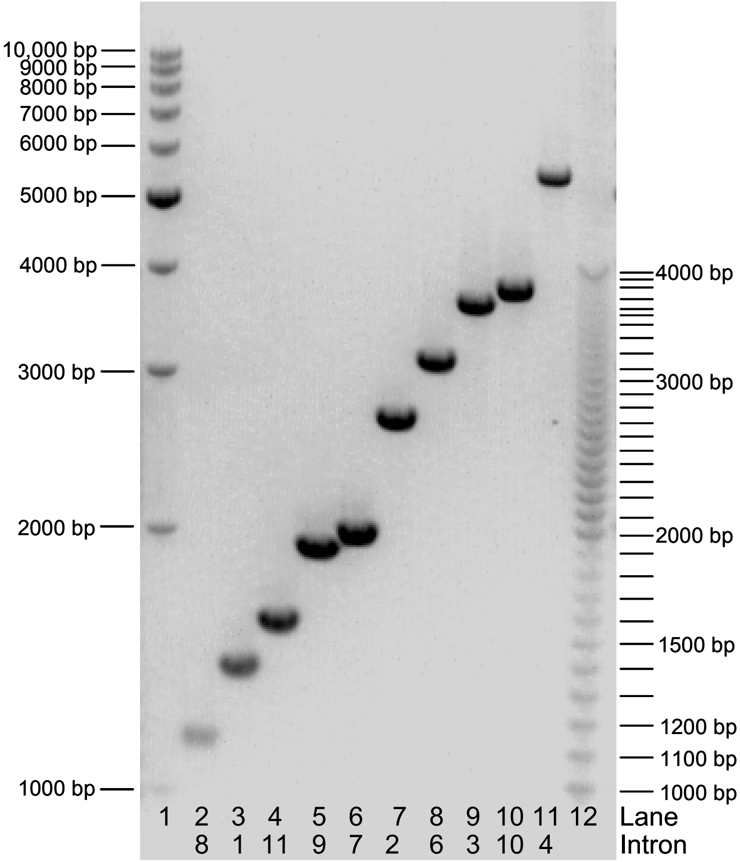


Figure 5. Bovine CARD15 intron amplicons sized by gel electrophoresis. Intron amplicons are ordered according to molecular weight in order to aid in the estimation of amplicon size. Lane 1 contains a 1 kb DNA step ladder (Promega Corp., Madison, Wisc.). Lane 12 contains a 100 bp step ladder (Promega Corp.).

Table 5. Intron and exon sizes and boundary sequences for bovine CARD15

Exons			*Introns*		*Splicing sites*			
No.	*cDNA location (bp)*	*Size (bp)*	*No.*	*Approx. Size (bp)*	*Donor sites*		*Acceptor sites*	
					Exon	*Intron*	*Intron*	*Exon*
1	1-140	140	1	1184[a]			cctcccag	ATTGT
2	141-607	467	2	2400	AGCGG	gtaagcac	ccgcacag	GCAAG
3	608-713	106	3	3150	TGAAG	gtatatat	tcttccag	ATGCT
4	714-2529	1816	4	5350	CTTTA	gtgagtga	ttgtttag	CTTGC
5	2530-2613	84	5	210[a]	TTAGC	gtaagtca	ctttccag	TCTTT
6	2614-2697	84	6	2950	TTGAG	gtgagcct	ccttccag	GCTAG
7	2698-2781	84	7	1950	TTGGG	gtaggtgg	actttcag	GTTCT
8	2782-2865	84	8	1025	CTCAG	gtaagcct	acatccag	CCTGG
9	2866-2949	84	9	1750	CTCTG	gtgagttt	cttgccag	CCTGG
10	2950-3033	84	10	3600	CTGAA	gtaaggaa	ctctctag	GCTGT
11	3034-3117	84	11	1500	GTCTG	gtaagatc		
12	3118-5105	1988						

[a] Indicates the exact intron size.

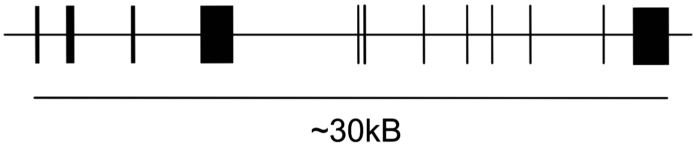


Figure 6. Genomic structure of bovine CARD15. Boxes represent the relative sizes and positions of exons within the genomic sequence.

Table 6. Bovine CARD15 expression by tissue using the comparative C_T method

Tissue	*CARD15 Avg. C_T* [a]	*Beta-actin Avg. C_T*	*ΔC_T* [b]	*$\Delta\Delta C_T$* [c]	*$2^{-\Delta\Delta C_T}$*
PBL[d]	28.98±0.03	32.59±0.05	-3.61±0.02	-2.25	4.75
Kidney	31.98±0.14	32.61±0.13	-0.63±0.01	0.73	0.60
Thyroid	32.30±0.23	33.96±0.41	-1.66±0.31	-0.3	1.23
Liver	30.45±0.50	31.81±0.34	-1.36±0.18	0	1.00
Adrenal	31.28±0.26	28.79±0.08	2.49±0.22	3.85	0.07
Lung	33.30±0.47	27.83±0.31	5.48±0.51	6.83	0.01
Anterior Pituitary	32.22±0.22	32.28±0.10	-0.06±0.17	1.3	0.41
Pancreas	31.24±0.26	30.49±0.03	0.75±0.23	2.10	0.23
Bone Marrow	33.63±0.72	30.22±0.16	3.41±0.64	4.77	0.04
Muscle	34.53±0.85	30.04±0.32	4.49±0.72	5.85	0.02
Heart	32.84±0.94	33.04±0.33	-0.20±0.79	1.16	0.45
Spleen	43.04±0.04	25.71±0.01	17.33±0.02	18.68	0.00
Hypothalamus	32.84±0.66	33.47±0.38	-0.63±0.40	0.73	0.60
Small Intestine	53.05±2.76	25.50±0.05	27.55±2.72	28.91	0.00

[a] C_T = Cycle Threshold: Cycle number where amplification exceeds the threshold determined by the geometric portion of the amplification curve.

[b] ΔC_T = Averaged CARD15 C_T – Beta-actin C_T : Normalization of RT-PCR cycles for CARD15 target to beta-actin house keeping gene.

[c] $\Delta\Delta C_T$ = Mean ΔC_T – Liver mean ΔC_T value: Liver was used as a calibrator to set the baseline for comparing mean differences in ΔC_T values across all tissues.

[d] Peripheral Blood Leukocytes.

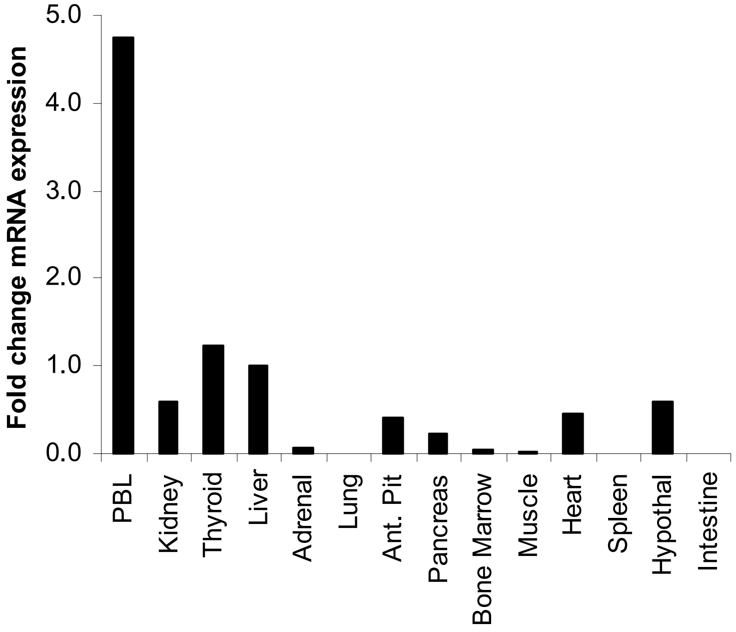


Figure 7. Bovine CARD15 mRNA expression profiles by tissue type. PBL = Peripheral Blood Leukocytes. Ant. Pit. = Anterior Pituitary.

to the region of BTA18 that is homologous to HSA16q21.1 and which harbors human CARD15. However, attempts to order bovine CARD15 within the bovine RH map revealed two possibilities with high levels of statistical support. The most likely order (Fig. 1) results in a rearrangement when compared to the homologous region in human, while the next best order results in conserved synteny with HSA16q21.1.

The gene order in human is ADCY7, CARD15, CYLD and MMP2. A comparative mapping project of BTA18 to homologous human and mouse chromosomes (Goldammer et al. 2002) revealed a small gap between linkage groups 1 and 2 in human that spans approximately 8 cR_{3000} on the human GB4 WGRH cR_{3000}. The most distal marker in group 1 is ADCY7 and the most proximal marker in group 2 is MMP2. The physical distance between these 2 genes in human is approximately 4.7 Mb (http://genome.ucsc.edu/cgi-bin/hgGateway?org=human). In the cattle RH map produced by Goldammer et al. (2002) ADCY7 is the most distal marker in linkage group 2 and MMP2 is the most proximal marker in linkage group 3. The distance between the two linkage groups was estimated to be 42 cR_{5000}. In this study, CYLD was identified within the same BAC containing bovine CARD15. Although not present in the human or cattle RH maps, CYLD is located approximately 45 kb from CARD15 in human. To test the hypothesis that all loci within this region possess conserved synteny between human and cattle and that the second most likely locus order was, in fact, the correct gene order, a set of primers was developed to amplify the bovine ADCY7 gene and used to screen two bovine BACs known to harbor CARD15 for the presence of ADCY7. Neither of these BACs were identified as containing ADCY7. Since one BAC was known to contain both bovine CARD15 and CYLD, the second most likely order obtained from the RH mapping is, in fact, the correct gene order and thus gene order is completely conserved among these genes between human and cattle. Since the retention frequency of the bovine CARD15 probe in the bovine SHC panel was 0.156 and considerably lower than the average panel retention frequency of 0.227, the greater likelihood for the incorrect locus order is a statistical artifact due, no doubt, to the small physical distance separating the loci.

Considerable difficulty was experienced in obtaining the sequence of the target gene even when the complete sequence for the human homolog was known. This was due in part to the complexity of structure of the target gene, but also because sequence for certain regulatory regions including the 5'UTR was important. Even though the coding regions of most protein coding genes are quite highly conserved across species, the development of primers from sequence in one species for use in another species is often a tenuous proposition. Clearly, the closer the two species are evolutionarily the greater are the chances for successful primer development. The likelihood of success of this approach may be further enhanced if sequence is available from two, or more, divergent species since primers may be designed from regions that are highly conserved among all species. Of course, this approach assumes that there is sufficient conservation of sequence identity among the species for the genomic regions that are to be sequenced from the target species. This will not always be the case for coding regions, and is unlikely to be the case for noncoding regions such as introns, 5'UTR and 3'UTR.

In this case, the strategy proved ineffective. The structure of the human CARD15 gene is reasonably complex comprising 12 exons, many of them short, thus allowing few opportunities for stretches of sequence with high levels of homology between human and mouse from which to design primers. Furthermore, it is a challenge to develop assays to amplify products for which size is unknown in the target species. Thus, the lack of success with this approach may have been due either to the unknown size of the target amplicon or with the sequence content of the primers themselves. In order to generate coding sequence, bovine cDNA clones were completely sequenced, from which bovine EST sequences had been produced and that had homology to human CARD15. This strategy allowed the generation of sequence data for the region of the bovine CARD15 gene representing a portion of exon 4 through to the 3'UTR. Suitable primer pairs were then developed to amplify across introns of the gene. To capture sequence from the 5' end of the gene, BACs were isolated harboring the cattle CARD15 homolog and then used to obtain sequence data by developing primers based on the human to mouse alignment. Contrary to the initial results, this primer development strategy was successful in obtaining sequence data for exons 2, 3 and the missing portion of exon 4. The success in this region of the gene can be attributed to the larger size of these exons and the fact that only one suitable primer had to be developed to prime the sequencing reactions using BAC DNA as a template. Although this method did produce sequence data, the primers were not successful in amplifying across introns in this region of the gene. However, the sequence data obtained from the flanking regions of exon 3 was used to develop primer pairs that would amplify across introns 2 and 3.

Exon 1 which includes the 5'UTR had virtually no homology between human and mouse and so no suitable primer pairs could be developed using the comparative sequence approach to acquire sequence data for this region of the bovine gene. Thus, 5' and 3' RACE was performed to obtain the bovine sequence for the 5'UTR and the remainder of the 3'UTR that was not included in the cDNA clones. While this approach was successful for generating sequence in the 5' region of bovine CARD15, the presence of a mononucleotide A_{14} repeat located approximately 1100 bases into the bovine 3'UTR and approximately 800 bases upstream of the actual PolyA site defeated the attempt to capture the complete 3'UTR sequence by this approach. The 3' RACE approach utilizes an adapter that ligates to the PolyA tail of the transcript which is then used to prime the subsequent amplification steps. In this application, the adapter preferentially ligated to the internal PolyA repeat site effectively isolating the remainder of the 3'UTR from the sequencing attempt. Fortunately, a recent deposit by the USDA MARC to the GenBank bovine EST sequence database was found to be homologous to the 3'UTR of the bovine CARD15 gene. The new sequence included the PolyA repeat as well as additional downstream sequence. The cDNA clone was sequenced from which the EST sequence was derived and thus the complete sequence for the entire 3'UTR of the bovine CARD15 gene was acquired.

While the difficulties encountered in obtaining sequence data for a target gene in a sequence-poor species are not unique, these experiences demonstrate the considerable resource cost that can be faced by researchers investigating genomic diversity in sequence-poor species. Fortunately, bovine has been identified as a high priority species by the NHGRI for sequencing and an 8× deep sequence coverage should be produced, assembled and annotated within the next two years. The rapid increase in availability of whole genome sequence assemblies should alleviate some of these problems for researchers in the future.

Overall, these data provide comprehensive sequence information for the bovine CARD15 gene which may provide a basis for future comparative studies of the gene. The map location of bovine CARD15 is presented which may prove beneficial for QTL or candidate gene studies targeted at identifying regions of the bovine genome harboring genes associated with disease resistance or susceptibility.

COMPARATIVE SEQUENCE ANALYSIS OF CARD15

Introduction

Comparative genomics is increasingly becoming an important tool for deciphering the information contained in the completed genome sequences of both closely and distantly related species and taxa. Previously, comparative genomics has played a large role in gene discovery using information from sequence-rich species to obtain sequence for application in sequence-poor species. It has also been extensively used to analyze the effects of genome organization in a multitude of species. The basic premise of comparative genomics assumes that features that are common between two organisms will likely reflect features encoded in the DNA that are conserved between the species. This includes both sequences that encode proteins and various RNAs as well as the DNA sequences that regulate the expression of coding genes. Similarly, features that differ between species should reflect a divergence in the DNA sequences that encode or control the expression of the proteins that are responsible for variation in these features.

With the current availability of whole genome sequences for a number of species, it is now possible to align complete genome sequences and begin to address the important questions concerning the roles of elements of the DNA sequence. It has long been recognized that regions of noncoding DNA with high levels of homology between divergent species are excellent candidates for being functional regions (Hardison et al. 1997). It has further been shown that the comparison of genomic sequences from more than two species provides even greater resolution for distinguishing regulatory regions (Thomas et al. 2003) since the likelihood of conservation of non-functional sequence is reduced exponentially assuming that the region is free to evolve independently among the compared species.

Two groups have now examined the structure and function of the Card15 gene in mouse and both studies have concluded a conserved role for the gene in human and mouse NF-kB activation (Iwanaga et al. 2003; Ogura et al. 2003). In both human and mouse, CARD15 has been shown to respond to bacterial components (LPS, PGN) (Ogura et al. 2001b; Ogura et al. 2003) and mutations in the gene that hinder responsiveness to bacterial components have been associated with susceptibility to Crohn's disease. However, there is currently no information available concerning the cis- or trans-acting elements that regulate the CARD15 gene. Plant R genes that are similar to CARD15 contain leucine rich repeats (LRR) which vary in copy number and provide protection/recognition of diverse types of bacteria. These genes have been extensively studied in terms of their structural organization, sequence evolution and genome distribution; however, information on the regulation of these genes is also limited.

An improved understanding of the elements responsible for the regulation of CARD15 should prove to be useful for elucidating the function of CARD15 in disease susceptibility. In this study, the bovine CARD15 sequence was exploited and a three species comparative analysis was performed in the hope of gaining further insight into the regulation of the gene. Results are provided from the alignment of the complete transcript of the gene in human, mouse and bovine. A comparison of the intronic sequences which closely flank exons conserved across the three species is also included. Finally, CARD15 sequence for *Bos indicus* and *Bison bison* is reported and a comparative analysis between the subspecies *Bos taurus* and *Bos indicus* and the closely related *Bison bison* is included.

Materials and Methods

Sequence alignment and analysis. All alignments were constructed using AlignX in the Vector NTI suite (Informax, Frederick, Mar.). Alignments for the comparative analysis used published sequence data available for human, mouse and bovine CARD15 (Accession numbers AF178930, AF520774, AY518737). Nucleotide and amino acid identities were computed using pairwise alignments. Intron sizes from human and mouse were calculated by aligning full length transcript sequences to the genome sequence.

Protein domain identification. To identify protein domains within the bovine gene, bovine sequence was analyzed using two domain analysis programs AnDom (http://www.bork.embl-heidelberg.de/AnDom; Schmidt et al. 2002) and ProDom (http://www.toulouse.inra.fr/prodom.html, Sonnhammer and Kahn 1994; Corpet et al. 2000). Additionally, the bovine CARD15 gene was aligned with the domains identified in human and mouse (Ogura et al. 2001a; Iwanaga et al. 2003; Ogura et al. 2003).

Identification of putative regulatory regions. The bovine sequence was searched for regulatory motifs in the 5'UTR and 3'UTR using the program UTRscan (http://www.ba.itb.cnr.it/BIG/UTRScan/; Pesole and Liuni 1999). The 5' and 3' flanking intronic regions for ~200 bp each side of each CARD15 exon were also aligned. Short motifs identified as being conserved between the three species in these intronic regions as well as in the 5'UTR and 3'UTR were then analyzed using the TFSCAN (http://zeon.well.ox.ac.uk/git-bin/tfscan) and NSITE (available through SoftBerry http://www.softberry.com/berry.phtml?topic=promoter) programs to identify putative regulatory motifs. Motifs selected for analysis required homology consisting of 6 or more bases with no more than 2 substitutions among the 3 species.

Bos indicus *and* Bison bison *CARD15 sequence.* Using the flanking intronic sequence that was previously obtained from the amplified intronic sequence generated for intron sizing, primers were designed to amplify each exon in its entirety (Table 7). These primer sets were used to amplify all exonic and some flanking intronic portions of CARD15 in one *indicus* and one *bison* animal. The PCR reactions were performed in a 50 μl reaction volume: 15 mM Tris, 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM of each dNTP (dATP, dCTP, dGTP and dTTP), 1 μM each primer, 50 ng DNA, 0.5 units of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, Calif.). The general PCR reaction included an initial hot start of 10 min at 95°C followed by 35 cycles for 30 s at 94°C, 30 s at the optimized annealing temperature and 45 s (or 1 min see Table 7) at 72°C, followed by a final elongation step at 72°C for 7 min. The PCR products were

Table 7. Primers used to amplify the exons of bovine CARD15

Exon	*Sense 5′ to 3’*	*Antisense 5′ to 3′*	*Anneal. Temp.*[a]	*Approx. product size*
1	ATGTCCCTGTTTTGTAGACAGACG	CAGAACACTGCTGTGAAGATGC	58/1min ext	900
2	CTCAGTTTGAACACCTGTACAATGG	AGGAGGACTATGACCCACATCTC	52	690
3	GGCCTTTTATTGTGGTGGAA	TGCTGATGTTTTTGCTCTTCA	55	300
4.1	TGGGTTCCTACCTGCAAAAC	CACTGAGGGCGAGGTAGAAT	52/1min ext	2250
4.2	CAGTCCAGAGTCTGCTCTTCAAC	GCCATTCCATGAATTTCAACTATC	55/1min ext	1360
5-6	GTCTCTGTGGGGGTTTTGTC	GTCCTTGTTTTCAGCGAGGT	55	500
7	GGGAGCAGTAAGGGTCCTC	CAGAGATCTTGGGGCTGAAG	55	300
8	CACTTGCTGGGACCTGAGT	CCCTCCTCACACTGGCTTC	53	200
9	GCATTTTGCCCTTCTTGAGT	ACGCAGTCATCCATCTTGGT	53	200
10	GGGCACATGGGTTCATCTT	CCCTCTCAAGGCCAATCAT	55	200
11	CCAGCTCCCAAAGTCTCCTT	GAGGCTCAGAGAGGTTAAAGAGG	53	200
12.1	AGGTTTACAAAGCAGCATCTTCC	ATGTCACCTGTCCCCTTTCA	58/1min ext	750
12.2	CACCTTTGATCTTTCCCTTGTACC	CCCTGGAGACACTTGGAGAG	56/1min ext	1000
12.3	CCAGCGTGGAGTTGTAAGTTATAG	AGACAAAGGACACAGAGACCAGAC	58/1min ext	700

[a] Extension (ext) time was 45 s unless otherwise indicated.

analyzed by electrophoresis on 2% agarose gels in 1.0× TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide. The PCR product was purified using QIAquick® (Qiagen, Valencia, Calif.) spin columns.

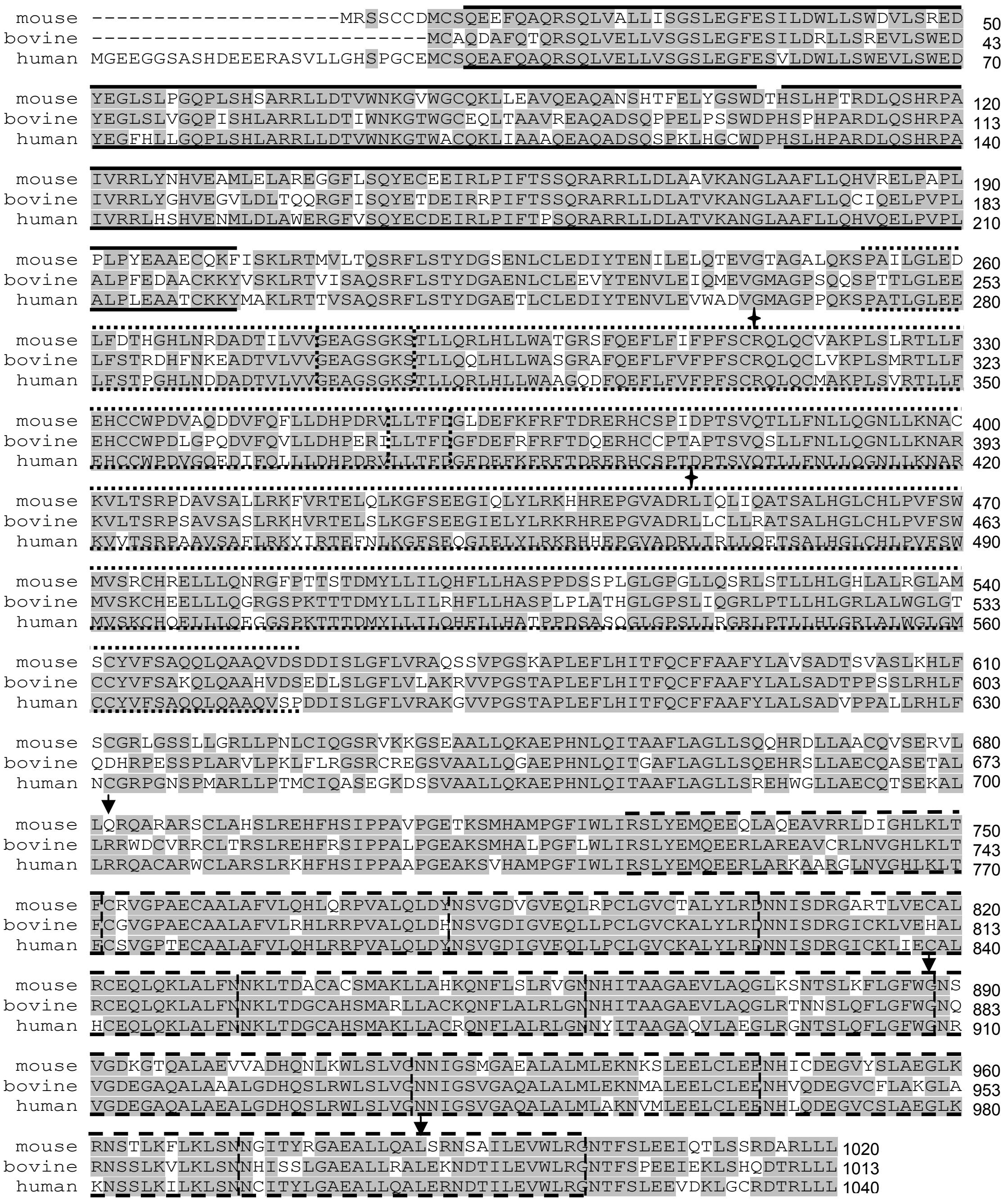
Cycle sequence reactions were performed in a 10 µl reaction volume: 1 µl Big Dye v3.0 (Applied Biosystems), 10 ng amplified DNA, 0.5 µl of 5% dimethylsulfoxide (DMSO), 0.5 µl of 20 µM primer. The cycle sequence reaction included an initial denaturation step of 2 min at 96°C, 35 cycles for 15 s at 96°C, 20 s at 50°C and 4 min at 60°C, followed by a final elongation step at 60°C for 15 min. Unincorporated nucleotides and primers were removed from the reactions using a BioMax® Spin-50 mini-column (Millipore Corp., Bedford, Mass.). Sequencing reactions were run on an ABI 3100 DNA analyzer (Applied Biosystems). The resulting sequence was aligned with *Bos taurus* CARD15 sequence (Accession number AY518737) and nucleotide and amino acid homology was inferred by pairwise comparison.

Results

Bovine CARD15 is highly homologous to human and mouse counterparts. The bovine CARD15 mRNA is 5105 base pairs and the protein it encodes is comprised of 1013 amino acids (Fig. 4). This compares to the human CARD15 gene with an mRNA length of 4485 bp and the mouse Card15 gene with an mRNA length of 4585 bp. The number and size of exons are completely conserved between human, mouse and bovine with the exception of exon 1 (Table 8). There is some variation in intron size which is primarily responsible for the variability in the genomic size of CARD15 among the 3 species (Table 8). Domain analysis of the bovine CARD15 gene and alignment with its mouse and human homologs, revealed that the bovine gene is comprised of two N-terminal caspase recruitment domains (CARD) (residues 4-93; 96-191), one centrally located nucleotide binding oligomerization domain (NOD) (residues 243-548) and 10 tandem LRR at the C-terminus (residues 715-992). A representation of the domain structure of bovine CARD15 in relation to the human and mouse homologs is presented in Fig. 8. The pairwise amino acid homology between the species for each of the

Table 8. Genomic structure of CARD15 in human, mouse and bovine

Chromosome	*HSA16*		*MMU8*		*BTA18*	
Genomic size kb	~36		~39		~30	
Transcript length bp:	4485		4585		5105	
Protein length aa:	1040/1013		1020/1013		1013	
No.	Exon	Intron	Exon	Intron	Exon	Intron
1	178	2169	110	1415	140	1184
2	467	7899	467	7342	467	~2400
3	106	2596	106	3053	106	~3150
4	1816	4192	1816	5239	1816	~5350
5	84	221	84	169	84	210
6	84	2948	84	1926	84	~2950
7	84	2613	84	2145	84	~1950
8	84	592	84	456	84	~1025
9	84	2104	84	5160	84	~1750
10	84	4242	84	5531	84	~3600
11	84	1844	84	2500	84	~1500
12	1330		1498		1988	



```
 mouse --------------------MRSSCCDMCSQEEFQAQRSQLVALLISGSLEGFESILDWLLSWDVLSRED   50
bovine ---------------------------MCAQDAFQTQRSQLVELLVSGSLEGFESILDRLLSREVLSWED   43
 human MGEEGGSASHDEEERASVLLGHSPGCEMCSQEAFQAQRSQLVELLVSGSLEGFESVLDWLLSWEVLSWED   70

 mouse YEGLSLPGQPLSHSARRLLDTVWNKGVWGCQKLLEAVQEAQANSHTFELYGSWDTHSLHPTRDLQSHRPA  120
bovine YEGLSLVGQPISHLARRLLDTIWNKGTWGCEQLTAAVREAQADSQPPELPSSWDPHSPHPARDLQSHRPA  113
 human YEGFHLLGQPLSHLARRLLDTVWNKGTWACQKLIAAAQEAQADSQSPKLHGCWDPHSLHPARDLQSHRPA  140

 mouse IVRRLYNHVEAMLELAREGGFLSQYECEEIRLPIFTSSQRARRLLDLAAVKANGLAAFLLQHVRELPAPL  190
bovine IVRRLYGHVEGVLDLTQQRGFISQYETDEIRRPIFTSSQRARRLLDLATVKANGLAAFLLQCIQELPVPL  183
 human IVRRLHSHVENMLDLAWERGFVSQYECDEIRLPIFTPSQRARRLLDLATVKANGLAAFLLQHVQELPVPL  210

 mouse PLPYEAAECQKFISKLRTMVLTQSRFLSTYDGSENLCLEDIYTENILELQTEVGTAGALQKSPAILGLED  260
bovine ALPFEDAACKKYVSKLRTVISAQSRFLSTYDGAENLCLEEVYTENVLEIQMEVGMAGPSQQSPTTLGLEE  253
 human ALPLEAATCKKYMAKLRTTVSAQSRFLSTYDGAETLCLEDIYTENVLEVWADVGMAGPPQKSPATLGLEE  280

 mouse LFDTHGHLNRDADTILVVGEAGSGKSTLLQRLHLLWATGRSFQEFLFIFPFSCRQLQCVAKPLSLRTLLF  330
bovine LFSTRDHFNKEADTVLVVGEAGSGKSTLLQQLHLLWASGRAFQEFLFVFPFSCRQLQCLVKPLSMRTLLF  323
 human LFSTPGHLNDDADTVLVVGEAGSGKSTLLQRLHLLWAAGQDFQEFLFVFPFSCRQLQCMAKPLSVRTLLF  350

 mouse EHCCWPDVAQDDVFQFLLDHPDRVLLTFDGLDEFKFRFTDRERHCSPIDPTSVQTLLFNLLQGNLLKNAC  400
bovine EHCCWPDLGPQDVFQVLLDHPERILLTFDGFDEFRFRFTDQERHCCPTAPTSVQSLLFNLLQGNLLKNAR  393
 human EHCCWPDVGQEDIFQLLLDHPDRVLLTFDGFDEFKFRFTDRERHCSPTDPTSVQTLLFNLLQGNLLKNAR  420

 mouse KVLTSRPDAVSALLRKFVRTELQLKGFSEEGIQLYLRKHHREPGVADRLIQLIQATSALHGLCHLPVFSW  470
bovine KVLTSRPSAVSASLRKHVRTELSLKGFSEEGIELYLRKRHREPGVADRLLCLLRATSALHGLCHLPVFSW  463
 human KVVTSRPAAVSAFLRKYIRTEFNLKGFSEQGIELYLRKRHHEPGVADRLIRLLQETSALHGLCHLPVFSW  490

 mouse MVSRCHRELLLQNRGFPTTSTDMYLLILQHFLLHASPPDSSPLGLGPGLLQSRLSTLLHLGHLALRGLAM  540
bovine MVSKCHEELLLQGRGSPKTTTDMYLLILRHFLLHASPLPLATHGLGPSLIQGRLPTLLHLGRLALWGLGT  533
 human MVSKCHQELLLQEGGSPKTTTDMYLLILQHFLLHATPPDSASQGLGPSLLRGRLPTLLHLGRLALWGLGM  560

 mouse SCYVFSAQQLQAAQVDSDDISLGFLVRAQSSVPGSKAPLEFLHITFQCFFAAFYLAVSADTSVASLKHLF  610
bovine CCYVFSAKQLQAAHVDSEDLSLGFLVLAKRVVPGSTAPLEFLHITFQCFFAAFYLALSADTPPSSLRHLF  603
 human CCYVFSAQQLQAAQVSPDDISLGFLVRAKGVVPGSTAPLEFLHITFQCFFAAFYLALSADVPPALLRHLF  630

 mouse SCGRLGSSLLGRLLPNLCIQGSRVKKGSEAALLQKAEPHNLQITAAFLAGLLSQQHRDLLAACQVSERVL  680
bovine QDHRPESSPLARVLPKLFLRGSRCREGSVAALLQGAEPHNLQITGAFLAGLLSQEHRSLLAECQASETAL  673
 human NCGRPGNSPMARLLPTMCIQASEGKDSSVAALLQKAEPHNLQITAAFLAGLLSREHWGLLAECQTSEKAL  700

 mouse LQRQARARSCLAHSLREHFHSIPPAVPGETKSMHAMPGFIWLIRSLYEMQEEQLAQEAVRRLDIGHLKLT  750
bovine LRRWDCVRRCLTRSLREHFRSIPPALPGEAKSMHALPGFLWLIRSLYEMQEERLAREAVCRLNVGHLKLT  743
 human LRRQACARWCLARSLRKHFHSIPPAAPGEAKSVHAMPGFIWLIRSLYEMQEERLARKAARGLNVGHLKLT  770

 mouse FCRVGPAECAALAFVLQHLQRPVALQLDYNSVGDVGVEQLRPCLGVCTALYLRDNNISDRGARTLVECAL  820
bovine FCGVGPAECAALAFVLRHLRRPVALQLDHNSVGDIGVEQLLPCLGVCKALYLRDNNISDRGICKLVEHAL  813
 human FCSVGPTECAALAFVLQHLRRPVALQLDYNSVGDIGVEQLLPCLGVCKALYLRDNNISDRGICKLIECAL  840

 mouse RCEQLQKLALFNNKLTDACACSMAKLLAHKQNFLSLRVGNNHITAAGAEVLAQGLKSNTSLKFLGFWGNS  890
bovine RCEQLQKLALFNNKLTDGCAHSMARLLACKQNFLALRLGNNHITAAGAEVLAQGLRTNNSLQFLGFWGNQ  883
 human HCEQLQKLALFNNKLTDGCAHSMAKLLACRQNFLALRLGNNYITAAGAQVLAEGLRGNTSLQFLGFWGNR  910

 mouse VGDKGTQALAEVVADHQNLKWLSLVGNNIGSMGAEALALMLEKNKSLEELCLEENHICDEGVYSLAEGLK  960
bovine VGDEGAQALAAALGDHQSLRWLSLVGNNIGSVGAQALALMLEKNMALEELCLEENHVQDEGVCFLAKGLA  953
 human VGDEGAQALAEALGDHQSLRWLSLVGNNIGSVGAQALALMLAKNVMLEELCLEENHLQDEGVCSLAEGLK  980

 mouse RNSTLKFLKLSNNGITYRGAEALLQALSRNSAILEVWLRGNTFSLEEIQTLSSRDARLLL 1020
bovine RNSSLKVLKLSNNHISSLGAEALLRALEKNDTILEVWLRGNTFSPEEIEKLSHQDTRLLL 1013
 human KNSSLKILKLSNNCITYLGAEALLQALERNDTILEVWLRGNTFSLEEVDKLGCRDTRLLL 1040
```

Figure 8. Comparative protein alignment for CARD15. Bold lines represent CARD domains; small dashed lines represent the NOD including the P-loop and Mg^{2+} binding site within boxes; larger dashed lines represent the 10 LRR. Shaded amino acids are identical. Arrow heads indicate amino acids associated with Crohn's disease. Four point stars indicate amino acids associated with Blau Syndrome.

domains is presented in Table 9.

In addition to the CARD15 domain structure, Fig. 8 also represents the protein alignment of the three species' sequences. One notable difference between the species is that both mouse and human have two in-frame translation initiation sites, while the bovine sequence contains only one translation start site. In human, these sites are separated by 81 nucleotides, whereas, in mouse the sites are separated by 21 nucleotides resulting in 2 different protein products consisting of 1040/1013 amino acids in human and 1020/1013 amino acids in mouse. In both species, the second translation initiation site corresponds to the unique bovine translation initiation site.

A comparison of the 1013 amino acid bovine protein product to the corresponding 1013 amino acid human and mouse products revealed an 83.7% and 76.8% homology at the nucleotide level and an 81.2% and 76% homology at the amino acid level, respectively. Additionally, the extent of conservation of the amino acids that have been shown to be associated with susceptibility to Crohn's disease and Blau Syndrome was examined. The consensus sequence of two of the amino acids associated with susceptibility to Crohn's disease has recently been reported as being conserved (G908R, 1007fs) between human and mouse, while that for a third is variable (R702W). The results of this study indicate that the consensus sequences for all three of these amino acids are conserved between cattle and human (Fig. 8). Furthermore, the consensus sequences of both amino acids associated with susceptibility to Blau Syndrome are conserved across all three species (Fig. 8).

The CARD15 5'UTR and 3'UTR are not conserved across species. Pairwise alignment of the 5'UTR revealed only a 35.4% homology between human and bovine, 42.1% homology between mouse and bovine and 33.3% homology between human and mouse (Fig. 9). The 3'UTR alignments also revealed low overall levels of homology, with only small contiguous regions of homology among the 3 species (Fig. 10). Pairwise alignments of just over 1000 bases of sequence from the region of highest homology within the 3'UTR revealed a 44.2% homology between bovine and mouse,

Table 9. Pairwise percent amino acid homology by CARD15 domain

Domain	*No. Amino Acids*	*human/mouse*	*human/bovine*	*bovine/mouse*
CARD1	94	76.6	76.6	72.3
CARD2	96	81.2	83.3	72.9
NOD	305	82.3	82.3	78.4
LRRs	277	81.2	88.1	80.1
LRR1	28	75	85.7	82.1
LRR2	28	89.3	85.7	89.3
LRR3	25	88	100	88
LRR4	28	82.1	89.3	85.7
LRR5	28	78.6	92.9	78.6
LRR6	28	78.6	82.1	85.7
LRR7	28	75	92.9	67.9
LRR8	28	82.1	89.3	85.7
LRR9	28	82.1	78.6	71.4
LRR10	28	82.1	85.7	67.9
No Domain	241	75.1	74.3	70.5

```
 Human    (1) GTAGACAGATCCAGGCTCACCAGTCCTGTGCCACTGGGCTTTTGGCGTTC
 Mouse    (1) ------------------------CCTTTTCTCCGGGTGTACTGGCTGGT
Bovine    (1) --------------------------------------CTCTGGACTCCG

 Human   (51) TGCACAAGGCCTACCCGCAGATGCCATGCCTGCTCCCCCAGCCTAATGGG
 Mouse   (27) TTTGTGTGTCAATTTGACACAGGCTG-GAG--TTATCACAGAGAAAGGAG
Bovine   (13) TGACCATCACAGGAGGGAACCTGCTGAGAGA-TTTCCTCTGCTGCGTGGG

 Human  (101) --CTTTGATGGGGGAAGAGGGTGGTTCAGCCTC-TCACGATGAGGAGGAA
 Mouse   (74) --CTTCAGTTGGGGAAGTGCCTCCATGAGATC----CAGCT---------
Bovine   (62) GACCCTGCCAGGGCTTGGAGCTCTGTGAGATCGCTTCCCACGGACTCCCA
                                                Exon 2
 Human  (148) AGAGCAAGTGTCCTCCTCGGACATTCTCCGGGTTGTGAAATG
 Mouse  (109) -------------------------------GTTGTGACATG
Bovine  (112) GGACCCAGAGTCTGAG--GGCTGAGCCCAGGATTGTGAAATG
```

Figure 9. CARD15 5'UTR alignment for human, mouse and bovine. Bases conserved among all three species are shaded. Translation initiation sites are indicated in bold face.

```
bovine3'   ------------------------------------------------------------
 human3'   ------------------------------------------------------------
 mouse3'   TGTCTCCGTTTGTGAGTGGACTGTAGGGGCCTGGACTCTGGAGGCTGAGTAACATCAGGC

bovine3'   ------------------------------------------------------------
 human3'   ------------------------------------------------------------
 mouse3'   AGAATCCCTCTGCTACGCAGGGCTGGTTTGCTTTTCTGGATGCAGTATAGTCACCTTCTG

bovine3'   ------------------------------------------------------------
 human3'   ------------------------------------------------------------
 mouse3'   TTAGCAGAGAAAGTCACCCCATTGCCGTCTGGAATTGACTTTTCCCGAGGAGTCGTGATG

bovine3'   ------------------------------------------TGTTTCCAGGCCAGTGTT
 human3'   ------------------------------------------AGTCTCCGGGAGGATGTT
 mouse3'   GTTGGTCTTGGTTGTTAACTGCACTGACTTAAGAGAGTCATGAGCCGAGAGGACCGCGTT

bovine3'   ----CAGCTCAGTGTGTTTGGGA-GGAGGCCATTGGTTTGGATCCC--AG-GATGGGACG
 human3'   ----CGTCTCAGTTTGTTTGTGA-GCAGGCTGTGAGTTTGGGCCCC--AGAGGCTGGGTG
 mouse3'   TCTGCCTCTAAAGAGGATCACCATGCAGAATTAGTGACTGGAAGGGGAAAGGCCCTCACT

bovine3'   ACATCTGAG---CACAGCCCACTCAGATGGAACCTGGATCTGCCCAGGGCC---------
 human3'   ACATGTGTT---GGCAGCCTCTTCAAAATGAGCCCTGTCCTGCCTAAGGCTGAACTTGTT
 mouse3'   GCATGTGGGTGACACAATCCTCTCTGGTTGCCCCGTGGGAGGATGATGGAGGAGGAGGAA

bovine3'   -------AACCC---AATAGGTCACC-TTTGTTCTGGCACAGGAAAGCACA--TCAGTGC
 human3'   TTCTGGGAACAC---CATAGGTCACC-TTTATTCTGGCAGAGGAGGGAGCA--TCAGTGC
 mouse3'   GACAGGGTATGCTTGCATATGTGAGCATTTCTTCTGAGTGAGGACAGCTGTGGTTGCTGC

bovine3'   CCTGTGGAGTAGACTTCA--CTGAATCCCAACTTTGCCATC-AACTTCTTGCCAAGATTC
 human3'   CCTCCAGGATAGACTTTT--CCCAAGCCTACTTTTGCCATT-GACTTCTTCCCAAGATTC
 mouse3'   CCTCATGTTTGAACAGCAGACTCCAGCGT-CTTTGGCCATTCAACATGGACCCA----TC

bovine3'   AATCCTGGGATGTTGAAGAGGGGCAGCCTGCCTGTACAGGATGGGGCTGGTCTCAAGTCA
 human3'   AATCCCAGGATGTACAAGGACAGC--CCCTCCTCCATAGTATGGGACTGGCCTCTGCTGA
 mouse3'   A--CCAGTGATGTTTTAGGGAGTTTTCAGGCCTGGAGCTTGATGGAACGTGAGGAGTCGA

bovine3'   AGCTGACATGC----------GTCAGGGAGGCCCATGGATGCC-------ACTGAGTATT
 human3'   TCCTCCCAGGCTTCCGTGTGGGTCAGTGGGGCCCATGGATGTGCTTGTTAACTGAGTGCC
 mouse3'   GCTTCTCCAGCTGAGAAGGTTCTCTGCCTCTCCACTGTTGGACTACTCTATCTGGATGGT

bovine3'   TATGGGTGTGGAG------------------AGCTCCC----CAC---------GAGGAG
 human3'   TTTTGGTGGAGAGGCCCGGC-CTCTCACAAAAGACCCCTTACCACTGCTCTGATGAAGAG
 mouse3'   ACTCCAAGCACTGGATGAGCTCAAGCAGGCCACCCCCACCCCCACCCCCATTACTATGTT

bovine3'   GGATGC--------------TCGGGAAGTAACTGTT--TGCTTTGT--CTTAGCTC---A
 human3'   GAGTACACAGAACACATAATTCAGGAAGCAGCT-TT--CCCCATGT--CTCGACTC---A
 mouse3'   GCTGGCCT----CTGGCTGGTCAAGCAGATTCCCCTGATACTATGTTGCTGGCCTCTGAA

bovine3'   TGGTCA----TCCA------TCAGGT-------TGAGTGGTTCGTCCACTCATCCTG---
 human3'   TC--CA----TCCAGGCCATTCCCCG-------TCTCTGGTTCCTCCCCTCCTCCTG---
 mouse3'   TGGCCAGGATTCCGAAGGGTTCAGTGGAAAGCATGGATGTTAACTGAACGCCTGTTGCTG

bovine3'   --GAC-CCTACACATGGC-ACTTCCTCTGTGGTTGAGATTCGGAATGTAGGCATTCTCAC
 human3'   --GACTCCTGCACACGCT-CCTTCCTCTGAGGCTGAAATTCAGAATATTAGTGACCTCAG
 mouse3'   TGGATCCACATACAAGACGCCTTCCTCCTCCTTTTTGAAGAGGAGTATTCAGGAAG-CAG
```

Figure 10. CARD15 3’UTR alignment for human, mouse and bovine. Bases conserved among all three species are shaded.

```
bovine3'   CTTTGATCTTTCCCTTGTACCCTGGCCCTGCTCCCACCCTCCCATCCGCCTC--AACCCT
 human3'   CTTTGATATTTCACTTA-------------------------CAGCACCCCC--AACCCT
 mouse3'   CTCTGGCCGTGTCTGGACTCACTCTCTAAGGT----------CATCCCACTCTGATACCT

bovine3'   CCCCCCATCCAGGGTGGGAGGGGCTACAACTCACC-CTGCTCTCCTT-CTGGTACTTAGG
 human3'   GGCAC---CCAGGGTGGGAAGGGCTACACCTTAGC-CTGCCCTCCTTTCCGGTGTTTAAG
 mouse3'   CCACTTGTCCTGCCTGAGGCCCAACAGGACTCAACTCCGACAACCAACATAATACCCCGC

bovine3'   ACAGTATTGAAAGGGGACAGGTGACATACATGTGTTCCTCAAGACATTCTAGAGTTTCAA
 human3'   ACATTTTTGGAAGGGGACACGTGACAGCCGTTTGTTCCCCAAGACATTCTAGGTTTGCAA
 mouse3'   TCTCCTTTCTTAAC--ACAAACAACAGTCCTTTGC-CCCCAGGATGTTCTAGATTTATAA

bovine3'   GAAAAA-TATGACTGC---CCAGC---------AACTGGACTTTTATTTCCAGTGAAATC
 human3'   GAAAAA-TATGACCACACTCCAGCTGGGATCACATGTGGACTTTTATTTCCAGTGAAATC
 mouse3'   GAAAAAATGTGACCACACTCCAGCTGGGATCACACATGGCCTTTGGTTTCCAGTAACATC

bovine3'   AATTACTCTTCAGTTA-AACCTTTGGGAACAACTC---TGTA------TCCAAATGCAAC
 human3'   AGTTACTCTTCAGTTA-AGCCTTTGGAAACAGCTCGACTTTAAAAAGCTCCAAATGCAGC
 mouse3'   AATTACTCTATAGTTTGAGCCTTTGGAATCAGTTC--CTC-------CCTCAACTGCAGC

bovine3'   TTTTAAAACT-AACCTAGGCCAGAATTTTGAACAGCCCCACCAGGTCTCTG----AAGCC
 human3'   TTTAAAAAATTAATCTGGGCCAGAATTTCAAACGGCCTCACTAGGCTTCTGGTTGATGCC
 mouse3'   TCTTTAGAATTAATCTGGGCCAGAATTTCAAACGACCCCACCAGGCCCGGAGTCGGTGCC

bovine3'   TGTGAACTGAACTCTGGCAGCAGACTTCCAAAATATATTCATAAGAGATGGTTTGGTTTT
 human3'   TGTGAACTGAACTCTGACAACAGACTTCTGAAATAGACCCACAAGAGGCAGTTCCATTTC
 mouse3'   TGTGAACTGAACTCTGACAACAGACTTCTGAAACATATTCCTGAAAGACAGCTCCATTTT

bovine3'   GTTTGTGCCAGGCCACTTTAGGATATA-AGTTATAGATCAAAAGTTTACAGGGCAAAATC
 human3'   ATTTGTGCCAGAATGCTTTAGGATGTACAGTTATGGATTGAAAGTTTACAGGAAAAAAAA
 mouse3'   ATTTATGACA--ATGCTTTAGGACATGGAGTTGGGGCTTGGAGGCTTGCAGGGGACAAAA

bovine3'   AAAGGCCCTTCCTTACAAAACAAATGTTTTTCTCTGAATTTTTCA--------------G
 human3'   TTAGGCCGTTCCTT-CAAAGCAAATGTCTTCCT--GGATTATTCAAAATGATGTATGTTG
 mouse3'   CCAAGCCTTTCTTT-CTGAGCAAG-GTCTTTCTGTGG-TTTTTCAAAAT----------G

bovine3'   AAGCTTCTGTAAACTGTCAGGTACTGTGCAAGTGTTATTATTTCAA-CACTGTTATTTGT
 human3'   AAGCCTTTGTAAATTGTCAGATGCTGTGCAAATGTTATTATTTTAAACATTATGATGTGT
 mouse3'   TCACCTTTGTAATTTTTAA---ATTGTACAAGTTTTACTACCTCCA-TTTTATTTAGCGT

bovine3'   GAAAAACTGGTTAATGTTTATAAACCACTTTGTTTTATTCTCCCTAGTTCATGATTTTAT
 human3'   GAAAA-CTGGTTAATATTTATAGGTCACTTTGTTTTACTGTCTTAAGTTTATACTCTTAT
 mouse3'   GAAAG-CGGGTTAATATTTATAAATGGCACTGTTTTATTTT-------------------

bovine3'   AAAAAAAAAAAAAATGACCATGAATGTTATGCTGTAAATAATCACAGAAGATAAAACTAT
 human3'   AGACAA------CATGGCCGTGAACTTTATGCTGTAAATAATCAGAGGGGAATAAACTGT
 mouse3'   ------------------------------------------------------------

bovine3'   TGAGTCACCAGAACTATCTTCATTGTGACCAAACACAATGAAGTATTTAAAAATACTCTG
 human3'   TGAGTCAAAAC-------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   AACATTATCACATATTAAAGCACAATATTCTCCTTGAAGGGAGGAGACATGATGTTTCAA
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------
```

Figure 10. Continued.

```
bovine3'   CCAGATAATTGATTGCTTAAGGCACAAGCAGTGTTTAGAAATAGCCTCGCAATCAAAACA
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   CATTTGGCTTCAGTTTAGAGAAGTCTAGCCCAGCGTGGAGTTGTAAGTTATAGAGGAACC
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   TCAGTGTCCCGGCAGAAACACAGATGAGAGAGACGCAAGCAGGCCCCTGGGCCTCCCTCC
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   ATTCTCTCCAAGTGTCTCCAGGGGAGAAGGATGGAGAAGACTGGGGAACAGTTCTCCTCT
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   GCAAGCAGCCTCGTGGGTAGGCCTTGGTGAAATAATTCTTAGCTGAATTTAATTAGCAAG
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   GACTCAGGTGGCTGCTCATCAAGGTAGAATCGGCTTCCTTGAATGGTTTCCTGTGTGTCT
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   GGTTGGTTAAATACTGTGGCATCTCCCTGGGCGCCTCCCCAGTAAGGGCATGTGTGTGGG
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   TTCTCTTCATTTGTTTGAATTTATTTTATTGAGGTATGGTTGATTTACAGCGTTGTGTTA
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   ATTTCTATACAGCTGAATGATTCTGTTATACATATGTGTACATACTCTTTTTCCATTATG
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   CTTTTTCACAGGATAAACTACTCATAGTTCCTTGTGCTATACATTAGGACTTTTTTATGT
 human3'   ------------------------------------------------------------
 mouse3'   ------------------------------------------------------------

bovine3'   ATCCATTCAATATATAATACATTTGCTAACCTCA
 human3'   ----------------------------------
 mouse3'   ----------------------------------
```

Figure 10. Continued.

47.5% homology between human and mouse and 65.7% homology between bovine and human. The length of the 3'UTR is also quite variable among the species; 1257 bp in human, 1425 bp in mouse and 1915 bp in bovine.

A search for putative 5'UTR regulatory motifs revealed two known motifs in the bovine gene. The first of these is a terminal oligopyrimidine tract (TOP) which includes the first five bases of the transcript and the second is an internal ribosome entry site or internal regulatory sequence (IRES) which includes bases 61-151 of the 5'UTR and the ATG translation initiation codon. The TOP motif is also present in mouse and includes the first 11 bases of the transcript, but the motif is not present in human. The IRES motif is not present in either human or mouse.

The search for putative regulatory motifs present in the 3'UTR resulted in the identification of three motifs in human: an alcohol dehydrogenase 3'UTR down regulation control element (ADH_DRE) (bases 187-194); a Brd-Box (bases 907-913) and a Gy-Box (bases 1114-1120). These motifs were not identified in mouse or bovine. However, the bovine and the mouse 3'UTR each contain a single 15-lipoxygenase differentiation control element (15 LOX-DICE) repeat. The 3'UTR sequence in all three species was searched for the presence of polyadenylation signals and 15, 7 and 3 putative sites were identified in bovine, human and mouse, respectively (Table 10).

CARD15 flanking intronic sequence is not conserved across species. The percent identity between species comparisons for ~200 bases of intronic sequence flanking each exon is presented in Table 11. Pairwise alignments between human and mouse revealed a range in percent identity from 39% (5' region of intron 1) to 75% (5' region of intron 7), between bovine and mouse from 34.7% (3' region of intron 7) to 65% (5' region of intron 7) and between human and bovine from 39% (5' region of intron 1) to 74% (5' region of intron 5). Percent identities calculated from the multiple alignment of bovine, human and mouse ranged from 21.8% (3' region of intron 7) to 52.5% (5' region of intron 5). In addition to examining the extent of overall homology in each intronic region flanking an exon, this intronic sequence was searched for conserved motifs. The hexamer TGCATG was found to be present in four of the 22

Table 10. Putative CARD15 polyadenylation signals by species

Bovine		*Human*		*Mouse*	
Putative Signal	*3'UTR position*	*Putative Signal*	*3'UTR position*	*Putative Signal*	*3'UTR position*
AAGAAA	664-669	AAGAAA	717-722	AAGAAA	994-999
AATCAA	709-714	AAAAAA	840-845	AATTAA	1114-1119
AATCAA	933-938	AATTAA	844-849	TATAAA	1405-1410
AACAAA	955-960	AATAGA	928-933		
TATAAA	1060-1065	AAAAAA	1009-1014		
TATAAA	1099-1104	AATAAT	1223-1228		
AATAAT	1138-1143	AATAAA[a]	1237-1242		
GATAAA	1151-1156				
AATGAA	1198-1203				
ATTAAA	1235-1240				
AATCAA	1332-1337				
AATAAT	1551-1556				
GATAAA	1832-1837				
AATATA	1890-1895				
AATACA	1897-1902				

[a] Most frequent and efficient form.

Table 11. Percent conservation of CARD15 intronic sequences flanking exons

Intron	*Bovine:Mouse:Human*	*Bovine:Human*	*Mouse:Human*	*Bovine:Mouse*
1-5'	24	39	39	43
1-3'	41.5	68.3	51.5	42.6
2-5'	41	61	60	57
2-3'	45.5	70.3	58.4	55.4
3-5'	37	62	55	45
3-3'	47.5	68.3	63.4	62.4
4-5'	46	63	63	63
4-3'	41.6	73.3	57.4	51.5
5-5'	39	74	40	44
5-3'	52.5	68.3	65.3	59.4
6-5'	43	67	55	58
6-3'	34.7	57.4	49.5	51.5
7-5'	56	73	75	65
7-3'	21.8	70.3	47.5	34.7
8-5'	49	80	59	55
8-3'	41.6	63.4	57.4	47.5
9-5'	35	69	45	49
9-3'	36.6	62.4	54.5	51.5
10-5'	40.6	70	48	54
11-3'	50	70.3	58.4	57
10-3'	no mouse data	68.3	no mouse data	no mouse data
11-5'	no mouse data	76	no mouse data	no mouse data

introns, including bovine intron 2 flanking exon 1, bovine and human intron 5 flanking exon 4, human and mouse intron 7 flanking exon 6 (1 bp mismatch in bovine) and in human and bovine intron 11 flanking exon 10 (no mouse data available). In addition to the multiple occurrences of the TGCATG hexamer, other short stretches of conservation within the human, mouse and bovine introns were identified. Considering only motifs with at least 6 bp of conservation and no more than two internal mismatches between species, conserved blocks ranging from 6 to 19 bp in length (Table 12) were identified.

***CARD15 is highly conserved in* Bos taurus, Bos indicus *and* Bison bison.** CARD15 sequence from one *Bos indicus* and one *Bison bison* animal by exon are presented in Figs. 11 and 12, respectively and have been deposited in GenBank as accession numbers AY518749-AY518770. Alignment of the nucleotide sequence revealed 26/5105 substitutions between *bison* and *indicus*, 21/5105 substitutions between *bison* and *taurus* and 17/5105 substitutions between *indicus* and *taurus*. All of these substitution sites were unique, occurring in only one of the species or subspecies. Alignment of the protein sequence revealed 1/1013 substitution between *indicus* and *taurus*, 1/1013 substitution between *indicus* and *bison* and 2/1013 substitutions between *taurus* and *bison* (Fig. 13). The amino acid substitution present in *bison* is located at residue 641 interstitial to the NOD and the LRR domains. This residue is conserved across human, mouse, *taurus* and *indicus*. The amino acid substitution present in *taurus* is located at residue 733 which is located within the first LRR. This residue is conserved between human, mouse, *indicus* and *bison*. In addition to surveying the coding portion of the gene for homology, ~200 bases of intronic sequence flanking each exon was examined and these regions were found to possess high levels of homology among the bovids, ranging from 95.2% to 100% in each pairwise comparison.

Discussion

An extensive comparison of the bovine, mouse and human CARD15 genes was performed. As expected, high levels of sequence conservation were found throughout

Table 12. CARD15 intronic sequence motifs identified as conserved between human, mouse and bovine

Intron	*Conserved motif*[a]	*Intron*[b]	*Conserved motif*
1-5'	none	8-3'	CTgAaATGGAG
1-3'	TCAgTTT		TCTtTGAaGTC
	AGAAGcC		TTTTgTT
	CTGACCT	9-5'	CTgGCCTCaTC
	CCTCCC		TGTGTG
2-5'	none		TGGATGA
2-3'	TTCAcTT		TTGtCTTT
	CCTTCtCACA		AGAgCCTGG
3-5'	none		TtGGtTTGtTGGTT
3-3'	AGcCAGGA		TtCCTcCA
	AGGTCCC	9-3'	AtTGAG
	ACCaTGG		TtTTCTC
	TCTTCTG	10-5'	ACAgTAAT
4-5'	none		AGGTGgC
4-3'	CCAGtGTTCTTTAGT	**10-3'**	CCTAAGgGAG
	GGGTgTCcA		CTACTTAAT
	TGgGGTgCTC		GTGAATGGA
	TGGGGG		GAGAGA
5	TCACTG		TACATTTCACT
	TGCTTT		TCATTGGGAATCTCAGACA
6-5'	CTGAGT		CAGGTGGGCTTCAG
	GTCTCA		AGTCTC
	ATGCTGTG		AAAACCAAG
	CTcTGGA		TCACCATT
	GCTgAgG		TATCTTC
	TCCTGcCcTTTG	**11-5'**	TGGGCAGGCCT
	TTTCcAGG		CCTCAGTTTT
6-3'	TCTTCC		GGGAGAGAGGAA
7-5'	AGAGGG		AGAATTT
	GCATGcAGG		GATCCCTT
	GGGgCTT		TTCTGCATG
	GATTTAGGaGCgGgTGA[c]		TTTAAG
	GGtTgGGG		TTTTTAAA
7-3'	ACTaAAAAgTCT		TCTCTG
	AGtTTGGcCAT	11-3'	CAGTGT
	CACtTTGCTGGGACC		AGTCATGGAGgCTTGtT
	TGAGgCC		CCCTGG
	AACACA		GGTAAAA
8-5'	GTTaTGAAGgTC		CAGGCA
	TGAACTTTaTtT		CACTCA
	TGgGCT		

[a] Base mismatches among species indicated in lower case.

[b] Bold face indicates that no mouse data were available.

[c] This sequence contains 3 mismatches but was included because it also contained at least 6 contiguous bases containing only 2 mismatches.

Exon 1
agacgctggagttcctcttcagttcccatgcctacgtccttttactttcgctttctgtgggcactgaacca
gcccacgtCTCTGGACTCCGTGACCATCACAGGAGGGAACCTGCTGAGAGATTTCCTCTGCTGCGTGGGGA
CCCTGCCAGGGCTTGGAGCTCTGTGAGATCGCTTCCCACGGACTCCCAGGACCCAGAGTCTGAGGGCTGAG
CCCAGGgtaagccattggcagatgtgacagcacacggtgtgtggtggtgggggagggcctgggcatttgg
ggaggggagcagtagatgggagtcagtgcatccataggcttgaaatgcaccacgagt

Exon 2
cctctccactaccctccgcctgctttctgagaagccctgccctgacctcattctcctcccagATTGTGAA**A**
TGTGCGCACAAGATGCTTTTCAGACACAGAGAAGCCAACTGGTGGAGTTGCTGGTCTCGGGGTCCCTGGAG
GGCTTTGAGAGTATTCTGGACCGGCTGCTTTCCCGGGAAGTCCTCTCCTGGGAGGACTATGAGGGGCTTAG
CCTCGTGGGGCAGCCCATCTCCCACTTGGCCAGGCGCCTCCTGGACACCATCTGGAATAAGGGTACTTGGG
GCTGTGAACAACTGACTGCAGCTGTGCGGGAGGCCCAGGCCGACAGCCAGCCCCCCGAGCTTCCCAGCTCC
TGGGACCCCCACTCACCCCACCCAGCCCGTGACCTGCAGAGTCACCGACCAGCCATTGTCAGGAGACTCTA
CGGCCACGTGGAGGGTGTGCTGGACCTGACACAGCAGCGGGGTTTCATCAGCCAGTACGAAACTGATGAAA
TCAGGCGGCCCATCTTCACTTCATCCCAGCGGgtaagcacttccctct

Exon 3
ataaataatgacttttgctttccgcacag**GCAAGAAGGCTCCTTGATCTCGCCACAGTGAAGGCGAATGGG**
TTGGCTGCCTTTCTTCTACAGTGTATTCAGGAATTACCGGTCCCATTGGCCCTGCCTTTTGAAGgtatata
tgtgttcttttcagttatcagaaaaggaaaggaagactttcagattc

Exon 4
accccttatcaggtcccattttcaccatggtcccagctcctcggtttcgtcttctgtcttccag**ATGCCGC**
CTGTAAGAAGTACATGTCCAAGCTGAGGACCGTTATATCAGCTCAGTCTCGTTTCCTGAGCACCTACGATG
GAGCAGAGAATCTTTGCCTGGAAGAAGTATATACAGAGAATGTTCTGGAAATCCAGATGGAGGTGGGCATG
GCTGGACCTTCGCAGCAGAGCCCTACCACCCTAGGCCTGGAGGAGCTCTTCAGCACCCGTGACCATTTCAA
CAAAGAGGCAGACACTGTGCTGGTGGTGGGCGAGGCGGGCAGCGGCAAGAGCACGCTCTTGCAGCAGCTGC
ACCTGCTCTGGGCTTCCGGGCGGGCCTTCCAGGAATTTCTCTTCGTCTTCCCATTTAGCTGCCGGCAGCTG
CAGTGCCTGGTGAAACCGCTGTCCATGCGGACGCTGCTCTTCGAACACTGCTGTTGGCCCGACCTTGGCCC
CCAGGACGTCTTCCAGGTCCTCCTTGACCACCCTGAGCGCATCCTCTTAACCTTTGATGGCTTTGATGAGT
TCAGGTTCAGGTTCACGGATCAGGAGCGTCACTGCTGTCCGACCGCCCCCACGTCAGTCCAGAGTCTGCTC
TTCAACCTTCTGCAGGGCAACCTGCTAAAGAATGCCCGCAAGGTGTTGACCAGCCGCCCCAGCGCGGTATC
GGCGAGCCTCCGAAAGCACGTGCGCACGGAACTCAGCCTCAAGGGCTTCTCGGAAGAGGGCATCGAACTGT
ACCTGAGGAAGCGGCATCGCGAGCCTGGGGTGGCCGACCGCCTCCTCTGCCTGCTCAGAGCCACCTCGGCC
CTGCACGGTCTGTGCCACCTGCCTGTCTTCTCCTGGATGGTGTCCAAGTGCCACGAGGAGCTGTTGCTGCA
GGGCCGGGGGTCCCCAAAGACCACCACGGATATGTACCTGCTGATCCTGCGGCACTTTCTGCTGCACGCCT
CCCCGCTACCCTTAGCCAACCATGGCCTGGGACCCAGCCTGATTCAGGGCAGGCTCCCCACACTCCTGCAT
CTCGGCCGCCTGGCTCTCTGGGGCCTGGGCACATGCTGCTACGTGTTCTCAGCCAAACAGCTGCAGGCGGC
ACATGTCGACAGTGAGGACCTTTCTCTTGGCTTCCTGGTGCTTGCCAAGAGGGTTGTACCTGGGAGTACAG
CCCCCCTGGAATTTCTGCATATCACTTTTCAGTGCTTCTTTGCTGCATTCTACCTCGCCCTCAGTGCCGAC
ACCCCGCCATCCTCGCTCAGACATCTCTTCCAAGATCACAGGCCTGAAAGCTCGCCACTGGCCAGGGTGCT
GCCCAAATTGTTCCTGCGGGGCTCCCGATGCAGAGAGGGCAGCGTGGCTGCTTTGCTGCAGGGGGCCGAGC
CGCACAACCTCCAGATCACAGGGGCCTTCCTGGCGGGGCTGTTGTCACAGGAGCACCGGAGCTTGCTGGCG
GAGTGCCAGGCCTCTGAGACGGCCCTGCTCCGGCGCTGGGATTGTGTCCAGCGGTGTCTGACCCGCAGCCT
CCACGAGCATTTCCGCTCCATCCCACCCGCCTTGCCGGGTGAGGCCAARAGCATGCACGCCCTGCCTGGCT
TCCTCTGGCTTATCCGGAGCCTGTATGAGATGCAGGAGGAGCGACTGGCGCGGGAGGCCGTTCGCAGGCTG
AACGTTGGGCACCTCAAGCTGACCTTCTGCGGTGTGGGCCCGGCCGAGTGTGCTGCCCTGGCCTTCGTGCT
GCGCCACCTCCGGCGGCCTGTGGCCCTGCAGCTGGACCACAACTCTGTGGGCGACATCGGCGTGGAGCAGC

Figure 11. CARD15 exonic sequence for *Bos indicus*. Lower case is flanking sequence; plain face is transcribed but not translated; bold face is translated. Start and stop codons are underlined.

TGCTGCCTTGCCTGGGTGTCTGCAAAGCTCTTTAgtgagtgatgctgggtgatgctggctgggcactgtgg
gaatgctgccatcctggtgcaggtaggggagccccctttatgctgggagccaggagtccaggcccagcccc

Exons 5 and 6
gtctttgtttag**CTTGCGAGATAACAACATCTCAGACCGAGGCATCTGCAAGTTGGTTGAACATGCTCTTC**
GCTGTGAGCAGCTGCAGAAGTTAGCgtaagtcaggctgtggacattgggccccatgtccccagactcagcc
caagccatgccgcctgggcagcaattgtgagcgagtgatgggctgggatgtttgcatggccagagggcaca
ggggactcctgagggagccagtgacccagggttggggtggtcactgtccactatgctttatctccatgtct
cttctctctcggaactttccag**TCTTTTCAACAACAAGTTGACCGATGGCTGTGCACACTCCATGGCCAGG**
CTCCTTGCGTGCAAGCAGAACTTCTTGGCTTTGAGgtgagcctggggcttccctactcctggagactttcg
tccccacaactgagtcagtttggtctggtctcgccccataatgcgcatatgac

Exon 7
ttcccttccag**GCTAGGAAACAACCACATCACGGCTGCGGAGCCGAGGTGCTTGCCCAGGGGCTCAGAAC**
TAACAACTCCTTGCAGTTTTTGGGgtaggtgggattctggggcagaggggcagcatgcaggggttgggggc
ttgcgaggatttaggagcgggtgaaaccgg

Exon 8
tgtttactctgttgaaactttcag**GTTCTGGGGCAACCAGGTGGGTGACGAGGGGGCCCAGGCCTTGGCTG**
CAGCCTTGGGTGATCACCAGAGCTTGAGGTGGCTCAGgtaagcctcagagttcgtcccgc

Exon 9
ttttgtttttgacgacatccag**CCTGGTGGGGAACAACATTGGCAGCGTGGGTGCTCAAGCCTTAGCATT**
GATGTTGGAAAAGAATATGGCCCTGGAAGAACTCTGgtgagtttaggggattcatctctctagggaggcag
acactagcctttttttcattctctggcctcatctgtggaagctgatgtgtgtgtaagaccaagatggacgac
tgcg

Exon 10
tgggttttctcctttattcttgccag**CCTGGAGGAGAACCACGTCCAGGATGAAGGTGTGTGTTTCCTCGC**
CAAAGGACTTGCAAGAAACTCAAGTCTGAAAGTCCTGAAgtaaggaatctgtaagcaagagctagac

Exon 11
tcaccattctatcttcctctctag**GCTGTCTAACAACCACATCAGCTCCCTAGGGGCAGAGGCCCTCCTGC**
GGGCCCTTGAAAAGAATGACACCATTCTGGAAGTCTGgtaagatcctgggcaggcctctttaacctctctg
agcctca

Exon 12
tgatctaaagcacagctctggtccagtcctcactcaaatctctctttgctttatccctttcag**GCTCCGAG**
GAAACACTTTCTCTCCGGAGGAAATTGAGAAACTCAGCCACCAGGATACCAGACTCTTGCTCTGATGTTTC
CAGGCCAGTGTTCAGCTCAGTGTGTTTGGGAGGAGGCCATTGGTTTGGATCCCAGGATGGGACGACATCTG
AGCACAGCCCACTCAGATGGAACCTGGATCTGCCCAGGGCCAACCCAATAGGTCACCTTTGTTCTGGCACA
GGAAAGCACATCAGTGCCTTGTGGAGTAGACTTCACTGAATCCCAACTTTGCCATCAACTTCTTGCCAAGA
TTCAATCCTGGGATGTTGAAGAGGGGCAGCCTGCCTGTACAGGATGGGGCTGGTCTCAAGTCAAGCTGACA
TGCGTCAGGGAGGCCCATGGATGCCACTGAGTATTTATGGGTGTGGAGAGCTCCCCACGAGGAGGGATGCT
CGGGAAGTAACTGTTTGCTTTGTCTTAGCTCATGGTCATCCATCAGGTTGAGTGGTTCGTCCACTCATCCT
GGACCCTACACATGGCACTTCCTCTGTGGTTGAGATTCGGAATGTAGGCATTCTCACCTTTGATCTTTCCC
TTGTACCCTGGCCCTGCTCCCACCCTCCCATCCGCCTCAACCCTCCCCCCATCCAGGGTGGGAGGGGCTAC
AACTCACCCTGCTCTCCTTCTGGTACTTAGGACAGTATTGAAAGGGGACAGGTGACATACATGTGTTCCTC
AAGACATTCTAGAGTTTCAAGAAAAATATGACTGCCCAGCAACTGGACTTTTATTTCCAGTGAAATCAATT
ACTCTTCAGTTAAACCTTTGGGAACAACTCCGTATCCAAATGCAACTTTTAAAACTAACCTAGGCCAGAAT
TTTGAACAGCCCCACCAGGTCTCTGAAGCCTGTGAACTGAACTCTGGCAGCAGACTTCCAAAATATATTCA

Figure 11. Continued.

```
TAAGAGATGGTTTGGTTTTGTTTGTGCCAGGCCACTTTAGGATATAAAGTTATAGATCAAAAGTTTACAGG
GCAAAATCAAAGGCCCTTCCTTACAAAACAAATGTTTTTCTCTGAATTTTTCAGAAGCTTCTGTAAACTGT
CAGGTACTGTGCAAGTGTTATTATTTCAACACTGTTATTTGTGAAAAACTGGTTAATGTTTATAAACCACT
TTGTTTTATTCTCCCTAGTTCATGATTTTATAAAAAAAAAAAAAATGACCATGAATGTTATGCTGTAAATA
ATCACAGAAGATAAAACTATTGAGTCACCAGAACTATCTTCATTGTGACCAAACACAATGAAGTATTTAAA
AATACTCTGAACATTATCACATATTAAAGCACAATATTCTCCTTGAAGGGAGGGGACATGATGTTTCAACC
AGATAATTGATTGCTTAAGGCACAAGCAGTGTTTAGAAATAGCCTCGCAATCAAAACACATTTGGCTTCAG
TTTAGAGAAGTCTAGCCCAGCGTGGAGTTGTAAGTTATAGAGGAACCTCAGTGTCCCAGCAGAAACACAGA
TGAGAGAGACGCAAGCAGGCCCCTGGGCCTCCCTCCATTCTCTCCAAGTGTCTCCAGGGGAGAAGGATGGA
GAAGACTGGGGAACAGTTCTCCTCTGCAAGCAGCCTCGTGGGTAGGCCTTGGTGAAATAATTCTTAGCTAA
ATTTAATTAGCAAGGACTCAGGTGGCTGCTCATCAAGGTGGATTCGGCTTCCTTGAATGGTTTCCTGTGTG
TCTGGTTGGTTAAATACTGTGGCATCTCCCTGGGCGCCTCCCCAGTAAGGGCATGTGTGTGGGTTCTCTTC
ATTTGTTTGAATTTATTTTATTGAGGTATGGTTGATTTACAGCGTTGTGTTAATTTCTATACAGCTGAATG
ATTCTGTTATACATATGTGTACATACTCTTTTTCCATTATGCTTTTTCACAGGATAAACTACTCATAGTTC
CTTGTGCTATACATTAGGACTTTTTTATGTATCCATTCAATATATAATACATTTGCTAACCTCataataca
tttgctagcctcaatctcttactccat
```

Figure 11. Continued.

Exon 1
tcagttcccatgcctacgtccttttactttcgctttctgtgggcactgcaccagcccacgtCTCTGGACTC
CGTGACCATCACAGGAGGGAACCTGCTGAGAGATTTCCTCTGCTGCGTGGGGACCCTGCCAGGGCTTGGAG
CTCTGTGAGATCGCTTCCCACGGACTCCCAGGACCCAGAGTCTGAGGGCTGAGCCCAGGgtaagccattgg
cagaggtgacagcacacggtgtgtggtggtgggggagggcctgggcatttggggaggggagcagtagatg
ggagtcagtgcatccataggcttgaaatgcaccacgagt

Exon 2
cctctccactaccctccgcctgctttctgagaagccctgccctgacctcattctcctcccagATTGTGAA**A**
TGTGCGCACAAGATGCTTTTCAGACACAGAGAAGCCAACTGCTGGAGTTGCTGGTCTCGGGGTCCCTGGAG
GGCTTTGAGAGTATTCTGGACCGGCTGCTTTCCCGGGAAGTCCTCTCCTGGGAGGACTATGAGGGGCTTAG
CCTCGTGGGGCAGCCCATCTCCCACTTGGCCAGGCGCCTCCTGGACACCATCTGGAATAAGGGTGCTTGGG
GCTGTGAACAACTGACTGCAGCTGTGCGGGAGGCCCAGGCCGACAGCCAGCCCCCCGAGCTTCCCAGCTCC
TGGGACCCCCACTCACCCCACCCAGCCCGTGACCTGCAGAGTCACCGACCAGCCATTGTCAGGAGACTCTA
CGGCCACGTGGAGGGTGTGCTGGACCTGACACAGCAGCGGGGTTTCATCAGCCAGTACGAAACTGATGAAA
TCAGGCGGCCCATCTTCACTTCATCCCAGCGGgtaagcacttccctctaagcacttttcagaggaaaagtg
tgcttagttactg

Exon 3
ataaataatgacttttgctttccccacag**GCAAGAAGGCTCCTTGATCTCGCCACAGTGAAGGCGAATGGG**
TTGGCTGCCTTTCTTCTACAGTGTATTCAGGAATTACCGGTCCCATTGGCCCTGCCTTTTGAAGgtatata
tgtgttcttttcagttatcagaaaaggaaaggaagactttcagattc

Exon 4
acccctcatcaggtcccattttcaccatggtcccagctcctcagtttcttcttctgtcttccag**ATGCCGC**
CTGTAAGAAGTACGTGTCCAAGCTGAGGACCGTTATATCAGCTCAGTCTCGTTTCCTGAGCACCTACGATG
GAGCAGAGAATCTTTGCCTGGAAGAAGTATATACAGAGAATGTTCTGGAAATCCAGATGGAGGTGGGCATG
GCTGGACCTTCGCAGCAGAGCCCTACCACCCTAGGCCTGGAGGAGCTCTTCAGCACCCGTGACCATTTCAA
CAAAGAGGCAGACACTGTGCTGGTGGTGGGCGAGGCGGGCAGCGGCAAGAGCACGCTCTTGCAGCAGCTGC
ACCTGCTCTGGGCTTCCGGGCGGGCCTTCCAGGAATTTCTCTTCGTCTTCCCATTTAGCTGCCGGCAGCTG
CAGTGCCTGGTGAAACCGCTGTCCATGCGGACGCTGCTCTTCGAACACTGCTGTTGGCCCGACCTTGGCCC
CCAGGATGTCTTCCAGGTCCTCCTTGACCACCCTGAGCGCATCCTCTTAACCTTCGATGGCTTTGATGAGT
TCAGGTTCAGGTTCACGGATCAGGAGCGTCACTGCTGTCCGACCGCCCCCACGTCAGTCCAGAGTCTGCTC
TTCAACCTTCTGCAGGGCAACCTGCTAAAGAATGCCCGCAAGGTGTTGACCAGCCGCCCCAGCGCGGTATC
GGCGAGCCTCCGAAAGCACGTGCGCACGGAACTCAGCCTCAAGGGCTTCTCGGAAGAGGGCATCGAACTGT
ACCTGAGGAAGCGGCATCGCGAGCCTGGGGTGGCCGACCGCCTCCTCTGCCTGCTCAGAGCCACCTCGGCC
CTGCACGGTCTGTGCCACCTGCCTGTCTTCTCCTGGATGGTGTCCAAGTGCCACGAGGAGCTGTTGCTGCA
GGGCCGGGGGTCCCCAGAGACCACCACGGATATGTACCTGCTGATCCTGCGGCACTTTCTGCTGCACGCCT
CCCCGCTACCCTTAGCTACCCATGGCCTGGGACCCAGCCTGATTCAGGGCAGGCTCCCCACACTCCTGCAT
CTCGGCCGCCTGGCTCTCTGGGGCCTGGGCACATGCTGCTACGTGTTCTCAGCCAAACAGCTGCAGGCGGC
ACATGTCGACAGTGAGGACCTTTCTCTTGGCTTCCTGGTGCTTGCCAAGAGGGTTGTACCTGGGAGTACAG
CCCCCTTGGAATTTCTGCATATCACTTTTCAGTGCTTCTTTGCTGCATTCTACCTCGCCCTCAGTGCCGAC
ACCCCGCCATCCTCGCTCAGACATCTCTTCCAAGATCACAGGCCTGAAAGCTCGCCACTGGCCAGGGTGCT
GCCCAAATTGTTCCTGCGGGGCTCCCGATGCAGAGAGGGCAGCGTGGCTGCTTTGCTGCAGGGGGCCGAGC
AGCACAACCTCCAGATCACAGGGGCCTTCCTGGCGGGGCTGTTGTCACAGGAGCACCGGAGCTTGCTGGCG
GAGTGCCAGGCCTCTGAGACGGCCCTGCTCCGGCGCTGGGATTGTGTCCGGCGATGTCTGACCCGCAGCCT
CCGCGAGCATTTCCGCTCCATCCCACCCGCCTTGCCGGGTGAGGCCAAGAGCATGCACGCCCTGCCTGGCT
TCCTCTGGCTCATCCGGAGCCTGTATGAGATGCAGGAGGAGCGACTGGCGCGGGAGGCCGTTCGCAGGCTG
AACGTTGGGCACCTCAAGCTGACCTTCTGCGGTGTGGGCCCGGCCGAGTGTGCTGCCCTGGCCTTCGTGCT

Figure 12. CARD15 exonic sequence for *Bison bison*. Lower case is flanking sequence; plain face is transcribed but not translated; bold face is translated. Start and stop codons are underlined.

GCGCCACCTCCGGCGGCCTGTGGCCCTGCAGCTGGACCACAACTCTGTGGGCGACATCGGCGTGGAGCAGC
TGCTGCCTTGCCTGGGCGTCTGCAAAGCTCTTTAgtgagtgatgctgggtgatgctggctgggcactgtgg
gaatgctgccatcctggtgcaggtaggggagcccccttatgctgggagccaggagtccaggcccagcccc

Exons 5 and 6
gtctttgtttag**CTTGCGAGATAACAACATCTCAGACCGAGGCATCTGCAAGTTGGTTGAACATGCTCTTC**
GCTGTGAGCAGCTGCAGAAGTTAGCgtaagtcaggctgtggacattgggccccatgtccccagactcagcc
caagccatgccgcctgggcagcaattgtgagcgagtgatgggctgggatgtttgcatggctagagggcaca
ggggactcctgagggagccagtgacccagggttggggtggtcactgtccactatgctttatctccatgtct
cttctctctcggaactttccag**TCTTTTCAACAACAAGTTGACCGATGGCTGTGCACACTCCATGGCCAGG**
CTCCTTGCGTGCAAGCAGAACTTCTTGGCTTTGAGgtgagcctggggcttccctactcctggagactttcg
tccccacaactgagtcagtttggtctggtctcgccccataatgcgcatatgac

Exon 7
ttcccttccag**GCTAGGAAACAACCACATCACGGCTGCGGGAGCCGAGGTGCTTGCCCAGGGGCTCAGAAC**
TAACAACTCCTTGCAGTTTTTGGGgtaggtgggattctggggcagaggggcagcatgcaggggttgggggc
ttgcaaggatttaggagcaggtgaaaccgg

Exon 8
tgaaactttcag**GTTCTGGGGCAACCAGGTGGGTGACGAGGGGGCCCAGGCCTTGGCTGCAGCCTTGGGTG**
ATCACCAGAGCTTGAGGTGGCTCAGgtaagcctcagagttcgtcccgc

Exon 9
ttttgtttttgacgacatccag**CCTGGTGGGGAACAACATTGGCAGCGTGGGTGCTCAAGCCTTAGCATT**
GATGTTGGAAAAGAATATGGCCCTGGAAGAACTCTGgtgagtttaggggattcatctctctagggaggcag
acactagcctttttttcattctctggcctcatctgtggaagctgatgtgtgtgtaagaccaagatggatgac
tgcg

Exon 10
tgggttttctcctttattcttgccag**CCTGGAGGAGAACCACGTCCAGGATGAAGGTGTGTGTTTCCTCGC**
CAAAGGACTTGCAAGAAACTCAAGTCTGAAAGTCCTGAAgtaaggaatctgtaagcaagagctagac

Exon 11
tcaccattctatcttcctctctag**GCTGTCTAACAACCACATCAGCTCCCTAGGGGCAGAGGCCCTCCTGC**
GGGCCCTTGAAAAGAATGACACCATTCTGGAAGTCTGgtaagatcctgggcaggcctctttaacctctctg
agcctca

Exon 12
tgatctaaagcacagctctggtccagtcctcactcaaatctgtctttgctttatccccttcag**GCTCCGAG**
GAAACACTTTCTCTCCGGAGGAAATTGAGAAACTCAGCCACCAGGATACCAGACTCTTGCTCTGATGTTTC
CAGGCCAGTGTTCAGCTCAGTGTGTTTGGGAGGAGGCCATTGGTTTGGATCCCAGGATGGGACGACATCTG
AGCACAGCCCACTCAGATGGAACCTGGATCTGCCCAGGGCCAACCCAATAGGTCACCTTTGTTCTGGCACA
GGAAAGCACATCAGTGCCCTGTGGAGTAGACTTCACTGAATCCCAACTTTGCCATCAACTTCTTGCCAAGA
TTCAATCCTGGGATGTTGAAGAGGGGCAGCCTGCCTGTACAGGATGGGGCTGGTCTCAAGTCAAGCTGACA
TGCGTCAGGGAGGCCCATGGATGCCACTGAGTATTTATGGGTGTGGAGAGCTCCCCACGAGGAGGGATGCT
CGGGAAGTAACTGTGTGCTTTGTCTTAGCTCATGGTCATCCATCAGGTTGAGTGGTTCGTCCACTCATCCT
GGACCCTACACATGGCACTTCCTCTGTGGTTGAGATTCGGAATGTAGGCATTCTCACCTTTGATCTTTCCC
TTGTACCCTGGCCCTGCTCCCACCCTCCCATCCGCCTCAACCCTCCCCCCATCCAGGGTGGGAGGGGCTAC
AACTCACCCTGCTCTCCTTCTGGTACTTAGGACAGTATTGAAAGGGGACAGGTGACATACATGTGTTCCTC
AAGACATTCTAGAGTTTCAAGAAAAATATGACTGCCCAGCAACTGGACTTTTATTTCCAGTGAAATCAATT
ACTCTTCAGTTAAACCTTTGGGAACAACTCCGTATCCAAATGCAACTTTTAAAACTAACCTAGGCCAGAAT

Figure 12. Continued.

```
TTTGAACAGCCCCACCAGGTCTCTGAAGCCTGTGAACTGAACTCTGGCAGCAGACTTCCAAAATATATTCA
TAAGAGATGGTTTGGTTTTGTTTGTGCCAGGCCACTTTAGGATATAAAGTTATAGATCAAAAGTTTACAGG
GCAAAATCAAAGGCCCTTCCTTACAAAACAAATGTTTTTCTCTGAATTTTTCAGAAGCTTCTGTAAACTGT
CAGGTACTGTGCAAGTGTTATTATTTCAACACTGTTATTTGTGAAAAACTGGTTAATGTTTATAAACCACT
TTGTTTTATTCTCCCTAGTTCATGATTTTATAAAAAAAAAAAATGACCATGAATGTTATGCTGTATATAAT
CACAGAAGATAAAACTATTGAGTCACCAGAACTATCTTCATTGTGACCAAACACAATGAAGTATTTAAAAA
TACTCTGAACATTATCACATATTAAAGCACAATATTCTCCTTGAAGGGAGGAGACATGATGTTTCAACCAG
ATAATTGATTGCTTAAGGCACAAGCAGTGTTTAGAAATAGCCTCGCAATCAAAACACATTTGGCTTCAGTT
TAGAGAAGTCTAGCCCAGCGTGGAGTTGTAAGTTATAGAGGAACCTCAGTGTCCCGGCAGGAACACAGATG
AGAGAGACGCAAGCAGGCCCCTGGGCCTCCCTCCATTCTCTCCAAGTGTCTCCAGGGGAGAAGGATGGAGA
AGACTGGGGAACAGTTCTCCTCTGCAAGCAGCCTCGTGGGTAGGCCTTGGTGAAATAATTCTTAGCTGAAT
TTAATTAGCAAGGACTCAGGTGGCTGCTCATCAAGGTGGAATCGGCTTCCTTGAATGGTTTCCTGTGTGTC
TGGTTGGTTAAATACTGTGGCATCTCCCTGGGCGCCTCCCCAGTAAGGGCATGTGTGTGGGTTCTCTTCAT
TTGTTTGAATTTATTTTATTGAGGTATGGTTGATTTACAGCGTTGTGTTAATTTCTATATAGCTGAATGAT
TCTGTTATACATATGTGTACATACTCTTTTTCCATTATGCTTTTTCACAGGATAAACTACTCATAGTTCCT
TGTGCTATACATTAGGACTTTTTTATGTATCCATTCAATATATAATACATTTGCTAACCTCataatacatt
tgctagcctcaatctcttactccatcc
```

Figure 12. Continued.

```
 Indicus     (1) MCAQDAFQTQRSQLVELLVSGSLEGFESILDRLLSREVLSWEDYEGLSLVGQPISHLARR
  Taurus     (1) MCAQDAFQTQRSQLVELLVSGSLEGFESILDRLLSREVLSWEDYEGLSLVGQPISHLARR
   Bison     (1) MCAQDAFQTQRSQLLELLVSGSLEGFESILDRLLSREVLSWEDYEGLSLVGQPISHLARR

 Indicus    (61) LLDTIWNKGTWGCEQLTAAVREAQADSQPPELPSSWDPHSPHPARDLQSHRPAIVRRLYG
  Taurus    (61) LLDTIWNKGTWGCEQLTAAVREAQADSQPPELPSSWDPHSPHPARDLQSHRPAIVRRLYG
   Bison    (61) LLDTIWNKGAWGCEQLTAAVREAQADSQPPELPSSWDPHSPHPARDLQSHRPAIVRRLYG

 Indicus   (121) HVEGVLDLTQQRGFISQYETDEIRRPIFTSSQRARRLLDLATVKANGLAAFLLQCIQELP
  Taurus   (121) HVEGVLDLTQQRGFISQYETDEIRRPIFTSSQRARRLLDLATVKANGLAAFLLQCIQELP
   Bison   (121) HVEGVLDLTQQRGFISQYETDEIRRPIFTSSQRARRLLDLATVKANGLAAFLLQCIQELP

 Indicus   (181) VPLALPFEDAACKKYMSKLRTVISAQSRFLSTYDGAENLCLEEVYTENVLEIQMEVGMAG
  Taurus   (181) VPLALPFEDAACKKYVSKLRTVISAQSRFLSTYDGAENLCLEEVYTENVLEIQMEVGMAG
   Bison   (181) VPLALPFEDAACKKYVSKLRTVISAQSRFLSTYDGAENLCLEEVYTENVLEIQMEVGMAG

 Indicus   (241) PSQQSPTTLGLEELFSTRDHFNKEADTVLVVGEAGSGKSTLLQQLHLLWASGRAFQEFLF
  Taurus   (241) PSQQSPTTLGLEELFSTRDHFNKEADTVLVVGEAGSGKSTLLQQLHLLWASGRAFQEFLF
   Bison   (241) PSQQSPTTLGLEELFSTRDHFNKEADTVLVVGEAGSGKSTLLQQLHLLWASGRAFQEFLF

 Indicus   (301) VFPFSCRQLQCLVKPLSMRTLLFEHCCWPDLGPQDVFQVLLDHPERILLTFDGFDEFRFR
  Taurus   (301) VFPFSCRQLQCLVKPLSMRTLLFEHCCWPDLGPQDVFQVLLDHPERILLTFDGFDEFRFR
   Bison   (301) VFPFSCRQLQCLVKPLSMRTLLFEHCCWPDLGPQDVFQVLLDHPERILLTFDGFDEFRFR

 Indicus   (361) FTDQERHCCPTAPTSVQSLLFNLLQGNLLKNARKVLTSRPSAVSASLRKHVRTELSLKGF
  Taurus   (361) FTDQERHCCPTAPTSVQSLLFNLLQGNLLKNARKVLTSRPSAVSASLRKHVRTELSLKGF
   Bison   (361) FTDQERHCCPTAPTSVQSLLFNLLQGNLLKNARKVLTSRPSAVSASLRKHVRTELSLKGF

 Indicus   (421) SEEGIELYLRKRHREPGVADRLLCLLRATSALHGLCHLPVFSWMVSKCHEELLLQGRGSP
  Taurus   (421) SEEGIELYLRKRHREPGVADRLLCLLRATSALHGLCHLPVFSWMVSKCHEELLLQGRGSP
   Bison   (421) SEEGIELYLRKRHREPGVADRLLCLLRATSALHGLCHLPVFSWMVSKCHEELLLQGRGSP

 Indicus   (481) KTTTDMYLLILRHFLLHASPLPLANHGLGPSLIQGRLPTLLHLGRLALWGLGTCCYVFSA
  Taurus   (481) KTTTDMYLLILRHFLLHASPLPLATHGLGPSLIQGRLPTLLHLGRLALWGLGTCCYVFSA
   Bison   (481) ETTTDMYLLILRHFLLHASPLPLATHGLGPSLIQGRLPTLLHLGRLALWGLGTCCYVFSA

 Indicus   (541) KQLQAAHVDSEDLSLGFLVLAKRVVPGSTAPLEFLHITFQCFFAAFYLALSADTPPSSLR
  Taurus   (541) KQLQAAHVDSEDLSLGFLVLAKRVVPGSTAPLEFLHITFQCFFAAFYLALSADTPPSSLR
   Bison   (541) KQLQAAHVDSEDLSLGFLVLAKRVVPGSTAPLEFLHITFQCFFAAFYLALSADTPPSSLR

 Indicus   (601) HLFQDHRPESSPLARVLPKLFLRGSRCREGSVAALLQGAEPHNLQITGAFLAGLLSQEHR
  Taurus   (601) HLFQDHRPESSPLARVLPKLFLRGSRCREGSVAALLQGAEPHNLQITGAFLAGLLSQEHR
   Bison   (601) HLFQDHRPESSPLARVLPKLFLRGSRCREGSVAALLQGAEQHNLQITGAFLAGLLSQEHR

 Indicus   (661) SLLAECQASETALLRRWDCVQRCLTRSLHEHFRSIPPALPGEAKSMHALPGFLWLIRSLY
  Taurus   (661) SLLAECQASETALLRRWDCVRRCLTRSLREHFRSIPPALPGEAKSMHALPGFLWLIRSLY
   Bison   (661) SLLAECQASETALLRRWDCVRRCLTRSLREHFRSIPPALPGEAKSMHALPGFLWLIRSLY

 Indicus   (721) EMQEERLAREAVRRLNVGHLKLTFCGVGPAECAALAFVLRHLRRPVALQLDHNSVGDIGV
  Taurus   (721) EMQEERLAREAVCRLNVGHLKLTFCGVGPAECAALAFVLRHLRRPVALQLDHNSVGDIGV
   Bison   (721) EMQEERLAREAVRRLNVGHLKLTFCGVGPAECAALAFVLRHLRRPVALQLDHNSVGDIGV

 Indicus   (781) EQLLPCLGVCKALYLRDNNISDRGICKLVEHALRCEQLQKLALFNNKLTDGCAHSMARLL
  Taurus   (781) EQLLPCLGVCKALYLRDNNISDRGICKLVEHALRCEQLQKLALFNNKLTDGCAHSMARLL
   Bison   (781) EQLLPCLGVCKALYLRDNNISDRGICKLVEHALRCEQLQKLALFNNKLTDGCAHSMARLL
```

Figure 13. Aligned CARD15 amino acid sequence for *Bos indicus*, *Bos taurus* and *Bison bison*. Amino acid substitutions are shaded.

```
Indicus   (841) ACKQNFLALRLGNNHITAAGAEVLAQGLRTNNSLQFLGFWGNQVGDEGAQALAAALGDHQ
 Taurus   (841) ACKQNFLALRLGNNHITAAGAEVLAQGLRTNNSLQFLGFWGNQVGDEGAQALAAALGDHQ
  Bison   (841) ACKQNFLALRLGNNHITAAGAEVLAQGLRTNNSLQFLGFWGNQVGDEGAQALAAALGDHQ

Indicus   (901) SLRWLSLVGNNIGSVGAQALALMLEKNMALEELCLEENHVQDEGVCFLAKGLARNSSLKV
 Taurus   (901) SLRWLSLVGNNIGSVGAQALALMLEKNMALEELCLEENHVQDEGVCFLAKGLARNSSLKV
  Bison   (901) SLRWLSLVGNNIGSVGAQALALMLEKNMALEELCLEENHVQDEGVCFLAKGLARNSSLKV

Indicus   (961) LKLSNNHISSLGAEALLRALEKNDTILEVWLRGNTFSPEEIEKLSHQDTRLLL
 Taurus   (961) LKLSNNHISSLGAEALLRALEKNDTILEVWLRGNTFSPEEIEKLSHQDTRLLL
  Bison   (961) LKLSNNHISSLGAEALLRALEKNDTILEVWLRGNTFSPEEIEKLSHQDTRLLL
```

Figure 13. Continued.

the protein coding regions of the gene. The overall amino acid homology was 81.2 % between bovine and human, 76% between bovine and mouse and 79.4% between human and mouse. However, the pairwise comparisons revealed that there was a slightly higher homology within the protein domains than in the inter-domain regions of CARD15. The 5'UTR of CARD15 is poorly conserved across the species with both the mouse and human transcripts containing two in-frame translation initiation sites within the first and second exons respectively, whereas, the bovine transcript contains only the translation initiation site located in the second exon. Studies in both human and mouse reveal that the 1013 amino acid protein produced from the second translation initiation site is most commonly detected. Similarly, there is little sequence conservation in the CARD15 3'UTR between the species, with the bovine 3'UTR being considerably longer than those for the human and mouse genes. The total genomic size of the gene is also variable among the three species: ~36 kb in human, ~39 kb in mouse and ~30 kb in bovine.

The 5'UTR and 3'UTR of genes often contain key regulatory elements involved in the post-transcriptional regulation of gene expression. While function has been ascribed to relatively few of the elements residing within 3'UTRs, these motifs generally enable interactions with RNA-binding proteins, and/or facilitate the formation of secondary structures in the 3'UTR of the mRNA. Effects mediated by these elements include the regulation of transcript stability, specification of subcellular transcript localization and the regulation of translation. Polyadenylation may also be a useful mechanism in the regulation of gene expression, since the efficiency of 3' mRNA end-processing may be impacted by the location and motif of the polyadenylation signal site within the 3'UTR. Because most pre-mRNAs in a cell are not efficiently processed, even small changes in the overall processing efficiency of a particular pre-mRNA may have a substantial effect on the overall level of gene expression. Greener et al. (2002) have shown that loss of the dystrophin gene 3'UTR results in a pronounced reduction in the level of dystrophin protein and is sufficient to cause Becker muscular dystrophy, indicating the importance and regulatory role of the 3'UTR. Polyadenylation signal motifs also determine the efficiency of addition of the polyA tail. In bovine, three putative alternative polyadenylation signal motifs within the last 82 bp of the transcript (Table 10) were identified. Of these, two were found within the last 25 bp of the sequenced transcript and from experimental results in fruit fly (Graber et al. 1999), each should have approximately the same efficiency of polyadenylation (20 and 22%). The third motif, which is located further from the end of the transcript has only a 5% efficiency of polyadenylation. Thus, it is likely that one (or both) of the two signal motifs located within the last 25 bases is most commonly used for the regulation of polyadenylation of the bovine CARD15 transcript.

A total of 15 putative polyadenylation signal motifs were identified in the bovine 3'UTR. Further upstream is located a motif with a 68% polyadenylation efficiency in fruit fly and another two motifs each with a 28% efficiency. In all three cases, a possible cleavage site (CA) and a downstream motif of 5 bases which included at least 4 uracils were found. Clearly, the most likely signal used to produce the transcript sequenced was one of the two motifs within the last 25 bases of the transcript, however, it seems likely that these additional three sites may also be utilized as alternative bovine polyadenylation signals. Several recent reports provide evidence that alternative polyadenylation signal sites with different polyadenylation efficiencies are used in the tissue specific regulation of gene expression. In particular, the *D. melanogaster* Su(f) gene possesses at least three distinct polyadenylation signal sites that are involved in mRNA processing, one of which produces a severely truncated transcript that is not translated (Audibert and Simonelig 1998). However, production of the nonfunctional transcript is correlated with levels of the functional protein, implying a self-regulation mechanism for the gene.

Two putative regulatory motifs were identified in the bovine 5'UTR. The identified terminal oligopyrimidine tract is a motif consisting of from 5 to15 pyrimidines and has been identified in vertebrate ribosomal protein and translation elongation factors. This tract is required for the coordination of translational repression and its deletion has been shown to result in unregulated translation (Levy et al. 1991). This suggests that this region is likely to be involved in the regulation of translation of the bovine and mouse genes. The fact that this motif is not found in human suggests that translation of the human gene is controlled by some other mechanism. In addition to the terminal oligopyrimidine tract, the bovine gene also contains an internal ribosome entry site or internal regulatory sequence, the function of which is currently under debate. This motif is believed to be involved in internal mRNA ribosome binding, which is a mechanism of translation initiation that is an alternative to the conventional 5’-cap dependent ribosome scanning mechanism. This mechanism is thought to be advantageous for the translation of specific mRNAs during the cell cycle, when under-phosphorylated eIF-4F prevents the conventional mechanism of initiation of translation. Thus, the bovine gene may allow translation to proceed by both mechanisms allowing production of the CARD15 protein even during the cell cycle when it is presumably not produced in human or mouse.

Three putative regulatory motifs were detected in the 3’UTR of human, but only one was found in bovine and mouse. The ADH_DRE found in human has been shown to produce a 2× down-regulation of ADH gene expression. The Brd-box motif mediates negative post-transcriptional regulation by affecting transcript stability and translational efficiency and the GY-box motif is believed to function in concert with the Brd-box motif in the mediation of post-transcriptional regulation. This suggests that one or more of these elements are associated with down-regulation of CARD15 expression in human and may thus be important elements in susceptibility to Crohn’s disease. The bovine and mouse 3’UTR both contain one 15 LOX-DICE repeat which is known to specifically bind regulatory proteins which inhibit mRNA translation. However, functional 15 LOX-DICE elements require at least two of these elements and because only a single element was found in bovine and mouse, this motif probably has no functional relevance.

Recent work comparing intronic sequence between human and mouse suggests that motifs within conserved sequence within 100 bp of exon boundaries may regulate the alternative splicing of exons (Sorek and Ast 2003). These authors found that intronic sequences flanking exons that were constitutively expressed in transcripts from both species were not conserved between the species. Approximately 200 bases of intronic sequence flanking both the 5' and 3' ends of each exon of the CARD15 gene in bovine, human and mouse were examined. All of the examined sequences contained low levels of conservation between human and mouse with homology ranging from 39-75%. The level of nucleotide conservation within the flanking intronic regions between human, mouse and bovine ranged from 21.8-56%. The hexamer TGCATG has been shown to be involved in the regulation of alternative splicing of exons when found in the downstream intron (Lim and Sharp 1998; Deguillien et al. 2001). Sorek and Ast (2003) found that this motif occurred 9× more often than expected in the intronic sequences downstream of alternatively spliced exons. The TGCATG motif was found in 4 of the 22 intronic flanking sequences and in each case the hexamer was detected within 200 bases downstream of an exon (130 bp from bovine exon 1, 104 bp from human and 98 bp from bovine exon 4, 27 bp from human, 27 bp from bovine and 27 bp from mouse exon 6 (AGCATG in bovine) and 134 bp from human and 130 bp from bovine exon 10). Transcript isoforms of the human CARD15 gene have recently been deposited in GenBank (Accession nos. AY187233 to AY187246). These represent partial sequences derived from cDNAs, however, they indicate that human CARD15 exons 3, 5, 6 and 7 are alternatively spliced. Thus, there must be additional elements, potentially located within the exons (such as Exonic Splice Enhancers) responsible for the alternate splicing of human exons 3, 5 and 7. Whether the detected motif is responsible for the regulation of alternate splicing of any of the designated exons in these species is a matter that will require further investigation.

In addition to harboring sequence motifs that are involved in the alternative splicing of exons, introns have also been shown to harbor other forms of regulatory elements. Giacopelli et al. (2003) found that sequence located in the first untranslated exon and first intron of the human osteopontin gene enhanced promoter activity. Brend et al. (2003) identified an enhancer located within the intron of the mouse Hoxb4 gene to be sufficient for appropriate temporal activation of expression and the establishment of the correct anterior boundary in the paraxial mesoderm. Borchert et al. (2003) identified negative regulatory elements located within the first intron of the human ph/snGPx gene which are involved in the joint regulation of the snGPx and phGPx GPx isoforms. A regulatory mutation within intron 3 of IGF2 results in a threefold increase in expression of mRNA when paternally inherited and has quantitative effects on muscle growth, fat deposition and size of the heart in pig (Van Laere et al. 2003). Several short stretches of conserved sequence within human, mouse and bovine introns were identified. These conserved regions could be functional regulatory RNA or protein binding regions, or could simply be ancient remnants of conserved sequence from the last common ancestor of these species. The unequivocal identification of regulatory elements is challenging because these elements are typically short (6–15 bp), they tolerate some degree of sequence variation and rules useful for their recognition are generally unknown. Since the divergence of human, mouse and bovine occurred at least 100 mya (Burt et al. 1999), it appears likely that unconstrained, nonfunctional genomic sequences would have diverged in at least one of the species within this time period. Consequently, the intronic regions found in this study to be conserved between all three species probably have functional relevance.

The mechanism by which CARD15 mutations confer susceptibility to Crohn’s disease remains poorly understood. Since Crohn’s disease is thought to involve an abnormal immune response to enteric bacteria in a genetically susceptible individual, it is possible that a defect in this signaling pathway is involved in the etiology of the disease. A deficit in CARD15 activity may lead to an impaired ability of the host to recognize and respond normally to enteric bacteria. The mutations associated with susceptibility to Crohn’s disease have been shown to result in a deficit in CARD15 activity in response to the bacterial component muramyl dipeptide. Consequently, further investigation into the existence of genetic variability within the identified putative regulatory regions seems warranted since variation in these regions may be associated with variation in susceptibility to Crohn’s disease.

MUTATION AND ASSOCIATION ANALYSES FOR BOVINE CARD15

Introduction

In vertebrates, the first line of defense against microbial pathogens is mediated by the innate immune system. Toll-like receptors (TLR) are responsible for the initial recognition of pathogens and these receptors have been shown to mediate the response to pathogen-associated molecular patterns such as lipopolysaccharide (LPS) and peptidoglycan (PGN). Toll-like receptors are composed of multiple extracellular leucine rich repeats (LRR) and an intracellular Toll-IL1 receptor domain which mediates a signaling cascade to the nucleus. The plant counterparts to the TLR are the disease resistance (R) genes. These genes also contain LRR and appear to recognize pathogen components both at the cell surface and within the cytoplasm. CARD15 has a structure similar to these two classes of genes consisting of an effector domain region (2 caspase recruitment domains), a centrally located nucleotide binding/oligomerization domain and carboxy-terminal LRR.

It has been shown that mutant forms of CARD15 lacking the LRR have an enhanced ability to activate NF-kB (Miceli-Richard et al. 2001). Additionally, the LRR of CARD15 have been shown to generally recognize bacterial products (Ogura et al. 2001b) and to confer responsiveness to LPS and PGN (Inohara et al. 2001). More specifically, muramyl dipetide (MDP) from peptidoglycan has been shown to be the key component selected by the CARD15 LRR (Girardin et al. 2003, Inohara et al. 2003). Specific mutations in the CARD15 gene in human have been shown to be associated with Crohn’s disease (Hugot et al. 2001; Ogura et al. 2001) and Blau Syndrome (Miceli-Richard et al. 2001) two very distinct granulomatous disorders. The R702W, G908R and L1007fsinsC variants associated with Crohn’s disease are associated with a reduced level of NF-kB activation in response to MDP in comparison to the normal CARD15 gene (Girardin et al. 2003; Inohara et al. 2003).

Whether there is a connection between Crohn’s disease in human and Johne’s disease in cattle has been a subject of contention since the diseases were first discovered in the late 1800s. There are many commonalities between the diseases which support the hypothesis that they are related. For example, both diseases result in diarrhea and weight loss and both result in granuloma formation in the ileum. Additionally, *M. ptb.* is known to be the causative agent for Johne's disease and in many cases mycobacteria (*M. ptb*. and other spp.) have been cultured from human Crohn's patient tissues (Chiodini et al. 1984; Coloe et al. 1986; Graham et al. 1987; Gitnick et al. 1989; Thorel et al. 1990; Collins et al. 2000; Schwartz et al. 2000). The belief, by some, that the two diseases are related and the recently discovered associations between mutations in the CARD15 gene and Crohn's disease make CARD15 an interesting candidate gene for susceptibility to Johne's disease. In order to evaluate CARD15 as a candidate gene for susceptibility to Johne's disease, it is necessary to elucidate the naturally occurring variation in the gene within the cattle population. DNA sequence including flanking intronic sequence for the bovine CARD15 gene was previously reported. Here, the results of a mutation analysis of the CARD15 gene in several breeds and subspecies of cattle and also an association study between these mutations and their haplotypes with Johne's disease is reported.

Materials and Methods

Breed panel. A breed panel of thirty unrelated cattle was assembled from nine domestic cattle populations (White 2003). The panel comprised: eight *Bos indicus* individuals including 6 Brahman and 2 Gir; and twenty *Bos taurus* individuals including 3 Angus, 7 Holstein, 3 Texas Longhorn, 2 Limousin, 3 Jersey and 2 N'Dama. Two Ankole-Watusi animals which represent an ancient *taurus* × *indicus* cross were also included in the panel.

Panel diagnosed with Johne's disease. In addition to the above mentioned breed panel, a panel of 11 DNA samples from animals that had been visually examined and diagnosed by a veterinarian with Johne's disease was assembled. Eight tissue samples were purchased from Dr. Michael Collins (University of Wisconsin) and included: three unrelated Holstein calves that had been experimentally infected with *M. ptb.* at 1 mo of age, one Holstein female that had been an embryo transfer recipient and was determined to have the disease after her calf was found to be infected; one Holstein that was a clinical Johne's case and that was found to be ELISA positive on repeated assay; one Holstein that was culture positive; and two Jerseys, one of which had a low grade infection based on tissue *M. ptb.* counts. Three samples were kindly provided by Dr. Allen Roussel (Texas A&M University) and included one Holstein, Jersey and Brahman. All of the latter animals were both ELISA and culture positive.

DNA extractions. DNA was extracted from tissue samples received from Dr. Michael Collins (ileum, spleen, lymph node, lung) using the following protocol. Approximately 500 mg of tissue that had been flash frozen was ground using a mortar and pestle and resuspended in 5 ml of saline EDTA. To this was added 360 µl of saline EDTA, 180 µl Proteinase K and the sample was incubated at 55°C while being rotated overnight. The samples were then extracted twice with phenol-chloroform and precipitated with 450 µl 3M sodium acetate and 14 mL of 95% ethanol. Samples were incubated at room temperature for 2 hr and then washed with 10 mL of ice-cold 70% ethanol. The samples were centrifuged for 5 min at 2000 rpm and the alcohol removed. The samples were then placed in a SpeedVac and allowed to dry for approximately 1 hr. The dried pellets were resuspended in 500 µl of TE and placed in a 37°C water bath overnight. The samples were then allowed to rotate for 1 hr and were quantified using a spectrophotometer.

PCR and sequencing. To screen the breed panel for the presence of polymorphisms, each exon of CARD15 along with a small amount of flanking intronic sequence was amplified in every individual and sequenced in the forward and reverse directions. The primers developed for the amplification of the individual exons in *Bos taurus*, *Bos indicus* and *Bison bison* (Table 7) were amplified in a 50 µl volume: 15 mM Tris, 50 mM KCl, 1.5 mM $MgCl^{2+}$, 200 µM of each dNTP (dATP, dCTP, dGTP and dTTP), 1 µM each primer, 50 ng DNA, 0.5 units of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, Calif.). The general PCR reaction included an initial hot start of 10 min at 95°C followed by 35 cycles for 30 s at 94°C, 30 s at the optimized annealing temperature and 45 s at 72°C, followed by a final elongation step at 72°C for 7 min. The PCR products were analyzed by electrophoresis on 2% agarose gels in 1.0× TBE buffer (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8.3) and stained with ethidium bromide. The PCR product was purified using QIAquick® (Qiagen, Valencia, Calif.) spin columns.

Cycle sequence reactions were performed in a 10 μl reaction volume: 1 μl Big Dye v3.0 (Applied Biosystems), 10 ng amplified DNA, 0.5 μl of 5% dimethylsulfoxide (DMSO), 0.5 μl of 20 μM primer. The general cycle sequence reaction included an initial denaturation step of 2 min at 96°C followed by 35 cycles for 15 s at 96°C, 20 s at 50°C and 4 min at 60°C, followed by a final elongation step at 60°C for 15 min. Unincorporated nucleotides and primers were removed from the reactions using a BioMax® Spin-50 mini- sephadex column (Millipore Corp., Bedford, Mass.). Sequencing reactions were run on an ABI 3100 DNA analyzer (Applied Biosystems).

All identified single nucleotide polymorphisms (SNPs) were validated by performing a second PCR and resequencing the produced amplicon in both the forward and reverse directions. This approach was used in order to reduce the risk of reporting PCR artifacts as polymorphisms.

SNP detection. The sequences acquired for each animal were aligned using Contig Express in Vector NTI (Informax, Frederick, Mar.) and were examined for sequence variation against each other and the consensus *Bos taurus* CARD15 sequence (Fig. 4). This approach allowed us to identify polymorphisms within both the coding and flanking intronic sequences for each individual exon.

Haplotype determination. Haplotyper (Niu et al. 2002) was used to predict haplotypes for each individual from the genotype data. This analysis was performed using all 41 animals from the breed and Johne's panels simultaneously. SNPtagger was then employed to determine the minimum number and position of SNP markers necessary to distinguish among the observed haplotypes (Ke and Cardon 2003).

Results

SNPs in the breed panel. Thirty-six SNPs were identified in a thirty animal breed panel from a total of 6176 bases examined (Fig. 14). Nineteen SNPs were located in the translated portion of the gene (cSNPs) and, of these, eight were nonsynonymous and caused amino acid substitutions (nsSNPs). Of the remaining SNPs, ten were located in intronic sequence and seven within the last 500 bases of the 3'UTR. All identified SNPs are presented in Table 13. Each SNP is described relative to the reference allele present in the originally generated *taurus* sequence. SNPs located within intons are identified by reference to the nearest coding region. For example, *E2(-32)* refers to a SNP 32 bp upstream of exon 2, while *E8(+12)* refers to a SNP 12 bp downstream of exon 8.

Included in the polymorphism survey was a total of 1071 bp of intronic sequence flanking the CARD15 exons. Table 14 indicates the number of flanking bases surveyed for each exon. Of the SNPs identified in the intronic regions of the bovine CARD15 gene, two appear to be within potential regulatory regions and therefore may have a functional role. The *E2(-32)* SNP is located within a 7 bp motif (A**G**AAGcC) detected by homology between human, mouse and bovine (Table 12). The bolded 'G' indicates the SNP site, while the lower case 'c' represents a mismatch in the sequence motif between human, mouse and bovine. Further, the *E3(-6)* SNP is located within the first 6 bases upstream of exon 3 and the high level of conservation of these closely flanking sequences is thought to be involved in the regulation of intron splicing (Sorek and Ast 2003). Complete sequence data for intron 5 and revealed that this intron was completely conserved among all individuals. This was a surprising finding and again, this high level of sequence conservation may be related to the regulation of splicing.

Of the thirty-six identified SNPs, twenty-four segregate in only one of the subspecies, eight segregate in both (in 3 cases only one *taurus* individual, a longhorn, was segregating) and four had subspecies specific alleles. For five of the loci, all *indicus* animals were alternate homozygotes at the SNP locus. Segregation of SNP loci by subspecies is reported in Table 13. Examining the polymorphisms by protein domain,

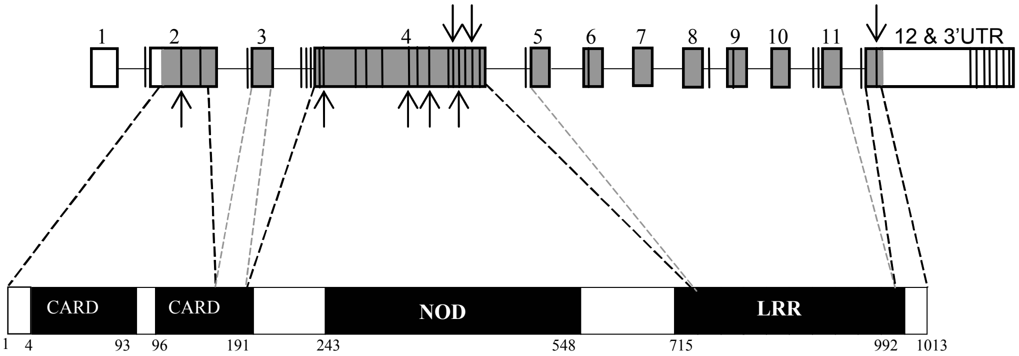


Figure 14. Location of SNP loci within bovine CARD15. Arrows indicate the location of nonsynonymous coding SNPs.

Table 13. Characteristics of bovine CARD15 SNP loci

SNP position and alleles[a]	*Domain affected*[b]	*Amino acid*	*Segregating*[c]	*Allele frequency*[d]	
				T	I
E2(-32) G/A	PRR		T&I	0.5	0.75
208 **A/G**	CARD1	**T70A**	T&I (LH)	0.98	0.5
363† C/T	CARD2	121	I	1	0.75
E3(-6) G/A/C	Splice site		T&I	0.8	0.38
E4(-58) C/T	Intron		T	0.92	0
E4(-22) A/G	Intron		T	0.92	0
E4(-16) T/G	Intron		T	0.92	0
570† T/C	CARD2	191	I vs T	1	0
586† **G/A**	LIBD	**V196M**	I	1	0.13
873† C/A	NOD	291	I	1	0.94
969† C/T	NOD	324	I	1	0.88
1194† C/T	NOD	398	AK (0.75)	1	1
1514† **C/A**	NOD	**T505N**	I	1	0.13
1569† T/C	NOD	523	I	1	0.31
1723† **T/C**	LIBD	**C575L**	I	1	0.44
1992† G/A	LIBD	664	I	1	0.81
2042† **G/A**	LIBD	**R681Q**	I	1	0.25
2066† **G/A**	LIBD	**R689H**	I	1	0.13
2145† C/T	LRR	715	I vs T	1	0
2197† **T/C**	LRR	**C733R**	I vs T	1	0
2364† C/T	LRR	788	T&I	0.98	0.19
E5(-11)† T/A	Intron		T	0.95	1
2481† G/A	LRR	827	AK (0.5)	1	1
E8(+12)† T/A	Intron		T	0.88	0
2787† C/T	LRR	929	I	1	0.88
E11(-14)† A/G	Intron		T	0.85	1
E11(-8)† C/T	Intron		T&AK(0.75)	0.97	1
E12(-6)† C/T	Intron		T&I	0.9	0.14
3020† **A/T**	LOD	**Q1007L**	T&AK(0.25)	0.88	1
4648 C/A	3'UTR (PRR)		T	0.83	1
4757 G/A	3'UTR (PRR)		T&I (LH)	0.97	0.25
4798 A/G	3'UTR (PRR)*		T&I	0.8	0.17
4801 A/T	3'UTR (PRR)*		T&I (LH)	0.95	0.17
4911 A/G	3'UTR (PRR)		I	1	0.83
5010 A/T	3'UTR		T	0.93	1
5098 C/T	3'UTR (PRR)		T	0.93	1

[a] Loci in bold face represent nonsynonymous mutations. † = Loci used to estimate haplotypes.

[b] PRR = Putative Regulatory Region. LIBD = Located in-between domains. LOD = Located outside of last domain. * = Located within the same regulatory motif.

[c] LH = Segregating in one Texas Longhorn (*taurus*) animal. AK = Ankole-Watusi.

[d] Frequency of the *Bos taurus* wild type allele in *Bos taurus* (T) and in *Bos indicus* (I).

Table 14. Size of sequenced bovine CARD15 intronic flanking regions

5' flank (bp)	*Exon*	*3' flank (bp)*
61	2	19
29	3	53
64	4	107
12	5	210[a]
210[a]	6	86
11	7	76
24	8	21
23	9	109
26	10	27
24	11	40
49	12	

[a] The entire 210 bp of intron 5 was sequenced.

reveals one site in CARD1, two sites in CARD2, 5 sites in the NOD and 5 sites in the LRR which harbor a total of 13 SNPs. Six SNP sites are located outside of domains and 7 sites were found within the 3'UTR (Table 13). Allele frequencies by subspecies reported in Table 13 are relative to the wild type allele detected within *Bos taurus* for each SNP. Sequence that included the SNPs located within intronic regions and in the 3' and 5'UTRs were analyzed using NSITE (available through SoftBerry at http://www.softberry.com/berry.phtml?topic=promoter) to identify putative regulatory motifs. Regions identified by the software to have homology to known regulatory regions are presented in Table 15.

A chi-squared test revealed that the frequency of synonymous and nonsynonymous (cSNPs) substitutions was greater in the *indicus* than the *taurus* population ($\chi^2 = 84.53$; 2 df). Chi-square tests were also used to reveal that the frequency of synonymous substitutions was the same as the frequency of nonsynonymous substitutions within *indicus* but the frequency of nonsynonymous substitutions greatly exceeded the frequency of synonymous substitutions in *taurus*. The numbers of synonymous and nonsynonymous cSNP substitutions within each subspecies are reported in Table 16.

Association tests. Within each subspecies, association tests were first performed by testing for SNP allele frequency differences between the group of animals in the breed panel (control group) and the group of sick animals (case group). Of the examined SNPs, 4 loci indicated a significant difference in allele frequency between the sick and control groups. Within *taurus*, the frequency of the wild type allele for *E2(-32)* was 1 in the case group and 0.5 within the control group ($\chi^2 = 10.8$; 1 df). Also within *taurus*, the frequency of the wild type allele for *3020* was 0.6 in the case group and 0.875 in the control group ($\chi^2 = 4.43$; 1 df). Both *E4(-58)* and *E4(-16)* revealed allele frequency differences between the *indicus* control and case groups($\chi^2 = 4.73$; 1 df); however, there was only a single animal within the case group and the genotype of this animal was atypical of the *indicus* genotypes within the control group for many of the loci.

Table 15. Putative regulatory motifs harboring SNPs within the bovine CARD15 3'UTR

SNP	*Motif*[a]	*TRANSFAC Accession #*
4648	CTC**c**CTCcaTTC	R02081
	TC**C**cTCCAtTCT	R02372
	CctCCATTCTCT	R04351
4757	**G**AATTTAATTAg	R00964
	TCTtAGcT**G**	R02003
	TAAtTCtTAgCT**G**A	R02819
4798	CAAGgT**A**gAAtCGG	R03055
4801	CAAGgTAgA**A**tCGG	R03055
4911	GA**A**TTTAtTTtA	R00964
5098	ATTtG**C**T	R01443
	TAcaTTTG**C**TaA	R02046

[a] Bold face nucleotide represents position of SNP in motif. Lower case represents sequence mismatches between bovine sequence and regulatory motifs.

Table 16. Synonymous and nonsynonymous substitutions at cSNPs within bovine CARD15

Subspecies[a]	*Synonymous substitutions*[b]	*Nonsynonymous substitutions*[c]
Bos taurus (N=30)	1/360 (0.28%)	9/240 (3.75%)
Bos indicus (N=9)	18/108 (16.67%)	16/69 (23.19%)

[a] Number of animals in parentheses.

[b] Number of synonymous substitutions/Number of genotypes at 12 sSNPs.

[c] Number of nonsynonymous substitutions/Number of genotypes at 8 nsSNPs.

Associations between CARD15 and disease status were next tested by examining differences in the frequencies of heterozygotes, homozygous mutants and heterozygotes + homozygous mutants across SNP loci. Tests were also performed for differences in the frequency of mutant alleles across SNP loci between the case and control groups. These analyses were performed at four levels of locus stratification: by considering all loci, ncSNPs, synonymous cSNPs and nonsynonymous cSNP loci. Within *Bos taurus*, there was no effect of multi-locus heterozygosity at CARD15 SNP loci on the likelihood of being classified within the case or control groups. However, there was a higher frequency of mutant alleles among the control group animals computed across all SNP loci (0.05 in control and 0.02 in case; $\chi^2 = 12.1$; 1 df) due in large part to a large difference in the ncSNPs (0.11 in control and 0.03 in case; $\chi^2 = 17.0$; 1 df). There was also a difference in allele frequency for sSNPs, but for this class the frequency of mutant alleles was greater in the case group (0.002 in control and 0.03 in case; $\chi^2 = 4.8$; 1 df). These allele frequency differences produced an increase in the frequencies of homozygous mutants and heterozygous + homozygous mutants in the case group for the ncSNPS and in homozygous mutants + heterozygotes for all SNPs. No associations were found between any mutant allele or genotype characteristic for nsSNPs. Within the *Bos indicus* population, there was a higher frequency of mutant alleles within the case group for all loci (0.11 in control and 0.36 in case; $\chi^2 = 32.5$; 1 df), ncSNPs (0.08 in control and 0.34 in case; $\chi^2 = 18.2$; 1 df), sSNPs (0.21 in control and 0.42 in case; $\chi^2 = 4.1$; 1 df) and nsSNPs (0.09 in control and 0.31 in case; $\chi^2 = 4.2$; 1 df). These differences also resulted in differences in genotypic frequencies, however, as previously discussed, there was only a single animal within this case group and the genotype of this animal was not typical of the control *Bos indicus* genotypes at most of the SNP loci.

Finally, the control and case groups were examined for differences in haplotype frequencies. For haplotype estimation, loci with missing genotypes were removed, leaving 23 of the original 36 loci which included all but one of the cSNPs (Table 13).

Table 17. Bovine CARD15 haplotypes estimated for 23 SNP loci

Population	*Control*	*Case*	*Total frequency*	*Haplotype*
Taurus	24	10	.415	00000000000000000000000
Taurus	2		.024	00000000000000000000010
Taurus	6	2	.098	00000000000000000001000
Taurus	4	8	.146	00000000000000000000001
N'Dama	2		.024	00000000000000000100000
Longhorn	1		.012	00000000000000100000101
N'Dama	2		.024	00000000000000010100010
Ankole	2		.024	00000000000000001000001
Ankole	1		.012	00000100000000000000101
Indicus	2		.024	01100011001111100100010
Indicus	3		.037	01100011111111100100010
Indicus	2		.024	01100010101111100100010
Indicus	1		.012	01100011101111100100010
Indicus	1		.012	01100011100111100100010
Indicus	1		.012	01100011100111000100010
Indicus	2		.024	11100011001111100100010
Indicus	1		.012	11100011001111100110000
Indicus	1		.012	11100010101111100100000
Indicus	1		.012	01001000000011000110000
Indicus	1		.012	01011000000011000100000
Indicus	0	1	.012	11001000111111000100010
Indicus	0	1	.012	11101010010011100110000

Twenty-two haplotypes were predicted by Haplotyper (Niu et al. 2002) from a total of 41 animals and these are reported in Table 17. All haplotypes were subspecies specific with four unique haplotypes explaining the diversity within 28 *Bos taurus* animals and 13 unique haplotypes explaining the diversity within 9 *Bos indicus* animals. One unique and three of the *Bos taurus* haplotypes were predicted in Texas Longhorn, while two unique haplotypes were each predicted in N'Dama and Ankole. Loci that were not included in the haplotype analysis include cSNP *208*, all of the 3'UTR SNPs and the SNPs from introns 1 and 3. It was impossible to complete genotypes for all animals for these loci due to depletion of the available DNA samples. To test for haplotypic associations with Johne's disease the animals were again separated by subspecies but excluded the Texas Longhorn, N'Dama and Ankole-Watusi animals from the *Bos taurus* control group due to the presence of haplotypes that were not represented within the *Bos taurus* case group, which comprised Holsteins and Jerseys. Despite the presence of unique *indicus* haplotypes within the case *Bos indicus* animal, no significant association between haplotypes and the risk of Johne's disease was found in either the *Bos taurus* or the *Bos indicus* animals.

Tagged SNPs. SNPtagger (Ke and Cardon 2003) was used to identify the minimum number of SNPs required to completely represent the diversity represented within the predicted haplotypes in Table 17. Only 14 of the 23 SNPs must be scored in order to completely predict these 22 haplotypes (Table 18). However only three loci are required to generate the haplotypes for *Bos taurus* if Texas Longhorn, N'Dama and Ankole-Watusi are excluded. Including these breeds requires the use of 7 SNP loci. A total of 7 SNPs are required to regenerate the 13 haplotypes found in *Bos indicus*.

Eight PCR reactions are required to amplify the fragments harboring the 14 SNPs required to distinguish all 22 predicted haplotypes. Three PCR reactions are required to amplify the 7 SNP sites represented in the *Bos indicus* assay and 5 PCR reactions are required to amplify the 7 SNP sites to distinguish all *Bos taurus* haplotypes. If N'Dama and Ankole are excluded, only 2 PCR reactions are required to assay the SNPs required to distinguish among the *Bos taurus* (including Texas Longhorn) haplotypes. A list of

Table 18. SNPs within bovine CARD15 predicted to discriminate among *Bos indicus* and *Bos taurus* haplotypes

SNP[a]	Sense 5' to 3'	Antisense 5' to 3'	Anneal Temp	Approx. product size
363**	CTCAGTTTGAACACCTGTACAATGG	AGGAGGACTATGACCCACATCTC	52	690
570	TGGGTTCCTACCTGCAAAAC	CACTGAGGGCGAGGTAGAAT	52/1min ext	2250
1194*	CAGTCCAGAGTCTGCTCTTCAAC	GCCATTCCATGAATTTCAACTATC	55/1min ext	1360
1569**	“	“	“	“
1723**	“	“	“	“
1992**	“	“	“	“
2042**	“	“	“	“
2364**	“	“	“	“
2481*	GTCTCTGTGGGGGTTTTGTC	GTCCTTGTTTTCAGCGAGGT	55	500
E8(+12)*	CACTTGCTGGGACCTGAGT	CCCTCCTCACACTGGCTTC	53	200
2787**	GCATTTTGCCCTTCTTGAGT	ACGCAGTCATCCATCTTGGT	53	200
E11(-14)*	CCAGCTCCCAAAGTCTCCTT	GAGGCTCAGAGAGGTTAAAGAGG	53	200
E11(-8)*	“	“	“	“
E12(-6)*	AGGTTTACAAAGCAGCATCTTCC	ATGTCACCTGTCCCCTTTCA	58/1min ext	750
3020*	“	“	“	“

[a] All loci except *1194* are used to distinguish the 22 haplotypes present in *Bos indicus* and *Bos taurus*. * = SNPs used in the *Bos taurus* assay including Texas Longhorn, N’Dama and Ankole. Bolded SNPs used in the *Bos taurus* assay excluding Texas Longhorn, N’Dama and Ankole. ** = SNPs used in the *Bos indicus* assay.

the SNPs required for each assay and of the primers required to amplify the regions of bovine CARD15 harboring these SNPS are presented Table 18.

Discussion

A panel of cattle was surveyed representing 9 different breeds and 2 subspecies in order to gain insight as to the extent of natural variation within the bovine CARD15 gene. All 12 exons were examined including the coding region of the gene (3042 bp), the 5'UTR which includes the first exon and 8 bp of the second exon (148 bp), the 3'UTR comprising the last 1988 bp of exon 12 and 1071 bp of flanking intronic sequence. Obtaining this sequence required 14 separate PCR reactions and 28 sequencing reactions. From this sequence, 36 SNP sites were identified in 6176 bp indicating a SNP, on average, every 172 bases (including flanking intronic sequence). Twenty-six SNPs were found in the 5105 bp transcript, indicating a SNP, on average, every 196 bp. These results are consistent with previous reports for the rate of occurrence of SNPs within bovine coding sequence (Heaton et al. 2001).

Genetic heterozygosity is thought to enhance the resistance of hosts to infectious diseases. For example, it is believed that heterozygosity at the MHC loci may enhance resistance to infectious diseases by increasing the diversity of antigens presented to T cells (Doherty and Zinkernagel 1975). Furthermore, it has been shown that diversity in the leucine rich repeat region (LRR) of the plant disease resistance genes allows for recognition of different pathogen components (Parniske et al. 1997). Although it is difficult to make definitive conclusions because of the small sample size employed in this study, a significantly higher rate of synonymous and nonsynonymous substitutions within *Bos indicus* than in *Bos taurus* was found which is consistent with previous findings (MacHugh et al. 1997) and the fact that *Bos indicus* cattle are known to possess greater disease and parasite resistance than *Bos taurus* cattle. There was also a significantly higher rate of nonsynonymous to synonymous substitutions in *Bos taurus* indicating that positive Darwinian selection is acting on the gene within this subspecies. Within *Bos indicus*, the rate of nonsynonymous substitutions the same as the rate of synonymous substitutions indicating selective neutrality. Due to the small sample size and admixture of breeds within both the *Bos taurus* and *Bos indicus* samples, the individual loci were not tested for Hardy-Weinberg equilibrium. However, visual inspection of the data suggested a deficit of heterozygotes within each of the loci. This was further reinforced when the program Haplotyper was used to infer haplotypes based upon all 36 SNP loci. The data contained missing data for animals whose DNA samples had been exhausted and Haplotyper invariably estimated the missing genotypes to be heterozygotes, presumably because the underlying model assumes panmixia, although Niu et al. (2002) indicate the program can be used for inference within stratified populations. For this reason, loci with missing data (predominantly ncSNPs) were dropped from the haplotype analysis reducing the number of SNPs to 23 used to infer haplotypes.

In human, mutations within the NBD are associated with Blau Syndrome and mutations in the LRR are associated with Crohn's disease. Five mutation sites were identified, including one nonsynonymous site, in each of the NOD and LRR domains. In human, the mutations associated with Crohn's disease are believed to result in a defect in recognition of pathogen components resulting in aberrant NF-kB activation. If this is the case, variation in gene expression mediated by the efficiency of transcription, mRNA stability or translation may also result in the aberrant activation of NF-kB. Variation in regulatory regions of the gene that result in a reduction of CARD15 protein may mimic the effects of the human CARD15 mutations in the LRR that are associated with Crohn's disease. This prompted the examination of the 5' and 3'UTRs which are known to be important regulatory regions of genes. Intronic sequences that flank CARD15 were also examined since these regions have recently been shown to harbor regions responsible for the regulation of alternative splicing of exons.

Two interesting SNPs were found within the CARD15 flanking intronic regions. *E2(-32)* is located in a short motif of 7 bp that is conserved in human, mouse and bovine. Known regulatory binding motifs consist of 5-15 nucleotides and have tolerance for mismatches located within the sequence. Thus, it is possible that the *E2(-32)* SNP creates variability within an important regulatory region in the cattle CARD15 gene. SNP *E3(-6)* was identified which falls within 6 bases of the start of the third exon. Sorek and Ast (2003) found evidence that the first seven bases flanking the 5' end of an exon are involved in the regulation of splicing of the intron which makes this an interesting SNP from the perspective of regulating the production of protein isoforms. No SNPs were found within the 5'UTR among the cattle tested which suggests that conservation of this region is necessary to preserve function in cattle. However, the 5'UTR was not conserved between human, mouse and bovine and since this region is important for the effective translation of the gene, the mechanisms for accomplishing this have evolved differently among the species. An additional 7 SNP sites were identified within the last 500 bases of the 3'UTR. The bovine 3'UTR does not align with human or mouse and contains additional sequence not found in these species. Thus, this region of the bovine gene may be free to evolve, or it may have evolved new functional elements for the regulation of transcript expression, localization or stability. While there are no recognized 3'UTR regulatory elements in this region, the sequence was scanned for known regulatory motifs and 6 of the 7 SNPs were within sequence motifs known to regulate transcription.

In the association studies, the frequency of the *E2(-32)* 'G' allele was significantly greater in the *Bos taurus* case group than in the control group. In fact, this allele was fixed within the case group but only at a frequency of 50% among the controls where all genotyped individuals were homozygotes. If this locus plays a role in susceptibility to Johne's disease, the lack of complete penetrance suggests that susceptibility is either polygenic, environmentally influenced or that the remaining *GG* and control animals were not exposed to the mycobacterium. However, it seems most likely that this locus is free to evolve and that it was the small sample size alone that led to the spurious association.

The lone Brahman animal within the *Bos indicus* case group had a multilocus genotype that was not typical of the *Bos indicus* animals with the control group. This animal tended to be either heterozygous (8 out of 23 loci included in the haplotype analysis) or homozygous for *Bos taurus* alleles suggesting that the animal may have been a hybrid at the CARD15 locus. This sample size effect made impossible any meaningful association analysis within *Bos indicus*. However, it is nevertheless interesting to note that the single *Bos indicus* animal with Johne's disease was genetically extremely dissimilar to its *Bos indicus* control counterparts at the CARD15 locus.

Haplotype analysis has recently gained favor for application in genome-wide association studies due to the increase in power that can be achieved through the integration of information at closely linked SNP loci. In this study, the *Bos taurus* haplotypes were fewer in number and completely distinct from the *Bos indicus* haplotypes. Only four haplotypes were found among the Angus, Limousin, Jersey and Holstein breeds and this number increased to 5 with the addition of Texas Longhorn, 7 with the addition of N'Dama and 9 with the addition of Ankole-Watusi. Conversely, 13 distinct haplotypes were found among the 9 *Bos indicus* animals. It is well known that there is more genomic diversity among *Bos indicus* than among *Bos taurus*, however, why this is the case is unclear. It may reflect the relatively recent history of development of the Brahman breed which involved crossing among Nellore, Gir and Guzerat along with grading-up from crosses to various *Bos taurus* breeds. On the other hand, the *Bos taurus* breeds represented in this study have been within the U.S. for many more generations than *Bos indicus* cattle, have been maintained as pure breeds and have been under effective selection for various production traits for at least 20 generations. Nevertheless, no associations between CARD15 haplotypes and the prevalence of Johne's disease were found in this study.

Finally it is important to consider the inherent difficulty of designing experiments to unequivocally determine the identities of genes involved (or not involved) in resistance to diseases which require some type of environmental exposure and where resistance may be polygenic. First consider the issues underlying phenotype definition in both the control and case animals. For many diseases, it is difficult to define the phenotype associated with disease. In the case of Johne's disease, some animals respond positively to ELISA and culture tests, but never display any clinical signs of the disease. These animals have clearly been infected with the causal agent and are able to propagate the spread of the pathogen but whether or not they will ever develop clinical indications of disease is uncertain. Should these animals be classified as having Johne's disease when the definitive diagnosis of Johne's disease requires necropsy? Similarly, animals will be identified that respond negatively to ELISA and culture tests and thus clearly have not been infected by *M. ptb.* and will not develop Johne's disease based upon their current status. However, it is not clear if these animals have ever been exposed to the pathogen or whether they were exposed at some stage and were resistant to infection. Thus, these control animals may either be susceptible animals that would have developed an infection had they been exposed or they are genetically resistant. Consequently, the control group of animals will usually represent an admixture of genotypes that are representative of susceptible and resistant animals.

It may be acceptable to characterize animals as being subclinically infected if they shed <10 cfu/g feces and clinically infected if they shed >100 cfu/g feces (Stabel pers. comm.). While this phenotype may be useful from the perspective of identifying genes responsible for the elimination of the mycobacterium if control of spread of disease is considered the goal, it may not be wholly appropriate for management of productivity loss if the infected animals progress to express clinical signs of the disease. Ideally, the phenotyping of animals should be based upon their development of granulomas in the ileum rendering the animal inefficient and resulting in weight loss and low productivity. However, the impracticality of this approach is not only economic, but also because of the slow onset of disease in infected animals. It is for these reasons that animals were only included in this study that had either been necropsied and confirmed to have Johne's disease, or that tested ELISA and culture positive and were exhibiting clinical signs of Johne's disease.

Even in the presence of the perfect experimental design including true controls (exposed but not infected) and true Johne's disease cases, there remains a significant problem in the event that disease resistance or susceptibility is polygenic, epistatic or requires some unknown environmental stimulus. In the case of polygenic or epistatic inheritance, several genetic architectures may lead to the expression of the phenotype of interest and very large populations are required to elucidate the genetic mechanisms underlying the disease phenotype. Of course, this may also require a whole genome association study to be performed to detect the effects of trans-acting mutations.

Considering all of these issues, the naturally occurring variation present in the bovine CARD15 gene in *Bos taurus* and *Bos indicus* animals is reported. A complete survey of SNPs in 11 diseased animals was performed and then characteristics of these loci between the diseased animals and those animals present in the breed panel were compared. There was no overwhelming evidence for a simple causal relationship between variation in the bovine CARD15 gene and Johne's disease. However, the small sample size, the admixture of breeds present in the control group and the fact that no animal within the control group was challenged with *M. ptb.* could easily have conspired to obscure a causal genetic relationship between CARD15 and disease. Consequently, it is important to utilize the SNP loci developed in this study to screen larger populations of animals in which a more robust experimental design has been possible in terms of breed composition and phenotype definition of the control group membership.

CONCLUSIONS AND DISCUSSION

In these studies, the bovine CARD15 gene has been characterized and tools have been made available for future researchers working in livestock disease resistance. The first study provided the nucleotide sequence and genomic localization for the previously unidentified bovine CARD15 gene. Both should prove useful in future QTL or candidate gene studies targeting genes associated with disease resistance or susceptibility. An assay has been developed mRNA expression profiles have been provided for the bovine CARD15 gene in a large variety of tissues, which should prove useful in future mRNA expression studies involving the bovine CARD15 gene.

In the second study, a comparative sequence analysis of the bovine CARD15 gene to its human and mouse homologs is provided. In this analysis sequence for flanking intronic regions that have not previously been examined is included. High levels of sequence conservation within the coding portion of the gene is reported, but very low levels of interspecies conservation in both of the UTRs and in the flanking intronic sequence with the exception of several short, conserved motifs (6-19 nucleotides). These conserved motifs could function as putative regulatory regions and homology in the regions harboring these motifs to previously identified regulatory motifs is reported. Information is provided concerning diversity of the UTRs among the species, including an additional 100 bases in the mouse 3'UTR and more than 700 additional bases in the bovine 3'UTR when compared to human. Multiple unique polyadenylation signal motifs in all three species were identified. Furthermore, it is reported that the bovine CARD15 gene possesses only one in-frame translation initiation site, whereas both the mouse and human forms of the gene have 2 such sites. Finally, it is demonstrated that the 3' and 5'UTRs in all three species contain different regulatory motifs. Considering that the UTRs of genes often contain key regulatory elements involved in the transcriptional and post-transcriptional regulation of gene expression, these interspecies differences are perhaps indicators that these genes have evolved to meet particular, but different regulatory needs in each organism.

In the third study, a survey of the extent of natural variation present in the bovine CARD15 gene is provided and association analyses between these mutations and Johne's disease is performed. This will prove a useful tool in future comparisons involving this gene and disease resistance or susceptibility. There is no overwhelming evidence for a simple causal relationship between variation in the bovine CARD15 gene and Johne's disease. However, a large number of polymorphisms located within the last 500 bases of the 3'UTR are reported. While homology in this region to typical 3'UTR regulatory elements was not detected, homology to short protein binding motifs typically found in enhancers was identified. It is possible these regions possess regulatory functions not commonly ascribed to the 3'UTR. Future studies testing hypotheses concerning the role of the 3'UTR in the enhancement or repression of the CARD15 transcript appear to be interesting and warranted. Sets of SNPs are provided which are completely predictive of multi-locus haplotypes in *Bos taurus* and *Bos indicus* which may be used to screen populations in future disease resistance association studies.

Overall, these studies provide a foundation for future studies addressing the role of CARD15 as an intrinsic disease resistance gene. Although no overwhelming evidence for a simple causal relationship between variation in the bovine CARD15 gene and Johne's disease was found, the relationship could have been masked due to the small sample size and lack of control animals known to have been exposed to *M. ptb.* Consequently, it is important to utilize the SNP loci developed in this study to screen larger populations of animals in which a more robust experimental design has been possible in terms of breed composition and phenotype definition of the control group membership. If the mutations identified in this study are not associated with susceptibility to Johne's disease then other regions of the gene such as the introns and the promoter should be interrogated for polymorphisms before discounting CARD15 as a candidate gene. Therefore, the gene remains an excellent candidate gene for disease resistance in cattle and its associations with other disease phenotypes should now be considered.

REFERENCES

1. Aderem A, Ulevitch RJ (2000) Toll-like receptors in the induction of the innate immune response. Nature 406, 782-787

2. Akira S, Takeda K, Kaisho T (2001) Toll-like receptors: critical proteins linking innate and acquired immunity. Nat Immunol 2, 675-680

3. Audibert A, Simonelig M (1998) Autoregulation at the level of mRNA 3' end formation of the suppressor of forked gene of *Drosophila melanogaster* is conserved in *Drosophila virilis*. Proc Natl Acad Sci U S A 95, 14302–14307

4. Bertin J, Nir WJ, Fischer CM, Tayber OV, Errada PR et al. (1999) Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF-kappaB. J Biol Chem 274, 12955-12958

5. Bishop SC, MacKenzie KM (2003) Genetic management strategies for controlling infectious diseases in livestock populations. Genet Sel Evol 35, Suppl 1, S3-17

6. Borchert A, Savaskan NE, Kuhn H (2003) Regulation of expression of the phospholipid hydroperoxide/sperm nucleus glutathione peroxidase gene. Tissue-specific expression pattern and identification of functional cis- and trans-regulatory elements. J Biol Chem 278, 2571-2580

7. Brend T, Gilthorpe J, Summerbell D, Rigby PW (2003) Multiple levels of transcriptional and post-transcriptional regulation are required to define the domain of Hoxb4 expression. Development 130, 2717-2728

8. Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A (1999) The dynamics of chromosome evolution in birds and mammals. Nature 402, 411-413

9. Cai L, Taylor JF, Wing RA, Gallagher DS, Woo SS et al. (1995) Construction and characterization of a bovine bacterial artificial chromosome library. Genomics 29, 413-425

10. Cavanaugh JA, Callen DF, Wilson SR, Stanford PM, Sraml ME et al. (1998) Analysis of Australian Crohn's disease pedigrees refines the localization for susceptibility to inflammatory bowel disease on chromosome 16. Ann Hum Genet 62, 291-298

11. Chevalet C, Corpet F (1986) Statistical decision rules concerning synteny or independence between markers. Cytogenet Cell Genet 43, 132-139

12. Chiodini RJ, Van Kruiningen HJ, Merkal RS (1984a) Ruminant paratuberculosis (Johne's disease): the current status and future prospects. Cornell Vet 74, 218-262

13. Chiodini RJ, Van Kruiningen HJ, Merkal RS, Thayer WR Jr, Coutu JA (1984b) Characteristics of an unclassified *Mycobacterium* species isolated from patients with Crohn's disease. J Clin Microbiol 20, 966-971

14. Chiodini RJ (1989) Crohn's disease and the mycobacterioses: a review and comparison of two disease entities. Clin Microbiol Rev 2, 90-117

15. Collins MT, Lisby G, Moser C, Chicks D, Christensen S et al. (2000) Results of multiple diagnostic tests for *Mycobacterium avium subsp. paratuberculosis* in patients with inflammatory bowel disease and in controls. J Clin Microbiol 38, 4373-4381

16. Coloe PJ, Slattery JF, Cavanaugh P, Vaughan J (1986) The cellular fatty acid composition of *Campylobacter* species isolated from cases of enteritis in man and animals. J Hyg (London) 96, 225-229

17. Cocito C, Gilot P, Coene M, de Kesel M, Poupart P et al. (1994) Paratuberculosis. Clin Microbiol Rev 7, 328-345

18. Corpet F, Servant F, Gouzy J, Kahn D (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. Nucleic Acids Res 28, 267-269

19. Crohn B, Ginzburg L, Oppenheimer G (1932) Regional ileitis, a pathological and clinical entity. JAMA 99, 1323-1329

20. Croucher PJ, Mascheretti S, Hampe J, Huse K, Frenzel H et al. (2003) Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. Eur J Hum Genet 11, 6-16

21. Cuvelier C, Mielants H, De Vos M, Quatacker J, Veys E (1994) Idiopathic inflammatory bowel diseases: immunological hypotheses. Acta Gastroenterol Belg 57, 292-299

22. Deguillien M, Huang SC, Moriniere M, Dreumont N, Benz EJ Jr et al. (2001) Multiple cis elements regulate an alternative splicing event at 4.1R pre-mRNA during erythroid differentiation. Blood 98, 3809-3816

23. Dell'Isola B, Poyart C, Goulet C, Mougenot JF, Sadoun-Journo E (1994) Detection of *Mycobacterium paratuberculosis* by polymerase chain reaction in children with Crohn's disease. J Infec Dis 169, 449-451

24. Dixon MS, Golstein C, Thomas CM, van Der Biezen EA, Jones JD (2000) Genetic complexity of pathogen perception by plants: the example of Rcf3, a tomato gene required specifically by Cf-2. Proc Natl Acad Sci U S A 97, 8807-8814

25. Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256, 50-52

26. Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3, 278-284

27. Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene L that determine differences in gene-for-gene specificity. Plant Cell 11, 495-506

28. Fidler HM, Thurrell W, Johnson NM, Rook GA, McFadden JJ (1994) Specific detection of *Mycobacterium paratuberculosis* DNA associated with granulomatous tissue in Crohn's disease. Gut 35, 506-510

29. Giacopelli F, Rosatto N, Divizia MT, Cusano R, Caridi G et al. (2003) The first intron of the human osteopontin gene contains a C/EBP-beta-responsive enhancer. Gene Expr 11, 95-104

30. Girardin SE, Boneca IG, Carneiro LA, Antignac A, Jehanno M et al. (2003) Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. Science 300, 1584-1587

31. Gitnick G, Collins J, Beaman B, Brooks D, Arthur M et al. (1989) Preliminary report on isolation of mycobacteria from patients with Crohn's disease. Dig Dis Sci 34, 925-932

32. Goldammer T, Kata SR, Brunner RM, Dorroch U, Sanftleben H et al. (2002) A comparative radiation hybrid map of bovine chromosome 18 and homologous chromosomes in human and mice. Proc Natl Acad Sci U S A 99, 2106-2111

33. Graber JH, Cantor CR, Mohr SC, Smith TF (1999) *In silico* detection of control signals: mRNA 3'-end-processing sequences in diverse species. Proc Natl Acad Sci U S A 96, 14055-14060

34. Graham DY, Markesich DC, Yoshimura HH (1987) Mycobacteria and inflammatory bowel disease. Results of culture. Gastroenterology 92, 436-442

35. Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci U S A 87, 1213-1217

36. Greener MJ, Sewry CA, Muntoni F, Roberts RG (2002) The 3'-untranslated region of the dystrophin gene - conservation and consequences of loss. Eur J Hum Genet 10, 413-420

37. Gui GP, Thomas PR, Tizard ML, Lake J, Sanderson JD et al.(1997) Two-year outcomes analysis of Crohn’s disease treated with Rifabutin and macrolide antibiotics. J Antimicrob Chemother 39, 393-404

38. Gutierrez O, Pipaon C, Inohara N, Fontalba A, Ogura Y et al. (2002) Induction of Nod2 in myelomonocytic and intestinal epithelial cells via nuclear factor-kappa B activation. J Biol Chem 277, 41701-41705

39. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S et al. (2001) Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 357, 1925-1928

40. Hardison RC, Oeltjen J, Miller W (1997) Long human-mouse sequence alignments reveal novel regulatory elements: a reason to sequence the mouse genome. Genome Res 7, 959-966

41. Heaton MP, Grosse WM, Kappes SM, Keele JW, Chitko-McKown CG et al. (2001) Estimation of DNA sequence diversity in bovine cytokine genes. Mamm Genome 12, 32-37

42. Hu Y, Ding L, Spencer DM, Nunez G (1998) WD-40 repeat region regulates Apaf-1 self-association and procaspase-9 activation. J Biol Chem 273, 33489-33494

43. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC et al. (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 379, 821-823

44. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP et al. (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411, 599-603

45. Inohara N, Koseki T, del Peso L, Hu Y, Yee C et al. (1999) Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB. J Biol Chem 274, 14560-14568

46. Inohara N, Nunez G (2001) The NOD: a signaling module that regulates apoptosis and host defense against pathogens. Oncogene 20, 6473-6481

47. Inohara N, Ogura Y, Chen FF, Muto A, Nunez G (2001) Human Nod1 confers responsiveness to bacterial lipopolysaccharides. J Biol Chem 276, 2551-2554

48. Inohara N, Nunez G (2003) NODs: intracellular proteins involved in inflammation and apoptosis. Nat Rev Immunol 3, 371-382

49. Iwanaga Y, Davey MP, Martin TM, Planck SR, DePriest ML et al. (2003) Cloning, sequencing and expression analysis of the mouse NOD2/CARD15 gene. Inflamm Res 52, 272-276

50. Jakobsen MB, Alban L, Nielsen SS (2000) A cross-sectional study of paratuberculosis in 1155 Danish dairy cows. Prev Vet Med 46, 15-27

51. Janeway CA Jr (1989) Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb Symp Quant Biol 54, 1-13

52. Ke X, Cardon LR (2003) Efficient selective screening of haplotype tag SNPs. Bioinformatics 19, 287-288

53. Koets AP, Adugna G, Janss LL, van Weering HJ, Kalis CH et al. (2000) Genetic variation of susceptibility to *Mycobacterium avium subsp. paratuberculosis* infection in dairy cattle. J Dairy Sci 83, 2702-2708

54. Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J et al. (2002) CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. Am J Hum Genet 70, 845-857

55. Levy S, Avni D, Hariharan N, Perry RP, Meyuhas O (1991) Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. Proc Natl Acad Sci U S A 88, 3319-3323

56. Lim LP, Sharp PA (1998) Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. Mol Cell Biol 18, 3900-3906

57. Lisby G, Andersen J, Engbaek K, Binder V (1994) *Mycobacterium paratuberculosis* in intestinal tissue from patients with Crohn's disease demonstrated by a nested primer polymerase chain reaction. Scand J Gastroenterol 29, 923-929

58. MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). Genetics 146, 1071-1086

59. McClure HM, Chiodini RJ, Anderson DC, Swenson RB, Thayer WR et al. (1987) *Mycobacterium paratuberculosis* infection in a colony of stumptail macaques (*Macaca arctoides*). J Infect Dis 155, 1011-1019

60. McFadden J, Collins J, Beaman B, Arthur M, Gitnick G (1992) Mycobacteria in Crohn's disease: DNA probes identify the wood pigeon strain of *Mycobacterium*

avium and *Mycobacterium paratuberculosis* from human tissue. J Clin Microbiol 30, 3070-3073

61. Miceli-Richard C, Lesage S, Rybojad M, Prieur AM, Manouvrier-Hanu S et al. (2001) CARD15 mutations in Blau syndrome. Nat Genet 29, 19-20

62. Mishina D, Katsel P, Brown ST, Gilberts EC, Greenstein RJ (1996) On the etiology of Crohn disease. Proc Natl Acad Sci U S A 93, 9816-9820

63. Moss MT, Sanderson JD, Tizard ML, Hermon-Taylor J, el-Zaatari FA et al. (1992) Polymerase chain reaction detection of *Mycobacterium paratuberculosis* and *Mycobacterium avium subsp silvaticum* in long term cultures from Crohn's disease and control tissues. Gut 33, 1209-1213

64. Murillo L, Crusius JB, van Bodegraven AA, Alizadeh BZ, Pena AS (2002) CARD15 gene and the classification of Crohn's disease. Immunogenetics 54, 59-61

65. Niu T, Qin ZS, Xu X, Liu JS (2002) Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. Am J Hum Genet 70, 157-169

66. O'Brien AD, Rosenstreich DL, Scher I, Campbell GH, MacDermott RP et al. (1980) Genetic control of susceptibility to *Salmonella typhimurium* in mice: role of the LPS gene. J Immunol 124, 20-24

67. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF et al. (2001a) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411, 603-606

68. Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S et al. (2001b) Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. J Biol Chem 276, 4812-4818

69. Ogura Y, Saab L, Chen FF, Benito A, Inohara N et al. (2003) Genetic variation and activity of mouse Nod2, a susceptibility gene for Crohn's disease. Genomics 81, 369-377

70. Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA et al. (1997) Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. Cell 91, 821-832

71. Pesole G, Liuni S (1999) Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNAs. Trends Genet 15, 378

72. Philipsson J, Thafvelin B, Hedebro-Velander K (1980) Genetic studies on disease recording in first lactation cows of Swedish dairy breeds. Acta Agric Scand 30, 327-334

73. Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF (1998) A family of human receptors structurally related to *Drosophila* Toll. Proc Natl Acad Sci U S A 95, 588-593

74. Sanderson JD, Moss MT, Tizard ML, Hermon-Taylor J (1992) *Mycobacterium paratuberculosis* DNA in Crohn's disease tissue. Gut 33, 890-896

75. Schmidt S, Bork P, Dandekar T (2002) A versatile structural domain analysis server using profile weight matrices. J Chem Inf Comput Sci 42, 405-407

76. Schwartz D, Shafran I, Romero C, Piromalli C, Biggerstaff J et al. (2000) Use of short-term culture for identification of *Mycobacterium avium subsp. paratuberculosis* in tissue from Crohn's disease patients. Clin Microbiol Infect 6, 303-307

77. Simianer H, Solbu H, Schaeffer LR (1991) Estimated genetic correlations between disease and yield traits in dairy cattle. J Dairy Sci 74, 4358-4365

78. Slonim D, Kruglyak L, Stein L, Lander E (1997) Building human genome maps with radiation hybrids. J Comput Biol 4, 487-504

79. Smith TP, Grosse WM, Freking BA, Roberts AJ, Stone RT et al. (2001) Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. Genome Res 11, 626-630

80. Sonnhammer ELL, Kahn D (1994) Modular arrangement of proteins as inferred from analysis of homology. Protein Sci 3, 482-492

81. Sorek R, Ast G (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13, 1631–1637

82. Stabel JR (1998) Johne's disease: a hidden threat. J Dairy Sci 81, 283-288

83. Sweeney RW (1996) Transmission of paratuberculosis. Vet Clin North Am Food Anim Pract 12, 305-312

84. Takeuchi O, Hoshino K, Kawai T, Sanjo H, Takada H et al. (1999) Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity 11, 443-451

85. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM et al. (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424, 788-793

86. Thompson DE (1994) The role of mycobacteria in Crohn's disease. J Med Microbiol 41, 74-94

87. Thorel MF, Blom-Potar MC, Rastogi N (1990) Characterization of *Mycobacterium paratuberculosis* and "wood-pigeon" mycobacteria by isoenzyme profile and selective staining of immunoprecipitates. Res Microbiol 141, 551-561

88. United States Department of Agriculture (1997) Johne’s disease on US dairy operations. N245.1097. United States Department of Agriculture Centers for Epidemiology and Animal Health, Animal and Plant Health Inspection Service, Veterinary Services. National Animal Health Monitoring System. Fort Collins, Colorado.

89. Van Kruiningen HJ, Chiodini RJ, Thayer WR, Coutu JA, Merkal RS et al. (1986) Experimental disease in infant goats induced by a *Mycobacterium* isolated from a patient with Crohn's disease. A preliminary report. Dig Dis Sci 31, 1351-1360

90. Van Kruiningen HJ, Ruiz B, Gumprecht L (1991) Experimental disease in young chickens induced by a *Mycobacterium paratuberculosis* isolate from a patient with Crohn's disease. Can J Vet Res 55, 199-202

91. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C et al. (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425, 832-836

92. Vermeire S, Wild G, Kocher K, Cousineau J, Dufresne L et al. (2002) CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure. Am J Hum Genet 71, 74-83

93. Wal S, Kunze ZM, Saboor S, Soufleri I, Seechurn P et al. (1993) Identification of spheroplast-like agents isolated from tissues of patients with Crohn’s disease and control tissues by polymerase chain reaction. J Clin Microbiol 30, 3070-3073

94. Warren BF, Watkins PE (1994) Animal models of inflammatory bowel disease. J Pathol 172, 313-316

95. White SN (2003) Comparative radiation hybrid mapping and haplotype analysis of bovine TLR4. Ph.D. Dissertation Texas A&M University

96. White SN, Kata SR, Womack JE (2003) Comparative fine maps of bovine toll-like receptor 4 and toll-like receptor 2 regions. Mamm Genome 14, 149-155

97. Whitlock RH, Buergelt C (1996) Preclinical and clinical manifestations of paratuberculosis (including pathology). Vet Clin North Am Food Anim Pract 12, 345-356

98. Womack JE, Moll YD (1986) Gene map of the cow: conservation of linkage with mouse and man. J Hered 77, 2-7

99. Womack JE, Johnson JS, Owens EK, Rexroad CE 3rd, Schlapfer J et al. (1997) A whole-genome radiation hybrid panel for bovine gene mapping. Mamm Genome 8, 854-856

100. Yang RB, Mark MR, Gray A, Huang A, Xie MH et al. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. Nature 395, 284-288

101. Yokoyama Y, Kino J, Okazaki K, Yamamoto Y (1994) Mycobacteria in the human intestine. Gut 35, 715-716

102. Zhang G, Ghosh S (2001) Toll-like receptor-mediated NF-kappaB activation: a phylogenetically conserved paradigm in innate immunity. J Clin Invest 107, 13-19

103. Zwick LS, Walsh TF, Barbiers R, Collins MT, Kinsel MJ et al. (2002) Paratuberculosis in a mandrill (*Papio sphinx*). J Vet Diagn Invest 14, 326-328.

VITA

Kristen Hawkins Taylor
3305 Westcreek Circle
Columbia, MO 65203

Educational Background:
Ph.D. Genetics, Texas A&M University, 2004
B.S. Secondary Education, Lamar University, 1989

Professional Experience:

1999-2004	Graduate Research Assistant, Texas A&M University
2000-2001	Research Associate, GenomicFX, Austin, TX
Summer 1998-Spring 1999	Graduate Teaching Assistant, Genetics 301, Texas A&M University
1994-1998	High school Biology teacher and coach, Bryan High School, Bryan TX
1990-1994	High school Biology teacher and coach, Milby High School, Houston, TX