

# BoLA class I allele diversity and polymorphism in a herd of cattle

Shawn Babiuk · Benjamin Horseman ·  
Chenhong Zhang · Mik Bickis · Anthony Kusalik ·  
Lawrence B. Schook · Mitchell S. Abrahamsen ·  
Reno Pontarollo

Received: 22 January 2006 / Accepted: 24 October 2006 / Published online: 6 January 2007  
© Springer-Verlag 2007

**Abstract** Major histocompatibility complex class I genes are among the most polymorphic genes characterized. The high level of polymorphism is essential for generating host immune responses. In humans, three distinct genomic loci encode human leukocyte antigen (HLA) class I genes, allowing individuals to express up to six different HLA class I molecules. In cattle, the number of distinct genomic loci are currently at least six, and the number of different bovine leukocyte antigens (BoLA) class I molecules that

are expressed in individual animals are variable. The extent of allele variation within the cattle population is unknown. In this study, the number and variety of BoLA class I sequences expressed by 36 individuals were determined from full-length BoLA class I cDNA clones. Twenty distinct BoLA class I alleles were identified, with only four being previously reported. The number of expressed BoLA class I alleles in individual animals ranged between one and four, with none of the animals having an identical complement of BoLA class I molecules. Variation existed in the number of BoLA class I alleles expressed as well as the composition of expressed alleles, however, several BoLA class I alleles were found in multiple individual animals. Polymorphic amino acid sites were analyzed for positive and negative selection using the ADAPTSITE program. In the antigen recognition sites (ARS), there were eight positions that were predicted to be under positive selection and three positions that were predicted to be under negative selection from 62 positions. In contrast, for non-antigen recognition sites (non-ARS), there were three positions that were predicted to be under positive selection and 20 that were predicted to be under negative selection from 278, indicating that positive selection of amino acids occurs at a greater frequency within the antigen recognition sites.

Nucleotide sequence data are available in GenBank databases under accession numbers DQ121139, DQ121144, DQ121149, DQ121150, DQ121161, QD121165, DQ121168, DQ121170, DQ121175, DQ121176, DQ121177, DQ121180, DQ121183, DQ121184, DQ121190, DQ121191, DQ121193, DQ121196, DQ121204, and DQ121206.

S. Babiuk · B. Horseman · R. Pontarollo  
Pyxis Genomics Canada,  
Saskatoon, Saskatchewan, Canada

*Present address:*

S. Babiuk (✉)  
CFIA/NCFAD,  
1015 Arlington Street, Suite T2300,  
Winnipeg, Manitoba, Canada R3E 3M4  
e-mail: babiuks@inspection.gc.ca

C. Zhang · A. Kusalik  
Department of Computer Science,  
University of Saskatchewan,  
Saskatoon, Saskatchewan, Canada

M. Bickis  
Department of Mathematics and Statistics,  
University of Saskatchewan,  
Saskatoon, Saskatchewan, Canada

L. B. Schook · M. S. Abrahamsen  
Pyxis Genomics Inc.,  
Chicago, IL, USA

**Keywords** Cattle · MHC class I · BoLA class I · Polymorphism · Allele

## Introduction

Major histocompatibility complex (MHC) genes encode surface glycoproteins that bind peptides derived from both the host and associated pathogens. MHC molecules present

these peptides to T cells through T cell receptors. There are two types of classical MHC molecules, MHC class I and MHC class II. MHC class I molecules present endogenous peptides between 8–10 amino acids in length to CD8+ cytotoxic T cells, whereas MHC II molecules present exogenous peptides between 12–15 amino acids in length to CD4+ helper T cells. The differences in antigen presentation by MHC class I and MHC class II molecules allow the immune system to differentiate intracellular and extracellular pathogens. In humans, there are three distinct genomic loci that encode classical MHC class I molecules, HLA-A, HLA-B, and HLA-C, allowing individuals to express between three and six functionally distinct MHC I molecules. Other animal species also express three different MHC class I loci, swine SLA-1 SLA-2, SLA-3 (Ando et al. 2003) and chimpanzees Patr-A, Patr-B, Patr-C (de Groot et al. 2000). In contrast, other animal species such as the rat, mouse, and rhesus macaque have less rigid MHC I expression, with less or more than three classical genes with variable numbers (Roos and Walter 2005). This evidence indicates that consistent expression of the same three genes may be unusual. The precise number of MHC class I loci are not known in cattle, although mapping studies have suggested that there are at least six loci (Ellis 2004). Complicating this issue in cattle is the suggestion that interlocus recombination occurs in cattle MHC I genes (Holmes et al. 2003). In addition, BoLA class I alleles have yet to be assigned to specific genomic loci, and the gene content of haplotypes (which alleles are found on the same chromosome) is variable (Ellis et al. 1999). Cattle MHC haplotypes are well conserved, for example, an A11 haplotype in one animal will almost invariably consist of the same two alleles in an unrelated animal, which allows serological typing to be reasonably effective in cattle (Ellis et al. 1999).

The ability to express several different MHC molecules by an individual allows an increased diversity of peptides displayed for antigen presentation. Increased diversity of antigen presentation has been demonstrated to be an advantage to the host against infectious disease. This is most clearly demonstrated in HIV where heterozygous individuals delayed acquiring AIDS, whereas homozygous individuals progress rapidly to AIDS (Carrington et al. 1999). MHC class I genes are among the most highly polymorphic genes found in higher vertebrates (Parham et al. 1988). It is thought that this high degree of polymorphism is a result of positive selection from infectious agents of protective MHC alleles. The diversity of MHC molecules within a population allows for the survival of the species against infectious disease, as disease susceptibility is influenced by the diversity and complement of MHC molecules.

Because MHC restriction plays a critical role in the generation of specific T cell responses, knowledge of the number of

expressed MHC class I genes, as well as the allele complement in different animal species, is critical. This information is particularly important in vaccine development for the identification of potential candidate peptide epitopes that can be used in vaccine components or as assay reagents for quantifying immune responses to vaccines. In addition, MHC class I allele frequency information at the population level is useful for determining the effectiveness of epitope-based vaccines within the population, as a restricted combination of class I alleles can be found in the majority of the population. HLA supertypes consist of different HLA alleles that are capable of binding identical peptide epitopes. For example, in humans, the A2 supertype includes A\*0201, A\*202, A\*203, A\*204, A\*205, A\*206, A\*207, A\*3802, and A\*6901 (del Guercio et al. 1995). In humans, nine HLA supertypes cover 90% of the population (Sette and Sidney 1999).

MHC class I polymorphisms were first determined using serology. However, serology is not as sensitive as molecular approaches such as DNA sequencing in detecting and analyzing MHC class I polymorphisms. Indeed, DNA sequence data have been used to identify additional polymorphisms in cattle from serologically identical samples (Ellis et al. 1998). Previous BoLA class I sequence data indicate that multiple BoLA class I alleles exist and that there is variation in the number of BoLA class I genes expressed in different animals (Ellis et al. 1999). However, there is only limited sequence information available for BoLA class I molecules, with only 58 validated full-length sequences (<http://www.ebi.ac.uk/ipd/mhc/bola/>) and several partial sequences available in GenBank (Bensaid et al. 1991; Brown et al. 1989; Ellis et al. 1992, 1996; Ennis et al. 1988; Garber et al. 1994; Pichowski et al. 1996; Russell et al. 1996; Ellis 2004). There is limited information regarding allele composition in individual animals and allele frequencies at the population level. Although, there have been individual cattle that have been typed both serologically and molecularly for MHC I (Ellis et al. 2005). To address these shortcomings, full-length BoLA class I alleles from 36 individual animals within a single herd were analyzed to determine individual allele expression and population frequencies. In addition, polymorphic amino acid positions were examined for positive and negative selection using the ADAPTSITE program (Suzuki and Gojobori 1999; Suzuki et al. 2001) using these BoLA alleles as well as the currently validated BoLA alleles.

## Materials and methods

### Animals

The 36 animals used in this study for determining BoLA class I allelic variation in individual animals as well as in

the herd were Charolais cross cattle from a farm outside Saskatoon, Canada. Animals were treated in compliance with the Canadian Council for Animal Care.

#### Cloning and sequencing of BoLA class I genes

Cattle blood was collected using Vacutainer tubes containing ethylenediaminetetraacetic acid (EDTA). Peripheral blood mononuclear cells (PBMCs) were isolated from whole bovine blood as follows. Blood was centrifuged at  $900\times g$  for 20 min. The buffy coat was removed and resuspended in phosphate buffered saline (PBS)–0.1 M EDTA. The cells were overlaid on isotonic 60% Percoll (Amersham Biosciences, Baie d'Urfe, QC) and centrifuged at  $1,400\times g$  for 30 min. The PBMCs were washed three times in PBS-EDTA.

Total RNA was isolated from the PBMCs using the following method. One milliliter of Trizol (Invitrogen, Burlington, ON) was placed on PBMCs in a loosened cell pellet. RNA was extracted from the Trizol by adding 200  $\mu$ l of chloroform, shaking and incubating for 3 min at room temperature, followed by centrifugation (12,000 for 10 min at 4°C). The aqueous phase was removed, and the RNA was precipitated with 500  $\mu$ l of isopropanol and incubated for 5 min at room temperature and resuspended in nuclease-free water. The RNA was further purified with an RNeasy MinElute column using the manufacturer's instructions (Qiagen Mississauga, ON) with DNA removal from the RNA using DNase 1 (Qiagen) treatment for 15 min at room temperature. RNA quality and quantity were analyzed using a 2100 Bio-analyzer (Agilent, Mississauga, ON), and there was no degradation of the RNA samples.

Total RNA was used to generate cDNA with Superscript III first strand cDNA synthesis for real-time polymerase chain reaction (RT-PCR) using the manufacturer's instructions (Invitrogen). Full-length cattle MHC class I amplicons were generated from cDNA by PCR using the primers Bo21 5' CATGGGGCCGCGAGC 3' and Bo3 5' GGATGAAGCATCACTCAG 3' (Ellis et al. 1999). PCR was done with Accuprime *Pfx* polymerase (Invitrogen) under the following PCR cycle conditions with [Mg 2.1 mM]: 95°C for 5 min followed by 30 PCR cycles, conditions of 95°C for 15 s, 51.4°C for 30 s, 68°C for 90 s, with a final extension reaction at 68°C for 7 min. Amplified DNA was run on a 1% agarose gel, and the DNA band was excised and purified using a gel purification kit (Qiagen). The PCR amplified DNA was then ligated into blunt-end TOPO cloning vectors following the manufacturer's instructions (Invitrogen), and competent *E. coli* cells were transformed with the ligation mixture and plated on LB kanamycin plates. Over 600 clones were sequenced, with multiple clones (up to 20) sequenced from each individual animal using both M13 forward and M13 reverse primers for each

clone. Potential sequences were generated from at least three clones. These potential sequences were confirmed by independent PCR by the discovery of the same sequence in two or more different animals.

#### Phylogenetic analysis of BoLA class I sequences

The BoLA class I protein sequences (amino acids 1–339) with the antigen recognition sites (ARS) removed (sites 5, 7, 9, 22, 24, 26, 58–60, 62–87, 95, 97, 99, 101, 114, 116, 146, 148–150, 152–156, 158–162, 164–166, 168–170, 172, and 174) were used to construct a phylogenetic tree. First, a multiple sequence alignment was performed using Clustal X (Thompson et al. 1997). Phylogenetic analysis was done using Phylip 3.5 c and based on the amino acid sequences from positions 1–339 with the ARS amino acid sites removed. Sequences were bootstrapped using 1,000 replicates, distance matrices computed using protein distance, unweighted pair group method with arithmetic mean clustering performed to obtain rooted trees, and a consensus tree determined. The results were displayed using TreeView (Page 1996).

#### Analysis of selection at single amino acid sites using ADAPTSITE

To detect natural selection for both negatively and positively selected amino acid sites, BoLA class I alleles starting at codon 1 were analyzed using ADAPTSITE (Suzuki and Gojobori 1999; Suzuki et al. 2001). Because ADAPTSITE requires aligned sequences that contain no gaps, gaps were removed from the gene sequences at position before the analysis was performed. This was done by excising the CACTGG DNA sequence corresponding to a TG insertion in the protein sequence and removing sequences that contained gaps in the consensus sequence. A phylogenetic tree was made from the 20 BoLA class I DNA sequences identified, as well as the validated BoLA class I DNA sequences obtained from the European Bioinformatics Institute (<http://www.ebi.ac.uk/ipd/mhc/bola/>) by the neighbor-joining method on the basis of Tamura–Nei distances (Tamura and Nei 1993). The human HLA-A, -B and -C validated sequences and swine SLA class I validated sequences obtained from the European Bioinformatics Institute were also assessed using the ADAPTSITE program.

Using the phylogenetic tree, the number of synonymous and non-synonymous substitutions and the average number of synonymous and non-synonymous sites for each codon site were determined using ADAPTSITE-P. The *p* value that a given site was positively or negatively selected was calculated using ADAPTSITE-T using a two-tailed test. Amino acid sites that potentially represent ARS (sites 5, 7,

9, 22, 24, 26, 58–60, 62–87, 95, 97, 99, 101, 114, 116, 146, 148–150, 152–156, 158–162, 164–166, 168–170, 172, and 174) were compared to non-ARS codon sites that were either positively or negatively selected, not significantly selected, not changed, or codon sites for which the statistical test could not be conducted (amino acids that have only one codon).

### Statistics

The statistical significance of the differences between ARS and non-ARS groups was confirmed using Prism GraphPad statistical software (GraphPad Software, San Diego, CA) using Fisher's exact test.

## Results

### Sequence analysis of transcribed BoLA class I genes in cattle

RT-PCR amplification was performed on PBMCs isolated from 36 individual animals with a single herd of Charolais cattle. Full-length BoLA class I cDNA sequences were amplified using the PCR primers Bo21 and Bo3 (Ellis et al. 1999). Figure 1 shows the level of polymorphism in the cattle BoLA class I genes. Twenty different BoLA class I alleles were identified from the sample population. A Blast search was performed on the putative BoLA class I gene sequences and corresponding BoLA class I protein sequences. The results from the Blast search revealed that there was significant sequence homology with other BoLA class I sequences. However, only 4 of the 20 DNA sequences (DQ121139, DQ121183, DQ121184, and DQ121191) shared 100% identity at the nucleotide level. These sequences have been previously described as pBoLA-1 for DQ121184 (Brown et al. 1989), as D18.2 for DQ121191 (Davies et al. 1997), 4221.1 for DQ121139 (Ellis et al. 1999), and as D18.1 for DQ121183 (Davies et al. 1997) and were all identified from the Holstein breed.

The level of polymorphism in the BoLA class I molecules was very high in certain regions of the BoLA class I molecule, primarily within the alpha 1 and alpha 2 regions (Parham et al. 1988). These regions contain specific amino acid positions that are involved in peptide binding and T cell receptor (TCR) recognition (Parham et al. 1988). Furthermore, the amino acid positions involved in peptide binding and TCR recognition display the highest level of polymorphism as illustrated in Fig. 1. These results are similar to studies with HLA and SLA class I molecules where the polymorphism occurs primarily at amino acid sites involved with antigen presentation.

**Fig. 1** Alignment of the predicted amino acid sequences of the 20 identified BoLA class I genes. A *dot* indicates a conserved amino acid codon, and a *dash* indicates a gap induced to aid alignment. The sequence alignment was generated using Clustal X

### Possible recombination events between BoLA class I molecules

There were two instances where there was possible recombination of BoLA class I molecules. Sequences DQ121149 and DQ121150 are identical, apart from a small region in the alpha 2 domain. In addition, sequences DQ121165 and DQ121170 are identical through the alpha 1 and part of the alpha 2 domain. Both of these instances are unusual if compared to the other full-length sequences and suggest recombination.

### BoLA class I typing of individual animals

The number of different BoLA class I alleles that were expressed in individual animals varied between 1–4 alleles (Table 1). The majority of animals displayed between 1–4 different BoLA class I alleles. The alleles identified in Table 1 show little evidence of haplotypes, illustrated by individual alleles occurring in different animals with different sets of accompanying alleles. Because most cattle MHC I haplotypes express two alleles, most heterozygous animals will generate at least four sequences. It is likely that it reflects the fact that probably only half the expressed alleles were detected in each case, as indicated by the low number of sequences from each animal. As only alleles that were obtained by at least two independent PCRs were included in the analysis, it is possible that additional alleles were present in individual animals but not identified and used in the analyses. This is very likely for animal 28 where no alleles were identified. Analysis of the allele complement of each of the 36 animals revealed that no two animals expressed the identical complement of BoLA class I alleles, indicating that this group of animals did not represent a high frequency of a limited number of alleles.

### BoLA class I frequency in the cattle herd

The frequency of the 20 different BoLA class I alleles was determined for the sample population. Eight alleles were found in two animals (DQ121165, DQ121170, DQ121175, DQ121180, DQ121184, DQ121190, DQ121196, and DQ121204), one allele (DQ121206) in three animals, four alleles (DQ121177, DQ121168, DQ121191, and DQ121161) in four animals, two alleles (DQ121139 and DQ121183) in five animals, two alleles (DQ121176 and DQ121149) in six animals, one allele (DQ121193) in eight animals, one allele (DQ121150) in nine animals, and

Alpha 1

	1	50
DQ121191	MGPRTL L L L L L S G V L V L T E T R A G S H M R Y F S T A V S R P G L G E P R Y L E V G Y V D D T Q F V R F D S D A R N P R M E P R K R W V E Q E G P E Y W D R E T Q R A K G	
DQ121149	..... F V . L A A . . . . . L . Y . . . . . F I S . . . . . P . E . . . A P . I . K . . . . . R I S . E	
DQ121150	..... F V . L A A . . . . . L . Y . . . . . F I S . . . . . P . . . E . . . A P . I . K . . . . . R I S . E	
DQ121144	..... F V . L A A . . . . . L K . H . . . . . D . F I T . . . . . P D . K . . . A P . I . K . . . . . D . R I S . E	
DQ121139	..... F V . L A . . . . . L K . H . . . . . D . F I T . . . . . P D . K . . . A P . I . K . . . . . R I S . E	
DQ121204	..... L . . . . . L . Y . G . . . . . F I A . . . . . T . . . . . P . . . E . . . V P . M . . . . . R I S . D	
DQ121180	..... L . . . . . L . Y . . . . . F I S . . . . . S A . . . E . . . A P . M . . . . . E Q . R I V . D	
DQ121206	..... V . . . . . L . . . . . F I A . . . . . T . . . . . P . . . E . . . A P . M . . . . . L . . . . . N . R I Y . D	
DQ121161	..... L . . . . . L . . . . . L . Y . . . . . F I A . . . . . R P D . . . . . A . . . . . E . R K . D	
DQ121177	..... L . . . . . L . . . . . Y . G . . . . . F I A . . . . . P D . T . . . V . I . . . . . R N L . D	
DQ121168	..... F . . . . . F . . . . . P . . . . . V . M . . . . . Q . . . . .	
DQ121170	..... F . . . . . F . . . . . P . . . . . A . . . . .	
DQ121165	..... F . . . . . F . . . . . P . . . . . A . . . . .	
DQ121184	..... F . . . . . F . . . . . P . . . . . A . . . . . Q . R K . . .	
DQ121190	..... F . . . . . F . . . . . P . . . . . T . . . K . . . . . N . R N . . .	
DQ121175	..... L . . . . . F . . . . . Q . . . . . P . . . . . A . . . . . N . R N . . .	
DQ121176	..... F . . . . . F . . . . . T . . . K . . . . . N . R N . . .	
DQ121193	..... F . . . . . F . . . . . P . . . . . A . . . . .	
DQ121196	..... L . Y . . . . . F I S . . . . . P . . . . . A . . . . . R N S . D	
DQ121183	..... L . . . . . E . . . . . F I I . . . . . S P . . . A . . . A P . M . . . . . E Q . R I V . D	

Alpha 2

	1	100	150
DQ121191	NAQIFRVSLNNLRGYNQSEAGSHTFQWMYGCDVGPDRLLGGYEQYGYDGRDYIALNEDLRSWTAGETEAIQTKRKWEAAGAAERQRNY		
DQ121149	.TLVY.E.....L.L.....R.R.DA.S.....E.....AD.A.....E.Y.SL..		
DQ121150	.TLVY.E.....NI.A.....S.SF.R.S.DA.....AD.A.....E.Y.SL..		
DQ121144	.TLLY.KD..T.....NI.A.....F.R.R.DA.....AD.A.....E.Y.SL..		
DQ121139	.TLVY.G.....Y.D.....D.....L.....AD.A.....V.F..		
DQ121204	T.S...G..T.....I.E.....F.R.....E.....AD.A.....DY..SL..		
DQ121180	T.S...G..T.....L.L.....F.R.FM.DA.....AD.A.....R..		
DQ121206	T.T...Y..T.....L.S.....R.FM.F.....Q.....AD.A.....DN..SE..		
DQ121161	T.T...D..T.....NI.A.....F.R.R.DA.....AD.A.....V.F..		
DQ121177	T.T...D..T.....I.E.....F.R.R.DA.....L.....AD.A.....E.F..		
DQ121168	.T...D..T.....L.Y.....R.FM.DA.....L.....AD.A.....E..		
DQ121170	.T.N.T.....I.Y.....F.R.R.DA.....AD.A.....E..		
DQ121165	.T.N.T.....I.Y.....F.R.R.DA.....AD.A.....V.F..		
DQ121184	T.T.AN.IAL.....RR.FM.....AD.A.....E..		
DQ121190	.T.G.....L.....RR.FT.....E.....AD.A.....E.SW..		
DQ121175	.S.N.T.....L.S.....RR.FM.....L.....AD.A.....Y.V..		
DQ121176	.S.G.T.....L.S.Y.....R.FM.....L.....AD.....E.Y.V..		
DQ121193	.T.N.T.....I.Y.....F.R.R.DA.....D.A.....E..GE..		
DQ121196	T.G...D..T.....L.L.Y.....S.RR.FT.....L.....V.M.S.V..V.K..		
DQ121183	T.T.AN.TAL.....NI.A.....S.SF.R.S.DA.....AD.A.....E.Y.SL..		

Alpha 3

	1	200
DQ121191	LEGRCVWELRRYLENGKDTLLRADPPKAHVTHHPISEREVTLRCWALGFYPEEISLTWQRNGEDQTDQDMELVETRPSGDGNFQKWAALVV	
DQ121149	.....A.....M.....S.....K.....E.....T.....	
DQ121150	.....A.....M.....S.....K.....E.....T.....	
DQ121144	.....G.....A.....S.....K.....S.....	
DQ121139	V.....G.....A.....K.....	
DQ121204	..E..G.....D.....K.....L.....	
DQ121180	.....G.....A.....E.....T.....	
DQ121206	..E.....A.....D.....T.....	
DQ121161	..L..G.....D.....E.....M.....T.....	
DQ121177	V..L..G.....H.....E.....T.....G..	
DQ121168	..T.....V.D.....E.....T.....	
DQ121170	..T.....V.D.....E.....T.....	
DQ121165	..L..G.....D.....E.....M.....T.....	
DQ121184	..T.....T.....S.GH.....D.....	
DQ121190	..T.....H.....DY.....M.....	
DQ121175	..E.....D.....HD.....V.....	
DQ121176	..E.....H.....D.....H.....T.....	
DQ121193	..T.....A.....G.....D.....V.....	
DQ121196	.....R.....G.....L.....	
DQ121183	..T.....S.G.....K.....E.....	



	Transmembrane domain				Cytoplasmic domain			
	250				300			
DQ121191	PSGEEQKYTCRVQHEGLQEPLTLKWEPPQPSFLT	MGIIVGLVLLVVTGAVVAGVVICMKKRS	GEKRGTYIQASS	SSDSAQGS	SDVSL	TPVKV		
DQ121149	.....R...H.....I.R.....T...I.....	.....A..WR.....Q..T..A.G..D.....						
DQ121150	.....R...H.....I.R.....T...I.....	.....A..WR.....Q..T..A.G..D.....						
DQ121144	.....R...H.....R.....T.....	.....A..WR.....Q..T..A.G..D.....						
DQ121139	.....R...H.....R.....T...I.....	.....A..WR.....Q..T..A.G..D.....						
DQ121204	.....R...H.....R.....T.....	.....A..WR.....Q..T..A.G..D.....						
DQ121180	.....R.....R.....T...I.....	.....L..A..WR.....GRI.T..A.....						
DQ121206	.....R.....R.....T...I.....	.....L..A..WR.....GRI.T..A.....T.....						
DQ121161	.....R.....R.....T.....	.....A..WR.....GRI.T..A.....						
DQ121177	.....R.....R.....T.....	.....A..WR.....GRI.T..A.N.....						
DQ121168	.....R.....R.....T...I.....	.....L..A..WR.....GRI.T..A.....						
DQ121170	.....R.....R.....T...I.....	.....L..A..WR.....GRI.T..A.....						
DQ121165	.....R.....R.....T.....	.....A..WR.....GRI.T..A.....						
DQ121184								
DQ121190								
DQ121175								
DQ121176							P.....L.....	
DQ121193								
DQ121196	.....Q.....							
DQ121183	.....K.....R.....T.....	.....V...A..W..R.....G.N.....						

**Fig. 1** (continued)

one allele (DQ121144) was found in ten animals. The five most frequent alleles occurred in 30/36 animals, indicating the possibility that these sequences might be BoLA super-types (Sette and Sidney 1999) or may have occurred due to the similar genetic backgrounds of the animals.

#### Phylogenetic analysis of BoLA class I alleles

The 20 BoLA class I protein coding sequences (amino acids 1–339) with the ARS sites removed were used to create a phylogenetic tree (Fig. 2). The phylogenetic analysis revealed that the BoLA class I sequences clustered into three groups. Three distinct groups were observed most notably in the transmembrane and cytoplasmic regions (Figs. 1 and 2). It is tempting to speculate that each group represents the alleles of a locus. However, this hypothesis is suspect, as the allele distribution found in some animals (8 and 18) contained three alleles in the same phylogenetic grouping. Therefore, it is unlikely that these groups correspond to discrete loci.

#### Detection of natural selection at single amino acid sites

To detect natural selection at individual amino acid sites on BoLA class I alleles, ADAPTSITE program was utilized on the 20 BoLA class I alleles that were identified from the cattle herd together with the validated BoLA alleles from the European Bioinformatics Institute. Amino acid sites were characterized as positively or negatively selected if they reached the 5% significance level. The BoLA class I amino acid positions were separated into sites potentially involved in antigen recognition (ARS) and sites that are not

involved in antigen recognition (non-ARS). The ARS codons chosen were similar to the ARS codons used by Ando et al. (2003) in a previous study and described in “Materials and methods”. In the BoLA class I alleles, positive selection was detected at 11 amino acid sites. Comparing the values for ARS codons (8 out of 62 positions) versus non-ARS codons (3 out of 278 positions) revealed that amino acid sites undergo significantly more positive selection than non-ARS sites (see Table 2). This result is similar to humans HLA-A, and -B where positively selected amino acids were found primarily in the ARS with 5 and 12 sites out of 62, respectively. However, for HLA-C, there were no positively selected amino acid sites detected at any sites. When the combined swine SLA-1, -2 and -3 class I sequences were analyzed using ADAPTSITE, it was demonstrated that amino acid sites in the ARS undergo more positive selection (7 out of 62) than amino acid sites in non-ARS (2 out of 281). These results are similar to the results determined in swine previously (Ando et al. 2003). Because amino acid changes in the ARS can affect peptide binding and TCR recognition, it would be expected that amino acids in these sites would undergo more positive selection than amino acid sites that are not involved in peptide binding and TCR recognition. In contrast, amino acid changes in the non-ARS region undergo less selection compared to the ARS region and, where selection occurred, it was predominantly negative selection. This was also observed with human and swine class I molecules.

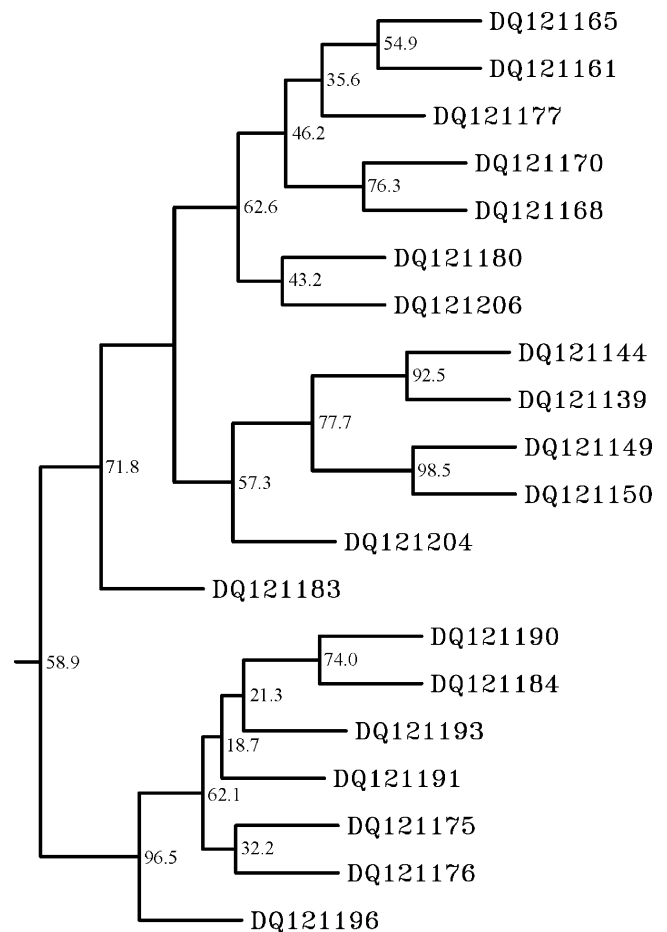
In cattle, 8 of the 11 positions that undergo positive selection were found in the ARS at positions 9, 67, 73, 77, 80, 97, 114, and 152. In humans, combining the results from HLA-A and -B class I molecules, positive selection

**Table 1** BoLA class I typing of individual animals

Animal	Total alleles	Alleles
2	2	DQ121168, DQ121193
3	4	DQ121168, DQ121170, DQ121175, DQ121193
4	2	DQ121139, DQ121176
5	4	DQ121161, DQ121180, DQ121184, DQ121193
6	1	DQ121176
7	2	DQ121139, DQ121149
8	3	DQ121139, DQ121144, DQ121150
9	2	DQ121150, DQ121191
10	1	DQ121177
11	2	DQ121149, DQ121183
12	3	DQ121149, DQ121177, DQ121193
13	3	DQ121168, DQ121170, DQ121193
14	3	DQ121150, DQ121177, DQ121196
15	2	DQ121150, DQ121175
16	1	DQ121150
17	3	DQ121161, DQ121165, DQ121193
18	4	DQ121176, DQ121191, DQ121193, DQ121204
19	2	DQ121144, DQ121149
20	3	DQ121144, DQ121180, DQ121184
21	1	DQ121150
22	3	DQ121176, DQ121204, DQ121206
23	3	DQ121150, DQ121161, DQ121193
24	2	DQ121190, DQ121206
25	1	DQ121149
26	3	DQ121165, DQ121190, DQ121193
27	2	DQ121144, DQ121150
28	0	
29	2	DQ121150, DQ121144
30	2	DQ121144, DQ121149
31	2	DQ121139, DQ121144
32	2	DQ121144, DQ121196
33	1	DQ121144
34	2	DQ121139, DQ121176
35	2	DQ121176, DQ121206
36	1	DQ121191

BoLA class I alleles that were found in each individual animal as well as the number of different alleles found in each animal are displayed in the table.

was previously detected in the ARS at 15 amino acid positions: 9, 62, 63, 67, 69, 80, 81, 82, 83, 95, 97, 114, 116, 152, and 156. In porcine, SLA class I molecules positive selection was detected in the ARS at seven amino acid positions: 9, 67, 74, 77, 95, 152, and 156 (Ando et al. 2003). A comparison of amino acid sites in the MHC class I molecules that undergo positive selection between cattle, human, and swine illustrates that there is overlap in the sites that undergo positive selection among the three species (Table 3).



**Fig. 2** Phylogenetic tree of BoLA class I amino acid sequences with the ARS sites removed. The BoLA class I protein sequences 1–339, with the ARS sites removed, were used to construct the phylogenetic tree. *Forks* of the tree are labeled with values, which are the percentage of times the group consisting of the species to the *right* of that *fork* occurred among the trees out of 1,000 bootstrap trees. All branch lengths are drawn to *scale*

## Discussion

There have been no studies that have characterized the diversity of BoLA class I molecules at the population level, however, these studies have been done with human HLA class I molecules (Ivanova et al. 2001). These population studies reveal that the high genetic diversity of the HLA class I molecules within the population and differences in the frequency of different HLA class I alleles vary between different ethnic groups (Mori et al. 1997). The limitations of this study were: (1) the possibility that the primers used might not amplify all the alleles; (2) the maximum number of 20 clones per animal was not very many; and (3) some alleles are only represented at a low level.

In human HLA class I molecules, there are approximately 1,000 alleles at the three class I loci. It is hypothesized that the level of diversity and polymorphism

**Table 2** Identification of amino acid sites that undergo positive and negative selection

	Positive selection	Negative selection	Excluded <sup>a</sup>	No proven selection <sup>b</sup>	No substitution <sup>c</sup>	Number of sites
Cattle						
ARS	8	3	1	39	11	62
Non-ARS	3	20	12	144	97	276
Total	11	23	13	183	108	338
Swine						
ARS	7	4	1	36	14	62
Non-ARS	2	14	10	93	160	279
Total	9	18	11	129	174	341
HLA-A						
ARS	5	1	3	25	28	62
Non-ARS	0	3	16	77	184	280
Total	5	4	19	102	212	342
HLA-B						
ARS	12	2	4	20	24	62
Non-ARS	2	15	9	60	191	277
Total	14	17	13	80	215	339
HLA-C						
ARS	0	0	2	17	43	62
Non-ARS	0	4	9	54	213	280
Total	0	4	11	71	256	342

Numbers of codon sites where positive, and negative selection was determined for ARS and non-ARS in BoLA class I molecules using ADAPTSITE. A 2×2 contingency table was constructed for positive and negative selection versus ARS and non-ARS. Differences between positive and negative selection for ARS and non-ARS were also analyzed using Fisher's exact test ( $p < 0.01$ ).

<sup>a</sup>Contains sites coding an amino acid for which there is only a single codon (Try, Met)

<sup>b</sup>Contains amino acid sites where no significant difference was detected between the numbers of synonymous and non-synonymous changes

<sup>c</sup>Contains amino acid sites where there was no substitution found

in cattle BoLA class I is likely high. Polymorphism was found in this cattle herd with 20 different BoLA class I sequences identified in 36 animals. From 20 different BoLA class I alleles, only four BoLA class I alleles shared 100% identity at the nucleotide level to any known BoLA class I sequences. It is interesting though that the four BoLA class I alleles that were previously identified have been among the first BoLA class I alleles reported and have been verified by different groups. The four previously identified BoLA class I alleles were all reported from Holstein cattle. Our study used Charolais cross, indicating that there are some shared BoLA class I alleles between Holstein and Charolais cattle. As only a small number of BoLA class I alleles identified in this study were previously identified, our results indicate that polymorphism is likely high in cattle BoLA class I alleles.

The pattern of polymorphism in cattle BoLA class I genes determined using ADAPTSITE showed a similar pattern of positive selection in the ARS as with human and porcine MHC I sequences. Analysis of human HLA-A, -B, and -C class I alleles by ADAPTSITE displayed significant differences between amino acid sites that were positively selected. One possible reason for these differences is that these loci experience different selection pressures leading to differences in amino acid sites that are positively selected. With cattle, it is currently unknown if there are different selection pressures on different BoLA class I loci. If BoLA class I alleles undergo the same selection pressures, it may explain the difficulties in assigning alleles to loci. The differences seen in amino acid sites that undergo positive selection between humans, swine, and cattle likely reflect

different selection pressures of MHC class I molecules caused by the different pathogens that infect each species.

The ADAPTSITE method has been shown to have a very low rate of false positives but little power for identifying positive selection (Wong et al. 2004). Despite these limitations of ADAPTSITE, it was still able to identify several positively selected sites in the limited

**Table 3** Overlap of positively selected amino acid sites in the ARS between cattle, humans, and swine

Species	Amino acid sites where positive selection was detected in the ARS					
Cattle	9	67	73	77	80	
Swine	9	67		74	77	
HLA-A	9	62			76	
HLA-B		63	67	69	74	
HLA-C						
Species	Amino acid sites where positive selection was detected in the ARS					
Cattle				97	114	152
Swine				95		152
HLA-A	81					152
HLA-B	81	82	83	95	97	116
HLA-C						156

Amino acid sites that were identified as sites that were positively selected using ADAPTSITE from cattle, human, and swine MHC I molecules were compared. Cattle versus human had 4 out of 19 possible sites identical, cattle versus swine had 4 out of 11 possible sites identical, and human versus swine had 6 out of 16 possible sites identical. Among the three species, there were three identical sites out of 19 total sites where positive selection was determined in the three species.



number of cattle BoLA class I alleles. Because BoLA class I alleles have not been assigned to loci, the ADAPTSITE program was run on the entire set of BoLA class I alleles without grouping the alleles. Running the data as pooled loci did not prevent the detection of amino acid sites that undergo selection.

Phylogenetic analysis on the amino acid sequences of the BoLA class I genes alpha 3 through cytoplasmic domains revealed three distinct groups, with notable groups of sequences in the cytoplasmic domain. However, recombination events over time likely cause mixing of these cytoplasmic domains, potentially limiting their usefulness in assigning alleles to loci. More studies generating haplotype data for cattle as well as gene mapping studies (Di Palma et al. 2002) to assign sequences to specific locations on chromosomes including sequencing of BoLA class I introns will be required to assign alleles to loci.

Between one and four, different BoLA class I alleles were expressed in individual animals at an unknown number of loci. Despite not being able to assign BoLA class I molecules to loci, it is clear that cattle can express several BoLA class I alleles similar to other animals including humans, although the numbers are less rigid with cattle. Further studies of the frequency of BoLA class I genes expressed at the population level will be useful to define BoLA class I supertypes, if they exist (Sette and Sidney 1999). Further sequencing of BoLA class I molecules from different cattle breeds should allow for the identification of BoLA class I supertypes. The identification of BoLA class I supertypes would be extremely useful for the development of epitope based vaccines for cattle. Serological data have suggested that there is a different repertoire of alleles expressed in different cattle breeds (Mackie et al. 1989; Stear et al. 1987). Understanding the frequency of different BoLA class I alleles in different cattle breeds will allow for improved design of vaccines.

This study illustrated that there are both similarities and differences between BoLA class I and class I MHC alleles from other species. The polymorphism data generated for BoLA class I alleles could reflect high polymorphism in one or more genes and little or none in another. Further studies will be needed to determine the level of polymorphism in BoLA class I genes. The results from the detection of natural selection at amino acid sites indicate that BoLA class I molecules undergo similar selective pressure as human HLA class I and swine SLA class I molecules, with positive selection being significantly higher in the ARS compared to non-ARS, indicating similarities in evolutionary selection pressure on amino acid sites where positive selection occurs predominantly in the ARS sites in cattle, swine, and humans, although there are subtle differences in the particular amino acid sites due to different selection pressures. The major differences between BoLA class I

alleles and MHC class I alleles from other species include the lack of clear loci and the variant class I haplotype composition in cattle. In this small sample of cattle, 16 out of 20 BoLA class I alleles identified were novel.

**Acknowledgment** We thank the animal care staff at the Vaccine and Infectious Disease Organization (University of Saskatchewan) for collecting the cattle blood samples. We also thank Juan Abrahante (Department of Veterinary and Biomedical Science, College of Veterinary Medicine, University of Minnesota) for the DNA sequencing preparation. Funding was provided by Pyxis Genomics Canada. Shawn Babiuk was a recipient of a Natural Science and Engineering Research Council of Canada (NSERC) industrial research fellowship with Pyxis Genomics Canada. Ben Horseman was a recipient of an NSERC undergraduate student internship with Pyxis Genomics Canada.

## References

- Ando A, Kawata H, Shigenari A, Anzai T, Ota M, Katsuyama Y, Sada M, Goto R, Takeshima SN, Aida Y, Iwanaga T, Fujimura N, Suzuki Y, Gojobori T, Inoko H (2003) Genetic polymorphism of the swine major histocompatibility complex (SLA) class I genes, SLA-1, -2 and -3. *Immunogenetics* 55:583–593
- Bensaid A, Kaushal A, Baldwin CL, Clevers H, Young JR, Kemp SJ, MacHugh ND, Toye PG, Teale AJ (1991) Identification of expressed bovine class I MHC genes at two loci and demonstration of physical linkage. *Immunogenetics* 33:247–254
- Brown P, Spooner RL, Clark AJ (1989) Cloning and characterization of a BoLA class I cDNA clone. *Immunogenetics* 29:58–60
- Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ (1999) HLA and HIV-1: heterozygote advantage and B\*35-Cw\*04 disadvantage. *Science* 283:1748–1752
- Davies CJ, Andersson L, Ellis SA, Hensen EJ, Lewin HA, Mikko S, MuggliCockett NE, vanderPoel JJ, Russell GC (1997) Nomenclature for factors of the BoLA system, 1996: report of the ISAG BoLA Nomenclature Committee. *Anim Genet* 28:159–168
- de Groot NG, Otting N, Arguello R, Watkins DI, Doxiadis GG, Madrigal JA, Bontrop RE (2000) Major histocompatibility complex class I diversity in a West African chimpanzee population: implications for HIV research. *Immunogenetics* 51:398–409
- del Guercio MF, Sidney J, Hermanson G, Perez C, Grey HM, Kubo RT, Sette A (1995) Binding of a peptide antigen to multiple HLA alleles allows definition of an A2-like supertype. *J Immunol* 154:685–693
- Di Palma F, Archibald SD, Young JR, Ellis SA (2002) A BAC contig of approximately 400 kb contains the classical class I major histocompatibility complex (MHC) genes of cattle. *Eur J Immunogenet* 29:65–68
- Ellis S (2004) The cattle major histocompatibility complex: is it unique? *Vet Immunol Immunopathol* 102:1–8
- Ellis SA, Braem KA, Morrison WI (1992) Transmembrane and cytoplasmic domain sequences demonstrate at least two expressed bovine MHC class I loci. *Immunogenetics* 37:49–56
- Ellis SA, Staines KA, Morrison WI (1996) cDNA sequence of cattle MHC class I genes transcribed in serologically defined haplotypes A18 and A31. *Immunogenetics* 43:156–159

- Ellis SA, Staines KA, Stear MJ, Hensen EJ, Morrison WI (1998) DNA typing for BoLA class I using sequence-specific primers (PCR-SSP). *Eur J Immunogenet* 25:365–370
- Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, Morrison WI (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. *Immunogenetics* 50:319–328
- Ellis SA, Morrison WI, MacHugh ND, Birch J, Burrells A, Stear MJ (2005) Serological and molecular diversity in the cattle MHC class I region. *Immunogenetics* 57:60–66
- Ennis PD, Jackson AP, Parham P (1988) Molecular cloning of bovine class I MHC cDNA. *J Immunol* 141:642–651
- Garber TL, Hughes AL, Watkins DI, Templeton JW (1994) Evidence for at least three transcribed BoLA class I loci. *Immunogenetics* 39:257–265
- Holmes EC, Roberts AF, Staines KA, Ellis SA (2003) Evolution of major histocompatibility complex class I genes in Cetartiodactyls. *Immunogenetics* 55:193–202
- Ivanova M, Spassova P, Michailova A, Naumova E (2001) Distributions of HLA class I alleles and haplotypes in Bulgarians—contribution to understanding the origin of the population. *Tissue Antigens* 57:208–215
- Mackie JT, Bath ML, Stear MJ, Davis WC (1989) Monoclonal antibodies to bovine major histocompatibility system antigens. *Exp Clin Immunogenet* 6:179–184
- Mori M, Beatty PG, Graves M, Boucher KM, Milford EL (1997) HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry. *Transplantation* 64:1017–1027
- Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12:357–358
- Parham P, Lomen CE, Lawlor DA, Ways JP, Holmes N, Coppin HL, Salter RD, Wan AM, Ennis PD (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc Natl Acad Sci USA* 85:4005–4009
- Pichowski JS, Ellis SA, Morrison WI (1996) Sequence of two cattle MHC class I cDNAs associated with BoLA A10 specificity. *Immunogenetics* 43:253–254
- Roos C, Walter L (2005) Considerable haplotypic diversity in the RT1-CE class I gene region of the rat major histocompatibility complex. *Immunogenetics* 56:773–777
- Russell GC, Oliver RA, Sawhney SM (1996) Cloning, transfection, and DNA sequence of a second gene from the BoLA-A11 haplotype. *Immunogenetics* 44:315–318
- Sette A, Sidney J (1999) Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50:201–212
- Stear MJ, Brown SC, Dimmock CK, Dufty JH, Hetzel DJ, Mackie JT, Nicholas FW, Tierney TJ, Wetherall JD (1987) Breed differences in the frequency of bovine lymphocyte antigens. *Exp Clin Immunogenet* 4:27–36
- Suzuki Y, Gojobori T (1999) A method for detecting positive selection at single amino acid sites. *Mol Biol Evol* 16:1315–1328
- Suzuki Y, Gojobori T, Nei M (2001) ADAPTSITE: detecting natural selection at single amino acid sites. *Bioinformatics* 17:660–661
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Wong WS, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041–1051