Published in final edited form as:

Immunogenetics. 2010 March ; 62(3): 159–172. doi:10.1007/s00251-009-0420-9.

Analysis of MHC class I genes across horse MHC haplotypes

Rebecca L. Tallmadge, Julie A. Campbell, Donald C. Miller, and Douglas F. Antczak

Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

Abstract

The genomic sequences of 15 horse Major Histocompatibility Complex (MHC) class I genes and a collection of MHC class I homozygous horses of five different haplotypes were used to investigate the genomic structure and polymorphism of the equine MHC. A combination of conserved and locus-specific primers was used to amplify horse MHC class I genes with classical and non-classical characteristics. Multiple clones from each haplotype identified three to five classical sequences per homozygous animal, and two to three non-classical sequences. Phylogenetic analysis was applied to these sequences and groups were identified which appear to be allelic series, but some sequences were left ungrouped. Sequences determined from MHC class I heterozygous horses and previously described MHC class I sequences were then added, representing a total of ten horse MHC haplotypes. These results were consistent with those obtained from the MHC homozygous horses alone, and 30 classical sequences were assigned to four previously confirmed loci and three new provisional loci. The non-classical genes had few alleles and the classical genes had higher levels of allelic polymorphism. Alleles for two classical loci with the expected pattern of polymorphism were found in the majority of haplotypes tested, but alleles at two other commonly detected loci had more variation outside of the hypervariable region than within. Our data indicate that the equine Major Histocompatibility Complex is characterized by variation in the complement of class I genes expressed in different haplotypes in addition to the expected allelic polymorphism within loci.

Keywords

horse; MHC class I gene; polymorphism; haplotype structure

INTRODUCTION

Major Histocompatibility Complex (MHC) class I molecules are membrane-bound glycoproteins that complex with small peptides and β -2-microglobulin. Classical MHC class I molecules provide immune surveillance against intracellular pathogens by presenting foreign peptides derived from intracellular protein degradation. Recognition of foreign peptides by the T cell receptor (TCR) on CD8⁺ T lymphocytes can initiate a specific immune response (Germain and Margulies 1993). MHC class I molecules can also inhibit Natural Killer (NK)-mediated cytolytic activity by interacting with receptors on NK cells, thereby playing a regulatory role in innate immunity (Lanier et al. 1997).

MHC class I genes encode a family of related molecules that consist of eight exons spanning approximately 4,000 bases and this gene structure is conserved in the horse (Carpenter et al.

Corresponding author: Douglas F. Antczak, Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, dfa1@cornell.edu, phone: (607) 256-5633, fax: (607) 256-5608.

Nucleotide sequence data reported are available in the GenBank database under the accession numbers DQ145583 – DQ145618, EF577062 – EF577072

2001; Moore et al. 1982; Steinmetz et al. 1981). Human classical MHC class I genes have many alleles, ranging from 463 to 1,249, whereas the non-classical genes have many fewer alleles, ranging from 9 to 44 (as of August 2009, <http://www.ebi.ac.uk/imgt/hla/intro.html>). This is an extraordinary level of diversity, and consequently *HLA-B* is the most polymorphic gene in the human genome (Mungall et al. 2003). The polymorphic nucleotides are often non-synonymous substitutions and are clustered into two main areas which comprise the peptide-binding groove (Bjorkman and Parham 1990). In addition to limited polymorphism, non-classical loci in humans can be further distinguished from classical loci by shorter transmembrane/cytoplasmic domains, lower surface expression, and limited tissue distribution (reviewed in Le Bouteiller 1994; Stroynowski 1990).

Analysis of human MHC class I gene sequences revealed that allelic variation is composed of point mutations and shared sequence motifs that arose from gene conversion events both within and between MHC class I loci during evolution (Lawlor et al. 1990; Parham et al. 1988). This mosaic of sequence motifs complicates the task of defining the origin of a given allele. However, analysis of human MHC class I alleles has uncovered locus-specific motifs, in the 5' upstream/promoter region (Cereb and Yang 1994), exons 1 through 8 (Parham et al. 1989), introns 1, 3 and 4 (Cereb et al. 1995; Swelsen et al. 2003) and the 3'UTR (Koller et al. 1984).

The number of expressed MHC class I loci has been shown to vary within a species, although thorough studies have not been undertaken in many species. Examples include (1) humans, who express three classical loci, *HLA-A*, *-B*, and *-C*, and two, *HLA-E* and *-F*, or three with *HLA-G*, non-classical loci (Ober et al. 1998); (2) pigs, which express up to three classical loci, *SLA-I*, *-2*, and *-3* and three non-classical loci *SLA-6*, *-7*, and *-8* (Smith et al. 2005); (3) chickens, which express two classical loci, *B-F major* and *minor* and one non-classical locus, *Y-F* (Kaufman et al. 1999); and (4) rats, which express one to three classical loci, *RT1-A1*, *-A2*, and *-A3* (Joly et al. 1996), and various combinations of non-classical loci, *RT1-C*, *RT1-E*, *RT-BM1* and *RT1-N1* (Kirisits et al. 1992; Lau et al. 2000). A pattern of variable numbers of expressed class I genes among MHC haplotypes has also been described for the rhesus macaque, pig, cow, and sheep (Babiuk et al. 2007; Birch et al. 2008; Ellis et al. 1999; Miltiadou et al. 2005; Otting et al. 2005; Tanaka-Matsuda et al. 2009).

Approximately 15 distinct horse MHC class I haplotypes can be discerned by serological assays (lymphocyte microcytotoxicity assays), designated ELA-A1, -A2, -A3, ..., -A19, but few advances have been made in antibody based detections methods in recent years (Lazary et al. 1988). Molecular studies of horse MHC class I sequences have been performed by cDNA library screening and RT-PCR experiments (Barbis et al. 1994; Chung et al. 2003; Ellis et al. 1995; Holmes and Ellis 1999; McGuire et al. 2003). In these studies, at least two and up to five classical MHC class I sequences were detected in each horse. While these studies provided a large number of coding sequences from different horses, the number of classical MHC class I loci expressed from each ELA haplotype was not defined but could range from one to five. The major hindrance to developing molecular typing methods is the inability to identify alleles of shared loci across haplotypes. Thus, MHC class I typing in horses remains dependent upon serological methods, despite the availability of nearly 60 horse MHC class I sequences.

An investigation of the evolutionary history of equine MHC class I genes identified five sequence groups, A to E (Holmes and Ellis, 1999). Group B sequences had high levels of polymorphism in the peptide binding region and several of these gene products are recognized by alloreactive T cells, consistent with classical loci (Holmes and Ellis, 1999). With the sequences and genomic resources available to them, Holmes and Ellis (1999) were unable to assign Group B sequences (putative classical MHC class I genes) to individual loci. Groups C, D, and E had acquired little allelic variation since the divergence of Equidae and

Rhinocerotidae, and therefore the authors suggested that these may be non-classical loci. Although group A sequences exhibit the hallmarks of non-classical genes (nonpolymorphic, truncated at the 3' end) their status was considered ambiguous (Holmes and Ellis, 1999).

Recently we used resources of the Horse Genome Project to characterize 15 MHC class I loci from a Bacterial Artificial Chromosome (BAC) contig of the ELA-A3 haplotype (Gustafson et al., 2003, Tallmadge et al. 2005). This haplotype was also carried by the DNA donor horse for the equine genome sequence (Wade et al., 2009). Seven of these loci are expressed in lymphocytes (locus 3.1 to 3.7, Tallmadge et al. 2005). Four of the expressed genes of the ELA-A3 haplotype were identified in the study by Holmes and Ellis (1999): locus 3.1 encodes *ECMHCB2* (3.1 indicates the haplotype (3 for ELA-A3) and locus (ELA locus 1)), 3.5 encodes *ECMHCA1*, 3.6 encodes *ECMHCC1*, and 3.7 encodes *ECMHCE1* (Supplementary Table 1).

In this paper we use the MHC class I gene framework that we developed (Tallmadge et al., 2005) and RT-PCR cloning to determine the pattern of polymorphism of horse MHC class I sequences and to identify alleles of shared loci from ten different serologically-defined ELA haplotypes.

MATERIALS AND METHODS

Animals: Blood samples, cDNA and, genomic DNA preparation

The horses sampled for this study were from either the Equine Genetics Center herd at Cornell University, from local farms in the Ithaca, NY area, or from Dr. Robert Mealey's research herd at Washington State University. Peripheral blood lymphocytes were isolated from heparinized samples of venous jugular blood using methods described previously (Antczak et al. 1982). Aliquots of peripheral blood lymphocytes were snap-frozen in liquid nitrogen and stored at -80° C until use. RNA isolation was performed with the RNeasy kit (Qiagen Inc., Valencia, CA). One microgram of RNA was treated with DNase I (Invitrogen, Carlsbad, CA) to degrade any contaminating genomic DNA. Synthesis of cDNA was performed as directed by M-MLV Reverse Transcriptase (USB, Cleveland, OH), using an oligo(dT) primer. Genomic DNA samples were isolated from the peripheral blood lymphocytes using the DNeasy Tissue Kit (Qiagen Inc., Valencia, CA).

MHC class I gene amplification, cloning, and sequencing

Oligonucleotide primers were designed with Primer3 on the web at: <http://www-genome.wi.mi.edu/cgi-bin/primer/prime3.cgi> and synthesized by Integrated DNA Technologies (Coralville, IA). Primers sequences are listed in Supplementary Table 2. Amplification was performed with *Pfu* DNA polymerase (Stratagene, La Jolla, CA) and reactions contained 1X PCR buffer, 0.4 mM dATP, dCTP, dGTP, dTTP, 1.25 U *Pfu*, 0.6 µM each primer and 50 ng of template DNA. Amplification reactions were performed with the following conditions: one cycle at 95°C, 45 seconds, thirty-three cycles of: 95°C for 30 seconds, 60°C for 30 seconds, 72°C for two minutes, and one cycle of 72°C for 10 minutes. Products were separated by electrophoresis on 1% agarose gels, stained with ethidium bromide and visualized under a UV Transilluminator (UVP, Upland, CA).

PCR products of MHC class I genes were cloned before sequencing. QIAquick kits (Qiagen Inc., Valencia, CA) were used for PCR product purification. The pCR® 4Blunt-TOPO vector was used for cloning *Pfu* DNA polymerase products, as described in the Zero Blunt® TOPO® PCR Cloning Kit cloning kit (Invitrogen, Carlsbad, CA).

Sequencing was performed by the Biotechnology Resource Center at Cornell University, Ithaca, NY with an ABI 3700 automated sequencer (Perkin Elmer, Foster City, CA). Templates

were sequenced at least once in each direction. All sequences generated for this work have been submitted to GenBank.

Sequence analysis

Sequencing electropherograms were viewed with EditView 1.0.1 software from Applied Biosystems (Foster City, CA). Sequences were analyzed with VectorNTI (Invitrogen, Carlsbad, CA). Alignments and sequence identity tables were generated by MegAlign from the LaserGene package (DNASTAR, Inc., Madison, WI). Primer sequences were removed prior to analysis. Most sequences were identified in four or more clones (Table 1). When discrepancies occurred, a second RT-PCR and cloning step was performed, and the sequences of these clones were used to resolve the discrepancies. Some sequences were confirmed by matching sequences from genomic subclones or sequences in the GenBank database. Other sequences were confirmed with sequences obtained from unrelated animals of the same ELA haplotype, as determined by serology.

Phylogenetic trees were constructed using the neighbor-joining method (Saitou and Nei 1987), as performed with human, mouse, and pig MHC sequences (Ando et al. 2003; Gu and Nei 1999). Sequences were analyzed on the basis of Tamura-Nei distances (Tamura and Nei 1993) using the *MEGA* version 3.1 program (Kumar et al. 2004). To determine the level of support for each node, bootstrap resampling was performed with 1,000 replications. HLA-A2*0101 sequence (HSA278305) was included in each analysis as an outgroup to root the phylogenetic tree.

The level of amino acid polymorphism at equine MHC class I loci was calculated from the variability index described by Wu & Kabat (1970) and plotted against the amino acid position of the mature protein. The variability is determined by the number of different amino acids at a given residue divided by the frequency of the most common amino acid at that residue. The complete cDNA sequence was not available for all alleles and no value was plotted for missing residues.

The mean number of synonymous (d_S) and non-synonymous (d_N) nucleotide substitutions per site was determined for the antigen recognition site (ARS) and conserved (CONS) regions among alleles of equine MHC class I loci using the SNAP program available at www.hiv.lanl.gov (Korber 2000). The significance of the difference between these rates was determined with a t-test; the p-values reflect the probability of rejecting the null hypothesis that $d_S = d_N$, * indicates $p < 0.05$, ** indicates $p < 0.001$.

Nomenclature

Horses homozygous for five distinct Equine Leukocyte Antigen (ELA) haplotypes commonly found in Thoroughbreds and Standardbreds were used in this study, denoted as ELA-A2, -A3, -A5, -A9 and -A10 (Antczak et al. 1986). To extend our findings, sequences from horses with additional MHC class I haplotypes and sequences available in GenBank were analyzed with the initial data set. Serological (ELA) types of horses were determined using clustered panels of alloantisera verified in international workshops (Lazary et al., 1988) combined in some cases with breeding history and results of mixed lymphocyte cultures.

Equine MHC class I sequences were initially given local names, and then renamed to reflect allelic series identified by phylogenetic analysis. Even though the sequences are not full-length they were named to conform to MHC allele naming conventions specified by Ellis et al. (2006). Allele names are based on the amino acid sequence and have five digits: digits 1 to 3 designate the allele group and digits 4 to 5 signify coding changes. Allele groups contain a maximum of four amino acid changes within the alpha 1 and alpha 2 domains plus up to 4

changes in the remaining coding sequence. Locus numbers prefix allele names when possible; otherwise allele names are prefixed with 'N' for not assigned. Within the N series, classical alleles start at N*00101 and non-classical alleles start at N*50001 (Supplementary Table 3).

RESULTS

Sequences of MHC class I genes were determined from horses of known ELA haplotypes determined by serological testing to characterize the genomic structure of horse MHC haplotypes and the polymorphism of horse MHC class I genes. Initially, samples of cDNA from five MHC homozygous horses were used (ELA-A2, -A3, -A5, -A9, -A10, Table 1). The sequences were analyzed in the context of seven expressed MHC class I genes previously identified in the ELA-A3 haplotype (Supplementary Table 1, Tallmadge et al., 2005). A combination of conserved and locus-specific primers was used for amplification (Supplementary Table 2), as described below.

Prevalence and polymorphism of horse non-classical MHC class I loci

Partial coding sequence (more than 700 bases) was obtained for each of the three non-classical loci and compared to the previously published sequences (Supplementary Table 1). ELA locus 5 sequences were detected in four of the five haplotypes assessed, and all haplotypes expressed ELA locus 6 and 7 products. High quality sequences were obtained from three or four ELA haplotypes for each locus, and nucleotide identity values were 100% for ELA locus 5, 97.7 – 100% for ELA locus 6, and 96.8 – 99% for ELA locus 7 (Table 1). From this limited set of sequences and the previously identified *ECMHCA1* (locus 5), *ECMHCC1* (locus 6), and *ECMHCE1* (locus 7) genes, locus 5 appears to be monomorphic (sequence *Eqca-5*00101*) whereas locus 6 and 7 appear bi-allelic (sequences *Eqca-6*00101*, *Eqca-6*00201*, *Eqca-7*00101*, and *Eqca-7*00201*). Phylogenetic analysis of these sequences, edited to exclude the antigen recognition site (ARS) residues, revealed clear clustering of allelic sequences by locus (Supplementary Figure 1).

MHC class I sequences from five ELA haplotypes

The remaining sequences were obtained with either ELA locus 2 specific primers or conserved primers (Supplementary Table 2) that amplify horse MHC class I genes other than the non-classical ELA loci 5, 6, and 7. Phylogenetic analysis was performed with the newly obtained sequences and perissodactyla sequences from Holmes and Ellis (1999), including horse, Przewalski's horse, donkey, onager, and Grevy's zebra sequences. Three functional classical MHC class I genes were included in group B by Holmes and Ellis (8–9, 1–29, *ECMHCB2*). Trees were constructed from an alignment of all sequences from exons 4 to 8, encoding the alpha 3, transmembrane, and cytoplasmic domains. All newly obtained sequences were placed within group B and similarly encoded a transmembrane domain of 35 amino acids in length (data not shown). The remainder of this paper describes our efforts to assign group B sequences to distinct loci.

ELA Locus 2 sequences

In a previous study, the MHC class I gene 3.2 (or ELA locus 2) was sequenced from a genomic subclone from the ELA-A3 MHC haplotype (DQ083408) and was found to have over 97% nucleotide identity to the cDNA sequence of the 8–5 gene (AY225157, Tallmadge et al. 2005). The identity between two random horse MHC class I sequences is usually below 92%. The 8–5 cDNA sequence was obtained from a horse of the ELA-A1/-A4 haplotype and protein expression of this gene was confirmed by transfection studies (McGuire et al. 2003). Because putative alleles of this class I gene were found in two different ELA haplotypes, we surveyed the ELA-A2, -A5, -A9, and -A10 haplotypes for mRNA expression of this gene. Locus-specific primers were designed in the hypervariable region and 3'UTR of the gene and used in RT-PCR

experiments (Supplementary Table 2). A product was detected and sequenced from all four haplotypes. Direct sequencing of the amplicon revealed the presence of only one product per horse, and each had from 95.4 – 99.6% nucleotide identity to the 3.2 gene sequence. The pattern of polymorphism of this locus was atypical because more nucleotide variation was found outside of the hypervariable region (4.1%) than within (1.7%). These sequences were named *Eqca-2*00101* to *Eqca-2*00303*, based on identity to the previously characterized ELA locus 2 (Table 1). Phylogenetic analysis of the conserved coding region (excluding the 57 amino acid positions predicted to comprise the ARS) clustered all six locus 2 sequences together with high bootstrapping values (Supplementary Figure 2A), and similar results were obtained with analysis of the 3'UTR sequence (Supplementary Figure 2B).

Remaining MHC class I group B classical sequences

For MHC class I group B sequences other than those encoded by locus 2 (above), we were unable to design locus-specific RT-PCR primers that would work across haplotypes. Thus, for the remaining MHC class I sequences, conserved primers that span ~80% of the horse MHC class I transcript, extending from the alpha 1 domain to the 3'UTR, were used in RT-PCR (Supplementary Table 2). Homozygous horses of five ELA haplotypes (ELA-A2, -A3, -A5, -A9, and -A10) were surveyed, and two to four sequences were identified per homozygote (Table 1). A phylogenetic tree was constructed based on the nucleotide alignment of the conserved regions of the MHC class I gene (Supplementary Figure 2A). Although clear locus clustering did not occur for all sequences, provisional locus assignments could be made. In addition to locus 2 described above, three groups were highly supported by the bootstrap values: (1) sequences *Eqca-16*00101*, *Eqca-16*00201*, and *Eqca-16*00301*, (2) sequences *Eqca-1*00101*, *Eqca-1*00201*, and *Eqca-1*00301*, and (3) sequences *Eqca-N*00201*, *Eqca-N*00301*, and *Eqca-N*00401*. Sequences *Eqca-N*00501* and *Eqca-N*00701* were paired but had low bootstrap support. The *Eqca-N*00101* sequence was placed in the tree but not tightly within a clade, also at low bootstrap values. Last, sequences *Eqca-3*00101* and *Eqca-4*00101* did not cluster with any other sequences. Although the *Eqca-3*00101* and *Eqca-4*00101* cDNA sequences were not found in the clones sequenced, it is known that they represent distinct loci and are expressed in the ELA-A3 haplotype (Tallmadge et al. 2005) and were included in all analyses.

We designate the cluster containing *Eqca-1*00101*, *Eqca-1*00201*, and *Eqca-1*00301* as ELA locus 1, the cluster containing sequences *Eqca-16*00101*, *Eqca-16*00201*, and *Eqca-16*00301* as ELA locus 16*; and the cluster containing sequences *Eqca-N*00201*, *Eqca-N*00301*, and *Eqca-N*00401* as ELA locus 17*. The first cluster is named locus 1 because the 3.1 gene (*Eqca-1*00101* allele, identical to *ECMHCB2*) has been physically mapped and fully sequenced. We chose locus designations 16 and 17 because a previous publication identified 15 horse MHC class I genes and pseudogenes from the ELA-A3 haplotype, which were named 3.1 through 3.15 (Tallmadge et al. 2005). The new designations are provisional because they lack confirmation at the genomic level and are not yet mapped. To reflect this provisional status, the locus numbers are given in italics and followed by an asterisk. Sequences *Eqca-N*00501* and *Eqca-N*00701* are not considered to be alleles of a shared locus, and like the other single sequences (*Eqca-N*00101*, *Eqca-3*00101*, *Eqca-4*00101*), they most likely represent additional loci.

Next, the 3'UTR sequences were analyzed for the classical MHC class I cDNA sequences. High overall nucleotide identity was evident, with the exception of single nucleotide polymorphisms distributed throughout the region. These polymorphisms were often shared between allelic sequences and portray locus-specific trends. The four groups identified in Supplementary Figure 2A are conserved in the phylogenetic tree generated from the 3'UTR sequence alignment (Supplementary Figure 2B), supporting the allelic assignments made

above. As in Supplementary Figure 2A, the *Eqca-N*00101* sequence was placed in the tree, but not tightly within a cluster, and sequences *Eqca-3*00101* and *Eqca-4*00101* consistently did not cluster with any other sequences.

MHC class I sequences from five additional ELA haplotypes

To extend these findings to additional ELA haplotypes (ELA-A1, -A4, -A6, -A7, and -W11), MHC class I sequences were obtained from six horses of various breeds that were either homozygous or heterozygous for the various ELA haplotypes (Table 1). Other published horse MHC class I sequences linked to specific ELA haplotypes were also included (Supplementary Table 1). We did not test these horses for products of the non-classical loci 5, 6, and 7. As for the ELA homozygous horses, locus 2 was tested separately and conserved primers were used for other group B sequences.

ELA Locus 2 sequences

Locus 2 specific primers successfully amplified sequences from these additional horses, adding three new sequences and assigning gene 8–5 to haplotype ELA-A4 (allele *Eqca-2*00301*) (Figure 1, Supplementary Figure 3). Two alleles of locus 2 were not always obtained from MHC heterozygous horses, most likely due to unequal amplification rather than absence of the second allele. Locus 2 amplification was not carried out on horses 0197 or 3832.

Remaining MHC class I group B sequences

Additional sequences were obtained from MHC heterozygous horses in this study with the conserved MHC class I primer set. These were added to previously published full-length or near full-length sequences in Genbank (Supplementary Table 1). After removal of the antigen recognition site, these sequences were analyzed together with the previous data set. The 3'UTR sequences determined in this study were also analyzed, but 3'UTR sequence was not available from alleles reported by Ellis and co-workers (1995). The phylogenetic tree of the coding sequences is shown in Figure 1, and the tree of the 3'UTR sequences is shown in Supplementary Figure 3. As above, the results from the coding sequence are consistent with those of the 3'UTR. The four locus groupings described above remained intact (locus 1, 2, 16*, 17*). Four sequences were added to locus 1: *Eqca-1*00401*, *Eqca-1*00501*, *Eqca-1*00601*, and *Eqca-1*00701*. Sequence *Eqca-1*00401* was obtained from an ELA-A1/-A4 horse, and was previously associated with the ELA-A4 haplotype by Chung et al. (2003). Sequence *Eqca-1*00501* was obtained from two heterozygous horses, both carrying the ELA-A6 haplotype, confirming the previously identified association (Chung et al. 2003). Sequence *Eqca-1*00601* was found in a homozygous ELA-A7 pony in this study. The identical sequence was previously described by Ellis et al. (1995) as *ECMHCB1*, from an ELA-A3/-A7 MHC heterozygous pony. The *ECMHCB1* assignment is of special interest because it was identified at the same time as the *Eqca-1*00101* cDNA sequence (MHC class I gene 3.1, identical to *ECMHCB2*), but the allelic relationship was not recognized at that time. *Eqca-1*00701* matches the 104 sequence detected in ELA-W11 heterozygous horses by Chung et al. (2003).

Three sequences were added to locus 16*: *Eqca-16*00401*, *Eqca-16*00501*, and *Eqca-16*00601*. Both *Eqca-16*00401* and *Eqca-16*00501* were identified from the ELA-A6/-A7 pony. These alleles are evidently expressed from the same locus, one from the ELA-A6 haplotype, and the other from the ELA-A7 haplotype. Sequence *Eqca-16*00501* was also identified in an ELA-A7/-A19 horse, confirming its presence in the ELA-A7 haplotype.

Only one sequence was added to locus 17*, *Eqca-N*00302*, which was found in two heterozygous horses that share the ELA-A1 haplotype. Another provisional locus, 18*, was identified from alleles *Eqca-N*00501* and *Eqca-N*00601* from ELA-A10 and -A1 haplotypes,

respectively. These alleles cluster with very high bootstrap support in both the conserved coding sequence and 3'UTR sequence.

Sequences *Eqca-3*00101*, *Eqca-4*00101*, *Eqca-N*00101*, *Eqca-N*00701*, *Eqca-N*00801*, *Eqca-N*00901*, *Eqca-N*01001*, and *Eqca-N*50001* remain ungrouped despite the additional data, but allelic sequences are likely to be identified when more sequences from other haplotypes are identified. The *Eqca-N*01001* and *Eqca-N*50001* alleles were originally identified as *ECMHCB4* and *ECMHCB3* by Ellis and co-workers (1995).

The amino acid alignment of all sequences used in these analyses is shown in Figure 2. In addition to variation in the alpha 1 and alpha 2 domains, differences were found throughout the alpha 3, transmembrane, and cytoplasmic regions even between alleles. This has contributed to the difficulty of defining horse MHC class I allelic series to date. Notably, the cytoplasmic region of sequences *Eqca-3*00101*, *Eqca-4*00101*, and *Eqca-N*50001* were very divergent compared to the other equine class I sequences. The low polymorphism and 3' end sequence differences that distinguish the non-classical loci are apparent for loci 5, 6, and 7. Putative locus-specific residues can be discerned. The nucleotide alignment of the 3'UTR sequences indicates that locus-specific residues may be present in this region also (Supplementary Figure 4).

Further support of our locus assignments were obtained from a limited study of intronic sequences. Intron 2 sequences were amplified with primers in the alpha 1 and alpha 2 coding regions and sequenced. Intron 2 sequences (approximately 240 bases) from the non-classical loci 5, 6, and 7 were compared between the ELA-A2 haplotype (GenBank DQ145590-2) and the ELA-A3 haplotype (Tallmadge et al. 2005). Within each locus, the intronic sequence identities were 96.3 to 100%, and the phylogenetic trees gave the same locus groupings as the coding sequence (Supplementary Figure 5). Intron 2 sequences from locus 2 also clearly clustered together. For the classical loci, all seven intron sequences were obtained from the 1–29 (*Eqca-16*00101*) gene (DQ145583 – 9) and were compared to the previously published intron sequences of the 8–9 (*Eqca-N*00101*) gene (AF207854, Carpenter et al. 2001), and 3.1 through 3.4 (DQ083407 – 10). These classical MHC class I sequences are not allelic and no phylogenetic clustering was detected (data not shown).

The pattern of amino acid sequence polymorphism of the designated loci is displayed by variability plots (Figure 3), determined using the formula published by Wu and Kabat (1970). The plots of loci 1 and *I6** are typical of classical, polymorphic class I loci, such as HLA-A and -B (Figure 3A). The plots of loci 2 and *I7** are atypical for classical loci because there is little variation in the antigen recognition sites of alpha 1 and alpha 2 domains and more variability in the conserved regions, especially the transmembrane domain. Further, the polymorphism found at each site is limited: often there are only two differences at a given polymorphic residue (each allele has one or the other) rather than many differences at a particular site. However, these limited differences do not create a bi-allelic locus because each allele is different, in contrast to the bi-allelic polymorphism detected in non-classical loci 6 and 7 (Figure 2). The patterns of polymorphism differ between the non-classical loci; the variable positions of locus 6 are spread across the MHC class I molecule whereas the variation of locus 7 is concentrated in the alpha 1 and alpha 2 domains (Figure 3B).

Although the variability plots reflect non-synonymous nucleotide substitutions, they are not informative for synonymous substitutions. The mean number of synonymous (d_S) and non-synonymous (d_N) nucleotide substitutions per site was determined for loci 1, 2, *I6**, *I7**, 5, 6, and 7. Comparisons were made between the antigen recognition site (ARS) and the remainder of the coding sequence (Figure 4). For loci 1 and *I6**, d_N was significantly higher than d_S in the ARS, and the converse was found in the remainder of the molecule. This pattern of

nucleotide substitution is consistent with that of the classical HLA loci (Hughes and Nei 1988) and it implies that the ARS is under positive selection for amino acid variation. In contrast, very few non-synonymous substitutions were identified in the ARS of loci 2 and 17*, and no synonymous changes were found. In the conserved region of loci 2 and 17*, the d_N was similar to the conserved region of loci 1 and 16*, and d_S was decreased. This pattern is atypical of classical loci and is more similar to non-classical MHC class I loci. The profile of non-classical locus 6 was very similar to locus 2, although locus 6 had many more substitutions in the conserved coding sequence. Locus 7 had more substitutions in the ARS than locus 6, and many fewer in the conserved region. Unlike the classical loci, d_S exceeded d_N in the ARS of locus 7, suggesting that the non-synonymous sites are under purifying selection at this locus. This is likely due to functional constraints of this molecule.

Collectively, we identified five allelic series comprised of 28 sequences obtained from ten ELA haplotypes (Figure 5). This number excludes the non-classical loci 5, 6, and 7 and putative non-classical sequence *ECMHCB3* (N^*50001). Seven sequences from this study remain ungrouped at this time; allelic series may be identified in the future or they might be unique to a particular haplotype. A model summarizing the genomic structure of these ten ELA haplotypes is presented in Figure 5, but the gene order is known only for the ELA-A3 haplotype (Tallmadge et al., 2005). This model does not imply gene location for the provisional loci and these assignments need to be confirmed with further studies on genomic DNA from multiple ELA haplotypes. Finally, further support for our allelic assignments comes from our observation that no horses tested carried more than two alleles assigned to any locus.

DISCUSSION

Horse MHC class I gene sequences were obtained from ten MHC haplotypes to investigate the genomic structure and polymorphism of the equine MHC. Comparisons of MHC class I coding and 3' untranslated region sequences from five homozygous ELA haplotypes and five additional ELA haplotypes were performed and provisional locus assignments were made. The horse MHC class I genes did not cluster into a small number of discrete loci: five allelic groups were created, accounting for 28 of the 35 alleles identified. Two of the remaining seven sequences represent known loci, and the other five may be individual representatives of additional loci. Together, the data suggests that at least seven expressed MHC class I loci are present among horses of these ten ELA haplotypes, and that the total number of expressed MHC class I loci varies among ELA haplotypes.

Additionally, the non-classical loci 5, 6, and 7 are found in nearly all haplotypes. On a per haplotype basis, including the individual sequences found, each haplotype in this study expresses at least three MHC class I loci in addition to the non-classical loci 5, 6, and 7. This is consistent with the range of classical loci (between one and five) predicted by previous cDNA studies (Barbis et al. 1994; Chung et al. 2003; Ellis et al. 1995; McGuire et al. 2003). Comparatively, this varied complement of expressed classical loci is consistent with the MHC haplotype organization described in the cow, sheep, and rhesus macaque, and unlike that of the human MHC in which a few MHC class I classical loci are common to all haplotypes (Ellis et al. 1999; Miliadou et al. 2005; Otting et al. 2005). Interestingly, this proposed structure appears similar to the human natural killer receptor gene cluster (Trowsdale et al. 2001).

This study identified 50 horse MHC class I sequences, of these 15 sequences represent non-classical loci and the remaining 35 sequences represent classical loci. Eighty percent (28/35) of the classical sequences were assigned to allelic series. Phylogenetic analysis of 35 classical sequences analyzed here and those identified by Holmes and Ellis (1999) placed them in group B, and all sequences in this group have the common trait of a 35 amino acid residue transmembrane domain. The sequences identified here tend to be poly-allelic, however the

level of polymorphism within loci varies. Some loci are similar to HLA-A and -B loci (1, 16*) and some have more variation outside the ARS (2, 17*). For the non-classical genes (ELA locus 5 (ECMHCA1), ELA locus 6 (ECMHCC1), and ELA locus 7 (ECMHCE1)) most individuals carry only a single allele. A second allele has been identified for loci 6 and 7, but the frequency of the second allele is unknown. Between alleles there are a small number of differences that usually occur outside the ARS.

Two loci blurred the traditional classical and non-classical paradigm established in other species. Loci 2 and 17* have a 35 amino acid transmembrane region and 3' end characteristic of group B sequences, which includes several functional classical class I genes (Holmes and Ellis 1999). These loci are polymorphic; many minor variants are apparent in the variability plots. However, the variability is not clustered in the ARS as is typical of classical loci. The alternative pattern of polymorphism may indicate that the antigenic repertoire presented by these molecules is limited in comparison with other classical loci, but functional studies have not been performed. These genes may be similar in nature to *HLA-C*, which exhibits much less polymorphism than *HLA-A* or *-B* and is expressed at lower mRNA levels (Johnson 2000). ELA locus 2 was the only class I gene detected in all horses tested.

Sequencing of MHC class I genes in several species has identified a greater level of allelic diversity and complexity than determined by serological characterization of haplotypes (Chardon et al. 1999; Ellis 2004; Ellis et al. 2005; Otting et al. 2005; Parham et al. 1995). We found evidence for extensive differences in haplotype structure within a single horse breed. In fact, horses appear to have more haplotype diversity than humans or mice, perhaps reflecting differences in evolutionary history or a more rapid rate of MHC class I gene duplication between horses and other species. Further studies of isolated breeds, such as Icelandic or Przewalski horses, may provide insight to causal evolutionary pressures.

Functional characterization has been performed for two of the sequences described here. *Eqca-1*00101* of locus 1 can present EHV-1 epitopes to CTLs and elicit target cell lysis (Kydd et al. 2006), and *Eqca-N*00601* in locus 18* can present equine infectious anemia virus (EIAV) epitopes to CTL effectors (Mealey et al., 2006). Mealey and co-workers (2006) also demonstrated that an allele of *Eqca-N*00601*, named 7-6, differed by only one amino acid. The 7-6 allele also presented EIAV epitopes to CTL, but epitope recognition differed from that presented by *Eqca-N*00601*. The functional analyses listed above confirm that these loci encode classical MHC class I molecules. No alleles of locus 1 or 16*, which encode highly polymorphic classical loci, were identified in the ELA-A1 haplotype. We hypothesize that in lieu of locus 1 or 16*, locus 18* encodes functional and polymorphic classical alleles in the ELA-A1 haplotype. Functional analysis of additional class I genes will be informative, especially those alleles of locus 2 and 17* that differ from the classical/non-classical definition.

In conclusion, our analysis suggests that ELA MHC class I haplotypes exhibit both polymorphism at shared loci and variation in the number of expressed MHC class I genes per haplotype. Further investigation of genomic DNA will be required to confirm whether the provisional assignments of alleles to distinct loci presented in this study are correct, and whether the absence of allelic mRNA expression equates with absence of the locus in genomic DNA. Recently, the horse genome sequence was determined from an individual closely related to the BAC library donor and these horses share the ELA-A3 haplotype (Wade et al., 2009). Analysis of the horse genome sequence confirmed the gene order of the expressed ELA-A3 MHC class I sequences along the chromosome (<http://genome.ucsc.edu/cgi-bin/hgGateway>).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

ARS	antigen recognition site
BAC	bacterial artificial chromosome
CTL	cytotoxic T lymphocyte
ELA	equine leukocyte antigen
HLA	human leukocyte antigen
MHC	major histocompatibility complex
NK	natural killer
UTR	untranslated region

Acknowledgments

This research was funded by NIH grants HD-34086, HD-15799, and HD-049545, the Harry M. Zweig Memorial Fund for Equine Research, the Dorothy Russell Havemeyer Foundation, Inc., and the USDA Regional Grants Program. We thank Dr. Shirley Ellis, Dr. Ashley Moffett, Dr. John Trowsdale, and Dr. Stephan Beck for helpful discussions, and Dr. Robert H. Mealey of Washington State University for generously providing samples from horses of known ELA serotypes.

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. Genetic polymorphism of the swine major histocompatibility complex (SLA) class I genes, SLA-1, -2 and -3. *Immunogenetics* 2003;55:583–593. [PubMed: 14605775]
- Antczak DF, Bailey E, Barger B, Guerin G, Lazary S, McClure J, Mottironi VD, Symons R, Templeton J, Varewyck H. Joint report of the Third International Workshop on Lymphocyte Alloantigens of the Horse, Kennett Square, Pennsylvania, 25–27 April 1984. *Anim Genet* 1986;17:363–373. [PubMed: 3826760]
- Antczak DF, Bright SM, Remick LH, Bauman BE. Lymphocyte alloantigens of the horse. I. Serologic and genetic studies. *Tissue Antigens* 1982;20:172–187. [PubMed: 6182639]
- Babiuk S, Horseman B, Zhang C, Bickis M, Kusalik A, Schook LB, Abrahamsen MS, Pontarollo R. BoLA class I allele diversity and polymorphism in a herd of cattle. *Immunogenetics* 2007;59:167–176. [PubMed: 17206412]
- Barbis DP, Maher JK, Stanek J, Klaunberg BA, Antczak DF. Horse cDNA clones encoding two MHC class I genes. *Immunogenetics* 1994;40:163. [PubMed: 8026865]
- Birch J, Codner G, Guzman E, Ellis SA. Genomic location and characterisation of nonclassical MHC class I genes in cattle. *Immunogenetics* 2008;60:267–273. [PubMed: 18431566]
- Bjorkman PJ, Parham P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev Biochem* 1990;59:253–288. [PubMed: 2115762]
- Carpenter S, Baker JM, Bacon SJ, Hopman T, Maher J, Ellis SA, Antczak DF. Molecular and functional characterization of genes encoding horse MHC class I antigens. *Immunogenetics* 2001;53:802–809. [PubMed: 11862413]
- Cereb N, Maye P, Lee S, Kong Y, Yang SY. Locus-specific amplification of HLA class I genes from genomic DNA: locus-specific sequences in the first and third introns of HLA-A, -B, and -C alleles. *Tissue Antigens* 1995;45:1–11. [PubMed: 7725305]
- Cereb N, Yang SY. The regulatory complex of HLA class I promoters exhibits locus-specific conservation with limited allelic variation. *J Immunol* 1994;152:3873–3883. [PubMed: 8144957]
- Chardon P, Renard C, Vaiman M. The major histocompatibility complex in swine. *Immunol Rev* 1999;167:179–192. [PubMed: 10319260]

- Chung C, Leib SR, Fraser DG, Ellis SA, McGuire TC. Novel classical MHC class I alleles identified in horses by sequencing clones of reverse transcription-PCR products. *Eur J Immunogenet* 2003;30:387–396. [PubMed: 14675391]
- Ellis SA, Bontrop RE, Antczak DF, Ballingall K, Davies CJ, Kaufman J, Kennedy LJ, Robinson J, Smith DM, Stear MJ, Stet RJ, Waller MJ, Walter L, Marsh SG. ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005. *Immunogenetics* 2006;57:953–958. [PubMed: 16389556]
- Ellis SA, Morrison WI, MacHugh ND, Birch J, Burrells A, Stear MJ. Serological and molecular diversity in the cattle MHC class I region. *Immunogenetics* 2005;57:601–606. [PubMed: 16086171]
- Ellis S. The cattle major histocompatibility complex: is it unique? *Vet Immunol Immunopathol* 2004;102:1–8. [PubMed: 15451610]
- Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, Morrison WI. Variation in the number of expressed MHC genes in different cattle class I haplotypes. *Immunogenetics* 1999;50:319–328. [PubMed: 10630296]
- Ellis SA, Martin AJ, Holmes EC, Morrison WI. At least four MHC class I genes are transcribed in the horse: phylogenetic analysis suggests an unusual evolutionary history for the MHC in this species. *Eur J Immunogenet* 1995;22:249–260. [PubMed: 8547231]
- Germain RN, Margulies DH. The biochemistry and cell biology of antigen processing and presentation. *Annu Rev Immunol* 1993;11:403–450. [PubMed: 8476568]
- Gu X, Nei M. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. *Mol Biol Evol* 1999;16:147–156. [PubMed: 10028282]
- Gustafson AL, Tallmadge RL, Ramlachan N, Miller D, Bird H, Antczak DF, Raudsepp T, Chowdhary BP, Skow LC. An ordered BAC contig map of the equine major histocompatibility complex. *Cytogenet Genome Res* 2003;102:189–195. [PubMed: 14970701]
- Holmes EC, Ellis SA. Evolutionary history of MHC class I genes in the mammalian order Perissodactyla. *J Mol Evol* 1999;49:316–324. [PubMed: 10473772]
- Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 1988;335:167–170. [PubMed: 3412472]
- Johnson DR. Differential expression of human major histocompatibility class I loci: HLA-A, -B, and -C. *Hum Immunol* 2000;61:389–396. [PubMed: 10715516]
- Joly E, Leong L, Coadwell W, Clarkson C, Butcher G. The rat MHC haplotype RT1c expresses two classical class I molecules. *J Immunol* 1996;157:1551–1558. [PubMed: 8759738]
- Kaufman J, Jacob J, Shaw I, Walker B, Milne S, Beck S, Salomonsen J. Gene organisation determines evolution of function in the chicken MHC. *Immunol Rev* 1999;167:101–117. [PubMed: 10319254]
- Kirisits MJ, Kunz HW, Hassett AL, Gill TJ 3rd. Genomic DNA sequence and organization of a TL-like gene in the grc-G/C region of the rat. *Immunogenetics* 1992;35:365–377. [PubMed: 1349585]
- Koller BH, Sidwell B, DeMars R, Orr HT. Isolation of HLA locus-specific DNA probes from the 3'-untranslated region. *Proc Natl Acad Sci U S A* 1984;81:5175–5178. [PubMed: 6591185]
- Korber, B. HIV Signature and Sequence Variation Analysis. In: Rodrigo, AG.; Learn, GH., editors. *Computational Analysis of HIV Molecular Sequences*. Dordrecht, Netherlands: Kluwer Academic Publishers; 2000. p. 55-72.
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 2004;5:150–163. [PubMed: 15260895]
- Kydd JH, Davis-Poynter NJ, Birch J, Hannant D, Minke J, Audonnet JC, Antczak DF, Ellis SA. A molecular approach to the identification of cytotoxic T-lymphocyte epitopes within equine herpesvirus 1. *J Gen Virol* 2006;87:2507–2515. [PubMed: 16894188]
- Lanier LL, Corliss B, Phillips JH. Arousal and inhibition of human NK cells. *Immunol Rev* 1997;155:145–154. [PubMed: 9059890]
- Lau P, Lorenzi R, Joly E. Comparison of RT-BM1 sequences from six different rat major histocompatibility complex haplotypes reveals limited variation, and alternate splicing in the 3' untranslated region. *Immunogenetics* 2000;51:148–153. [PubMed: 10663577]
- Lawlor DA, Zemmour J, Ennis PD, Parham P. Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. *Annu Rev Immunol* 1990;8:23–63. [PubMed: 2188663]

- Lazary S, Antczak DF, Bailey E, Bell TK, Bernoco D, Byrns G, McClure JJ. Joint Report of the Fifth International Workshop on Lymphocyte Alloantigens of the Horse, Baton Rouge, Louisiana, 31 October-1 November 1987. *Anim Genet* 1988;19:447–456. [PubMed: 2466424]
- Le Bouteiller P. HLA class I chromosomal region, genes, and products: facts and questions. *Crit Rev Immunol* 1994;14:89–129. [PubMed: 7702749]
- McGuire TC, Leib SR, Mealey RH, Fraser DG, Prieur DJ. Presentation and binding affinity of equine infectious anemia virus CTL envelope and matrix protein epitopes by an expressed equine classical MHC class I molecule. *J Immunol* 2003;171:1984–1993. [PubMed: 12902502]
- Mealey RH, Lee JH, Leib SR, Littke MH, McGuire TC. A single amino acid difference within the alpha-2 domain of two naturally occurring equine MHC class I molecules alters the recognition of Gag and Rev epitopes by equine infectious anemia virus-specific CTL. *J Immunol* 2006;177:7377–7390. [PubMed: 17082657]
- Miltiadou D, Ballingall KT, Ellis SA, Russell GC, McKeever DJ. Haplotype characterization of transcribed ovine major histocompatibility complex (MHC) class I genes. *Immunogenetics* 2005;57:499–509. [PubMed: 16028041]
- Moore KW, Sher BT, Sun YH, Eakle KA, Hood L. DNA sequence of a gene encoding a BALB/c mouse Ld transplantation antigen. *Science* 1982;215:679–682. [PubMed: 7058332]
- Mungall AJ, Palmer SA, Sims SK, Edwards CA, Ashurst JL, Wilming L, Jones MC, Horton R, Hunt SE, Scott CE, Gilbert JG, Clamp ME, Bethel G, Milne S, Ainscough R, Almeida JP, Ambrose KD, Andrews TD, Ashwell RI, Babbage AK, et al. The DNA sequence and analysis of human chromosome 6. *Nature* 2003;425:805–811. [PubMed: 14574404]
- Ober C, Aldrich C, Rosinsky B, Robertson A, Walker MA, Willadsen S, Verp MS, Geraghty DE, Hunt JS. HLA-G1 protein expression is not essential for fetal survival. *Placenta* 1998;19:127–132. [PubMed: 9548178]
- Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE. Unparalleled complexity of the MHC class I region in rhesus macaques. *Proc Natl Acad Sci U S A* 2005;102:1626–1631. [PubMed: 15665097]
- Parham P, Adams EJ, Arnett KL. The origins of HLA-A,B,C polymorphism. *Immunol Rev* 1995;143:141–180. [PubMed: 7558075]
- Parham P, Lawlor DA, Lomen CE, Ennis PD. Diversity and diversification of HLA-A,B,C alleles. *J Immunol* 1989;142:3937–3950. [PubMed: 2715640]
- Parham P, Lomen CE, Lawlor DA, Ways JP, Holmes N, Coppin HL, Salter RD, Wan AM, Ennis PD. Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc Natl Acad Sci U S A* 1988;85:4005–4009. [PubMed: 3375250]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425. [PubMed: 3447015]
- Smith DM, Lunney JK, Martens GW, Ando A, Lee JH, Ho CS, Schook L, Renard C, Chardon P. Nomenclature for factors of the SLA class-I system, 2004. *Tissue Antigens* 2005;65:136–149. [PubMed: 15713212]
- Steinmetz M, Frelinger JG, Fisher D, Hunkapiller T, Pereira D, Weissman SM, Uehara H, Nathenson S, Hood L. Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. *Cell* 1981;24:125–134. [PubMed: 6786753]
- Stroynowski I. Molecules related to class-I major histocompatibility complex antigens. *Annu Rev Immunol* 1990;8:501–530. [PubMed: 2188672]
- Swelsen WT, Voorter CE, van den Berg-Loonen EM. Polymorphism of intron 4 in HLA-A, -B and -C genes. *Tissue Antigens* 2003;61:475–483. [PubMed: 12823771]
- Tallmadge RL, Lear TL, Antczak DF. Genomic characterization of MHC class I genes of the horse. *Immunogenetics* 2005;57:763–774. [PubMed: 16220348]
- Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993;10:512–526. [PubMed: 8336541]
- Tanaka-Matsuda M, Ando A, Rogel-Gaillard C, Chardon P, Uenishi H. Difference in number of loci of swine leukocyte antigen classical class I genes among haplotypes. *Genomics* 2009;93:261–273. [PubMed: 18996466]

- Trowsdale J, Barten R, Haude A, Stewart CA, Beck S, Wilson MJ. The genomic context of natural killer receptor extended gene families. *Immunol Rev* 2001;181:20–38. [PubMed: 11513141]
- Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blocker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MC, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, Della Valle G, Fryc S, Guerin G, Hasegawa T, Hill EW, Jurka J, Kiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Roed KH, Ryder OA, Searle S, Skow L, Swinburne JE, Syvanen AC, Tozaki T, Valberg SJ, Vaudin M, White JR, Zody MC, Lander ES, Lindblad-Toh K. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 2009;326:865–867. [PubMed: 19892987]
- Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* 1970;132:211–250. [PubMed: 5508247]

Figure 1. Phylogenetic tree of horse MHC class I sequences from ten equine MHC haplotypes
The tree was constructed based on the coding sequence of horse MHC class I genes excluding residues of the antigen recognition site (Tables 1 and Supplementary Table 1). All of these sequences clustered with group B sequences defined by Holmes and Ellis (1999). Allele names are followed by associated serological ELA haplotype and exclude the prefix 'Eqca'. Putative locus assignments shown on the right, provisional locus assignments are given in italics and followed by an asterisk (*). Sequences were analyzed using the neighbor-joining method (Saitou and Nei 1987) on the basis of Tamura-Nei distances (Tamura and Nei 1993) using the *MEGA* 3.1 program. Percent of supporting bootstrap replications (n=1000) are shown next to the branches. HLA-A*0101 sequence included as an outgroup to root the phylogenetic tree.

Leader	$\alpha 1$	$\alpha 2$																		Consensus	MHC (ELA class II locus)	
		10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180			
	VRFDSDAASPRMEPRAPWMEQGEPEYMERETIRAGAAQTRVNLNTLGYNGOEA	SHLTLQRMGYCDGVPDGRLLRGYSQYATGADYIALNEDLRSTAADTAAGITRRKWEAAVGRWRNYL																				
--MPMTFILLLSGALILTETWAGSHSMRYKTAVSFRPGGEPRFIVSGYVDDTQF	R.V.....Q..E...T..H.....G..R.TAL.....	F.P.W.C.....R..P.F.....F.....E..QH.....	1*00101	A3																		
	Q..E...S.V.D.T.....G..N.....	F.P.W.C.....R..P.F.....A.....HD.....	1*00401	A4																		
--MPMTFILLLSGALILTETWAGSHSMRYFYTAVSFRPGGEPRFIVAGYVDDTQF	EN.KE.....Q.....N..R...H...S...E.D.....	Y.E.E.....S.R.....I.F.....A.....S.....QE.....	1*00501	A6																		
	EN.KE.....N.....Y..D.T...Y.E..R.TAL.....	W.....D.....HD.....	1*00601	A7																	1	
	EN.KE.....EN..R...H...S...E.....S.....	Y.E.....S.R.....I.F.....A.....L.....QR.....	1*00201	A9																		
	G.....V.....E...T..H.....G..N.....	F.P.W.C.....R..P.F.....HD.....	1*00301	A10																		
	D.I.....Q.....NM.E.T...G..R.A.L.S.....	F.P.W.C.....R..P.F.....KE.....	1*00701	W11																		
	C...L...P.....D.....T.....N..H.S...V..H.....	C...L...P.....D.....T.....N..H.S...V..H.....	2*00202	A1																		
MKVMPPTFILLLGGSLSTETWAGSHSLRYFYTAVSRPGGEPRFIVSGYVDDTQF	KE.....K.....N...Y.E.....A.....	C...L...R.....D.....S.....H.....	2*00201	A2																		
	C...L...R.....D.....T.....N..H.S...V..H.....	C...L...R.....D.....T.....N..H.S...V..H.....	2*00101	A3																		
	C...R.....D.....N..S.....H.....	C...R.....D.....N..S.....H.....	2*0301	A4																		
	C...R.....D.....N..S.....H.....	C...R.....D.....N..S.....H.....	2*00302	A5																	2	
	C...R.....D.....T.....N..S.....H.....	C...R.....D.....T.....N..S.....H.....	2*00401	A7																		
	KE.....K.....N...Y.E.....	C...R.....D.....T.....N..S.....H.....	2*00102	A9																		
	V.....Q..D..D..H.R.N.....	C...R.....D.....T.....N..S.....H.....	2*00303	A10																		
MKVMPPTFILLLGGSLSTETWAGSHSLRYFYTAVSRPGGEPRFIVAGYVDDTQF	V.....Q..D..D..H.R.N.....	C...R.....D.....T.....N..S.....H.....	2*00304	W11																		
MVMPMTFILLLSSGILTETWAGSHSMRYFYTAVSRPGGEPRFIVSGYVDDTQF	A.....L.....MS..N.N.S...G.....	E.....F.....F.....A.....T..E..D...F.....	3*00101	A3																	3	
--MPFTFILLLSSGILTETWAGSHSMRYFYTAVSRPGGEPRFIVSGYVDDTQF	SV.H.....	L..R.H.R.SASF.....A.D.F.....	4*00101	A3																	4	
	A.....KE..L.....EN..Y.N...E.....	L..R.H.R.SASF.....R.....S..L.....GY.....	16*00201	A5																		
	A.....KE.....N...Y.E.....D.....	L..R.H.R.SASF.....D.....A.....D.....	16*00501	A6																		
	A.....KE.....N...Y.E.....S.N.....	L..R.H.R.SASF.....E..D.....S.....GL.....	16*00501	A7																		
	AR.....KD..L...RV.....S.E.....	L..R.H.R.SASF.....S..L.....KE.....	16*00301	A9																	16*	
	V.....LN.QR..DN.....N.....	G.C.Y.....V.....S.....K.....ED.....GF.....	N*00302	A1																		
	LN.QR..DT.....N.....	G.C.Y.....V.....S.....K.....ED.....GF.....	N*00201	A5																		
	V.....LN.QR..DT.....N.....	G.C.Y.....V.....S.....K.....ED.....GF.....	N*00301	A9																	17*	
	LN.QR..DT.....N.....	G.C.Y.....V.....S.....K.....ED.....GF.....	N*00401	A10																		
	V.....NM.E.T.N...G...H.....	E.....S.R.....E.F.....L.....T..E..GY.....	N*00601	A1																		
MKVMPPTFILLLSSGILSTETLWAGSHSMRYFYTAVSRPGGEPRFIVAGYVDDTQFA	K.R.....N..P..D.T...S.N.....	D.....E.....S.R.....E.F.....L.....T..E..GY.....	N																		18*	
	V.....Q.....M.E.N...G..N.....	Y.L...E.L...QS.....E.F.....A.....ME.H.....	N*00101	A2																		
	BQ..R..DH.....	W.C...R.....Q.....F.....N.....D.....	N*00701	A9																		
	EN.KE.....V.....Q...BROV...YD.....	Y.L...E.L...QS.....E.F.....A.....ME.H.....	N*00801	A4																		
--MPMTFILLLSGALILTETWAGSHSMRYFTAVSFRPGGEPRFIVAGYVDDTQF	E...V..N..ID..D.....	Y.Y.....R.....R.D.....HD.....	N*00901	A7																		
	GYDDIUF.....	N..T..DN..A...S..N.....	5*00101	A2																		
--MPATFILLLSGTILTETWAGSHSMRYFLTAVSRPGGEPRFIVSGYVDDTQF	N..T..DN..A...S..N.....	F...E..T.....A...SK..K.E..D..ND..A...D.....	5*00101	A3																		
	GRGEPRFIVSGYVDDTQF.....	N..T..DN..A...S..N.....	5*00101	A9																		
--MPATFILLLSGTILTETWAGSHSMRYFLTAVSRPGGEPRFIVSGYVDDTQF	N..T..DN..A...S..N.....	F...E..T.....A...SK..K.E..D..ND..A...D.....	5*00101	A9																		
MKVTAPTKVLVLSGALALTKTWAGSHSMRYFTAVSFRPGGESWVLEVGYYDDTQF	E..QR..A.L.LHS..G..N.....	L.....H...SASF.....L.....SK..K.V..N..RY.....	6*00101	A3																	5	
	E..QR..A.L.LHS..G..N.....	L.....H...SASF.....L.....SK..K.V..N..RY.....	6*00101	A5																		
	E..QR..T.L.LHS..G..N.....	L.....H...SASF.....L.....LP.....SK..K...N..RY..G.N.....	6*00201	A9																		
MKVTAPTKVLVLSGALALTKTWAGSHSMRYFTAVSFRPGGESWVLEVGYYDDTQF	F.....E..QR..A.L.LHS..G..N.....	L.....H...SASF.....L.....SK..K.V..N..RY.....	6*00101	A10																		
	E..QR..A.L.LHS..G..N.....	L.....H...SASF.....L.....SK..K.V..N..RY.....	6*00101	A3																	6	
--MPFTFILLLSGALILTKTWAGSHSLRYENTALSRPGGEPRFLIVGYVDDTQF	EN.....V.....EN..T.I..ST..N...S..N.....	E.....E.....S.....S.....QD..A...D.....ES..QD..A...D.....	7*00101	A2																		
	T.I..ST..N...G.....	Y.WLC...L.R...H.....EN.....QD..A...D.....	7*00101	A5																		
	T.I..ST..N...G.....	Y.WLC...L.R...H.....EN.....QD..A...D.....	7*00101	A9																		
	T.I..ST..N...G.....	Y.WLC...L.R...H.....EN.....QD..A...D.....	7*00101	A5																		
MKVTAPTKVLVLSGALALTKTWAGSHSMRYFTAVSFRPGGESWVLEVGYYDDTQF	T.I..ST..N...G.....	Y.WLC...L.R...H.....EN.....QD..A...D.....	7*00101	A9																		
MKVTAPTKVLVLSGALALTKTWAGSHSMRYFTAVSFRPGGESWVLEVGYYDDTQF	T.I..ST..N...G.....	Y.WLC...L.R...H.....EN.....QD..A...D.....	7*00101	A9																		
	BQ..W..DS..NC..S..D...DN..H.....	LRF.....R.S.....	ECMHCB3	A3or7																		

$\alpha 3$										Transmembrane				Cytoplasmic			Consensus	MHC (ELA) class I locus	
190	200	210	220	230	240	250	260	270	280	290	300	310	320	330	340	350	360		
GTCTVNLRLYLENGKETLQRA	PPKTHVTHHPISDREVTLRC	WALGFYPAEILT	WQRGDELTDQDEL	VELTRPAGDGT	QKWA	VVPVSGEEQRYTCHQ	HEGLPEFVTLR	PEPPQSTILIVGLVGLVLL	---	VAVVAGAVIWRKR	SKRGSTVQ	ANNSDAQSGSDASLPQ	NV*						
.R.	.	.S.	.H.	.	.F.	.A.	.	.R.	.	.S.II.	.	.G.	.	.I.	.	.	.	1*00101 A3	
.R.	.	.	.	.	.	.	.	.	.	.IV.	.	.G.	.A.	.	.	.	.	1*00401 A4	
.E.	.	.H.	.	.	.F.	.	.	.R.	.	.I.	.	.G.	.	.I.	.	.	.	1*00501 A6	
.S.	.R.	.	.	.HE.	.	.	.	.R.	.	.VII.	.	.G.	.M.	.	.	.	.	1*00601 A7	1
.E.	.H.	.	.	.	.F.	.	.	.R.	.	.I.	.	.G.	.	.I.	.	.	.	1*00201 A9	
.R.	.	.H.	.	.	.F.	.	.	.Q.	.R.	.	.IV.	.	.G.	.A.	.	.	.	1*00301 A10	
.G.	.R.	.	.H.	.	.F.	.	.	.Q.	.R.	.	.IV.	.	.G.	.A.	.	.	.	1*00701 W11	
.	.N.	.	.	.A.	.	.V.	.	.	.V.	.S.II.	.	.G.	.A.	.M.	.	.T.	.	2*00202 A1	
.	.N.	.	.	.	.	.	.	.	.	.S.II.	.	.G.	.A.	.M.	.	.	.	2*00201 A2	
.	.R.	.	.H.	.	.A.	.	.	.L.	.	.	.G.P.	.	.V.	.	.T.	.	.	2*00101 A3	
.	.R.	.	.M.	.	.F.	.	.	.L.	.	.I.	.MG.	.	.	.M.	.	.	.	2*00301 A4	
.	.R.	.	.M.	.	.F.	.	.	.L.	.	.I.	.MG.	.	.	.M.	.	.	.	2*00302 A5	2
.	.R.	.	.H.	.	.	.	.	.L.	.	.G.P.	.	.V.	.	.T.	.	.	.	2*00401 A7	
.	.R.	.	.H.	.	.	.	.	.L.	.	.G.P.	.	.V.	.	.T.	.	.	.	2*00102 A9	
.	.R.	.	.A.	.	.M.	.	.F.	.	.	.I.	.G.	.	.	.M.	.	.	.	2*00303 A10	
.	.R.	.	.A.	.	.M.	.	.F.	.	.	.I.	.G.	.	.	.M.	.	.	.	2*00304 W11	
.	.H.	.W.	.M.	.	.C.	.NH.	.	.S.HQ.	.E.	.	.A.	.I.	.	.M.II.	.G.	.G.	.H.	3*00101 A3	3
.R.	.K.	.H.	.	.	.	.	.	.	.	.RK.	.R.	.P.	.I.	.G.	.	.H.	.GRA.	4*00101 A3	4
.R.	.S.	.R.	.	.E.	.	.AI.	.	.V.	.	.	.	.L.	.	.	.	.	.	16*00101 A2	
.E.	.	.V.	.	.E.	.	.A.	.	.M.	.	.	.	.L.	.	.	.	.	.	16*00201 A5	
.P.	.	.H.	.	.	.	.	.	.	.	.	.	.L.	.	.	.	.	.	16*00401 A6	16*
.	.	.	.	.AI.	.	.	.	.	.	.	.	.L.	.	.	.	.	.	16*00501 A7	
.R.	.	.V.	.	.AI.	.	.	.	.	.	.RK.	.	.L.	.	.	.	.	.	16*00301 A9	
.V.	.	.E.	.	.	.	.	.	.L.	.	.Q.	.R.	.	.	.	.	.	.	16*00601 W11	
.	.R.	.I.	.	.	.E.	.S.S.	.	.	.S.W.	.	.	.S.IIV.	.	.	.	.	.	N*00302 A1	
.	.R.	.I.	.	.	.E.	.S.S.	.	.	.	.	.S.	.S.IIV.	.	.	.P.	.	.	N*00201 A5	
.	.R.	.I.	.	.	.E.	.S.S.	.	.	.S.	.	.	.S.IIV.	.	.	.	.	.	N*00301 A9	17*
.	.R.	.I.	.	.	.	.	.	.K.	.R.	.	.I.	.	.G.	.G.	.M.	.	.	N*00401 A10	
.	.N.	.A.	.	.H.	.	.E.	.S.S.	.V.	.F.	.R.	.	.A.	.G.	.	.	.	.	N*00601 A1	
.	.N.	.A.	.	.H.	.	.E.	.S.S.	.V.	.R.	.	.A.	.L.	.	.G.	.	.A.	.I.	N*00501 A10	18*
.	.S.	.F.	.	.S.	.	.P.	.S.S.	.HQ.	.E.	.	.	.	.	.G.	.	.	.	N*00101 A2	
.	.	.	.	.E.	.S.S.	.HQ.	.E.	.	.D.	.N.	.	.T.	.R.	.	.	.	.	N*00701 A9	
.P.	.	.H.	.	.	.	.	.	.	.	.	.	.IIV.	.	.G.	.L.	.	.	N*00801 A4	
.	.	.	.	.Q.	.N.	.	.V.	.F.	.E.	.	.	.Q.	.I.	.L.	.G.	.	.	N*00901 A6	
.	.RM.	.	.	.	.	.	.	.	.	.	.	.S.V.	.PS.DL.	.F.GANVI.	.A.	.CKR.	.G.G.	5*00101 A2	
.	.RM.	.	.	.H.	.I.	.	.F.	.	.	.S.	.R.	.	.S.P.	.PS.DL.	.F.GANVI.	.A.	.CKR.	5*00101 A3	5
.	.RM.	.	.	.H.	.I.	.	.F.	.	.	.S.	.R.	.	.S.P.	.PS.DL.	.F.GANVI.	.A.	.CKR.	5*00101 A9	
.	.K.	.S.	.	.S.	.A.	.	.HI.	.M.	.S.	.	.II.	.SHE.I.H.L.IV.	.	.LVTG.	.	.	.H.DG.G.	ECMHCA1	
.	.K.	.S.	.	.S.	.A.	.	.HI.	.M.	.S.	.	.II.	.SHE.I.H.L.IV.	.	.LVTG.	.	.	.H.DG.G.	6*00101 A3	
.	.K.	.S.	.	.S.	.A.	.	.HI.	.M.	.S.	.	.II.	.SHE.I.H.L.IV.	.	.LVTG.	.	.	.H.DG.G.	6*00101 A5	
.	.K.	.S.	.	.S.	.A.	.	.HI.	.M.	.S.	.	.II.	.SHE.I.H.L.IV.	.	.LVTG.	.	.	.H.DG.G.	6*00201 A9	6
.	.K.	.S.	.	.S.	.A.	.	.HI.	.M.	.S.	.	.II.	.SHE.I.H.L.IV.	.	.LVTG.	.	.	.H.DG.G.	6*00101 A10	
.	.K.	.S.	.	.S.	.A.	.	.HI.	.M.	.S.	.	.II.	.SHE.I.H.L.IV.	.	.LVTG.	.	.	.H.DG.G.	ECMHCC1	
.	.K.	.V.	.R.	.I.	.M.	.	.	.R.	.	.I.R.	.	.G.F.	.MVT.	.	.V.	.	.G.	7*00101 A2	
.	.K.	.V.	.R.	.I.	.M.	.	.	.R.	.	.I.R.	.	.G.F.	.MVT.	.	.V.	.	.G.	7*00201 A3	
.	.K.	.V.	.R.	.I.	.M.	.	.	.R.	.	.I.R.	.	.G.F.	.MVT.	.	.V.	.	.G.	7*00101 A5	
.	.K.	.V.	.R.	.I.	.M.	.	.	.R.	.	.I.R.	.	.G.F.	.MVT.	.	.V.	.	.G.	7*00101 A9	7
.	.K.	.V.	.R.	.I.	.M.	.	.	.R.	.	.I.R.	.	.G.F.	.MVT.	.	.V.	.	.G.	ECMHCE1	
.	.H.	.	.	.R.	.	.E.	.S.E.	.	.S.	.	.K.	.	.I.P.	.ICII.	.G.	.	.G.	ECMHCB3	

Figure 2. Predicted amino acid alignment of 50 horse MHC class I sequences

Classical loci are listed in numerical order, non-classical loci are grouped below. Allele names are followed by associated serological ELA haplotype and exclude the prefix 'Eqca'. The consensus sequence is shown at the top, dots '.' represent identities, dashes '-' represent gaps inserted to optimize the alignment, asterisks (*) indicate stop codon. Exon boundaries are indicated by vertical bars, and horizontal gray bars separate loci, listed on the right. Sequences are from Tables 1 and Supplementary Table 1, not all sequences are the same length.

Figure 3. Variability plots of horse MHC class I sequences

Variability plots of A) classical (group B) and B) non-classical (groups A, C, E) MHC class I amino acid sequences. Groups defined by Holmes and Ellis (1999). Only mature amino acid sequence was included and the variability index was determined as described by Wu & Kabat (1970).

Figure 4. Synonymous and non-synonymous nucleotide substitutions in horse MHC class I loci
 Mean number of synonymous (d_S) and non-synonymous (d_N) nucleotide substitutions per site in the antigen recognition site (ARS) and conserved (CONS) regions among alleles of equine MHC class I loci. d_S and d_N were estimated separately for the ARS and CONS regions of the coding sequence. Tests of hypothesis that $d_S = d_N$: * $p < 0.05$, ** $p < 0.001$. Groups A, B, C, and E were defined by Holmes and Ellis (1999).

Figure 5. Model of the genomic structure of equine MHC class I loci across ELA haplotypes

Schematic representation of the gene content of equine MHC class I loci in ten serologically defined ELA haplotypes. Boxes contain sequences obtained from individual horses and sequence names exclude the prefix 'Eqca' (Table 1). Columns represent class I loci identified from genomic clones and provisional loci (*) identified by phylogenetic clustering. Rows represent ELA haplotypes, sequences from ELA-A2, -A3, -A5, -A9, and -A10 haplotypes were determined from MHC homozygous horses (upper half of figure), remaining sequences were determined from MHC heterozygous horses (lower half of figure). Only alleles sequenced in this study are included. Actual location and order of loci is not known for haplotypes other than ELA-A3. Loci 1 and 16*, and 2 and 17*, are placed together because they contain the largest numbers of alleles of any of the horse MHC class I loci. Loci 1 and 16* are typical polymorphic classical loci, while 2 and 17* have atypical patterns of polymorphism. Amplification of locus 5, 6, and 7 was not performed for ELA haplotypes A1, A4, A6, A7, and W11.

Table 1

MHC class I sequences amplified from ten horse MHC class I haplotypes.

MHC class I Haplotype (ELA) ^a	Horse Acc No. ^b	Breed ^c	Allele Eqca prefix ^d	Associated Haplotype	Locus ^e	No. of clones	No. of horses ^f	GenBank Accession	Matches ^g
A2/A2	0834	TB	2*00201	A2	2	>4	2	DQ145607	
			5*00101	A2	5	4	1	DQ145611	
			7*00101	A2	7	4	1	DQ145616	
			16*00101	A2	I6*	>4	2	M95410	
			N*00101	A2	N	>4	2	M95409	
A3/A3	3474	TB	1*00101	A3	1	>4	2	DQ083407	
			2*00101	A3	2	>4	2	DQ083408	
			5*00101	A3	5	4	1	DQ083411	
			6*00101	A3	6	4	1	DQ083412	
			7*00201	A3	7	4	1	DQ083413	
A5/A5	1406	SB	2*00302	A5	2	4	1	DQ145608	
			6*00101	A5	6	4	1	DQ145613	
			7*00101	A5	7	4	1	DQ145617	
			16*00201	A5	I6*	>4	1	DQ145594	
			N*00201	A5	I7*	>4	1	DQ145593	
A7/A7	3832	Pony	1*00601	A7	1	>4	2	EF577072	X79890
			N*01001	A7	N	>4	2	DQ145602	X79892
A9/A9	1536	TB	1*00201	A9	1	4	1	DQ145597	
			2*00102	A9	2	4	1	DQ145609	
			5*00101	A9	5	4	1	DQ145612	
			6*00201	A9	6	4	1	DQ145614	
			7*00101	A9	7	4	1	DQ145618	
			16*00301	A9	I6*	4	1	DQ145598	
			N*00301	A9	I7*	4	1	DQ145595	
			N*00701	A9	N	4	1	DQ145596	

MHC class I Haplotype (ELA) ^a	Horse Acc No. ^b	Breed ^c	Allele Eqca prefix ^d	Associated Haplotype	Locus ^e	No. of clones	No. of horses ^f	GenBank Accession	Matches ^g
A10/A10	2446	SB	1*00301	A10	1	>4	2	DQ145600	
			2*00303	A10	2	4	1	DQ145610	
			6*00101	A10	6	4	1	DQ145615	
			N*00401	A10	I7*	>4	2	DQ145601	
			N*00501	A10	I8*	4	1	DQ145599	
A6/A10	0197	SB	1*00501	A6	1	>4	2	DQ145606	AY176099
			N*00901	A6	N	>4	2	DQ145604	AY176109
			1*00301	A10	1	>4	2	DQ145600	
			N*00401	A10	I7*	>4	2	DQ145601	
A6/A7	2470	Pony	1*00501	A6	1	>4	2	DQ145606	AY176099
			16*00401	A6	I6*	4	1	DQ145605	
			N*00901	A6	N	>4	2	DQ145604	AY176109
			2*00401	A7	2	2	2	EF577071	
			16*00501	A7	I6*	>4	2	DQ145603	
A7/A19	2590	TB	2*00401	A7	2	4	2	EF577071	
			16*00501	A7	I6*	>4	2	DQ145603	
A1/A4	3824	Arabian	2*00202	A1	2	2	2	EF577062	
			N*00302	A1	I7*	3	2	EF577063	
			N*00601	A1	I8*	2	2	EF577064	AY374512
			1*00401	A4	1	>4	1	EF577065	AY225153
			2*00301	A4	2	>4	1	EF577066	AY225157
A1/W11	3823	Arabian	N*00801	A4	N	4	1	EF577067	AY225151
			2*00202	A1	2	2	2	EF577062	
			N*00302	A1	I7*	3	2	EF577063	
			N*00601	A1	I8*	2	2	EF577064	AY374512
			1*00701	W11	1	>4	1	EF577068	AY176092
			2*00304	W11	2	1	1	EF577069	

MHC class I Haplotype (ELA) ^a	Horse Acc No. ^b	Breed ^c	Allele Eqca prefix ^d	Associated Haplotype	Locus ^e	No. of clones	No. of horses ^f	GenBank Accession	Matches ^g
			16*00601	W11	I6*	>4	1	EF577070	

^aDetermined by serology;

^bPrimary donor horse used for MHC class I sequencing;

^cBreed of donor horse, TB is an abbreviation for Thoroughbred, SB is for Standardbred;

^dAll alleles were assigned the prefix Eqca;

^eN denotes a sequence that has not been assigned to a locus, for other loci see text and Figure 5;

^fMHC class I sequences were identified in two unrelated horses for most haplotypes;

^gGenbank Accession number of sequences identical to alleles identified in this study.