J Mol Evol (1999) 49:316–324

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary History of MHC Class I Genes in the Mammalian Order Perissodactyla

E.C. Holmes,[1] S.A. Ellis[2]

[1] The Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
[2] Institute for Animal Health, Compton, Nr Newbury, RG20 7NN, UK

Received: 17 November 1998 / Accepted: 7 April 1999

Abstract. We carried out an analysis of partial sequences from expressed major histocompatibility complex (MHC) class I genes isolated from a range of equid species and more distantly related members of the mammalian order Perissodactyla. Phylogenetic analysis revealed a minimum of six groups, five of which contained genes and alleles that are found in equid species and one group specific to the rhinoceros. Four of the groups contained only one, or very few sequences, indicating the presence of relatively nonpolymorphic loci, while another group contained the majority of the equid sequences identified. These data suggest that a diversification of MHC genes took place after the split between the Equidae and the Rhinocerotidae yet before the speciation events within the genus *Equus.*

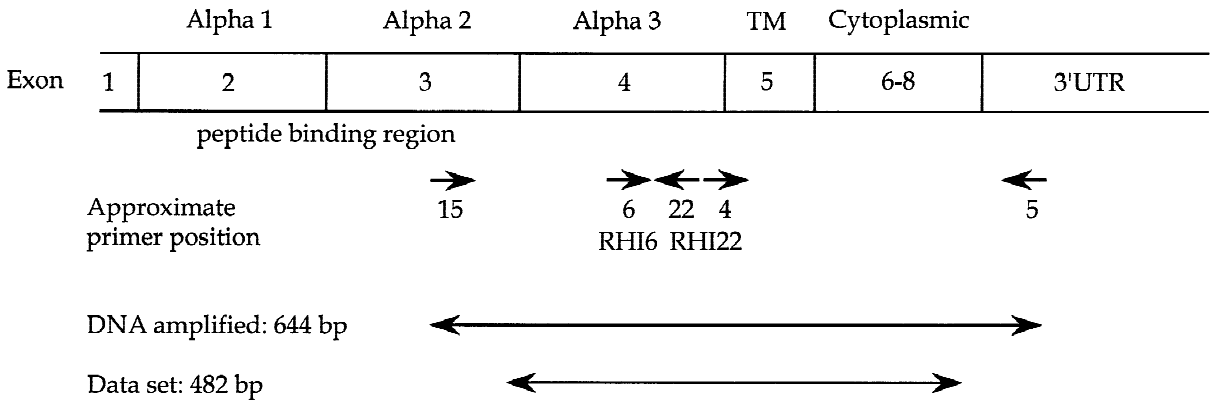
Key words: Major histocompatibility complex — Perissodactyls — Rhinoceros — Phylogeny — Polymorphism

Introduction

Major histocompatibility complex (MHC) class I molecules are highly polymorphic glycoproteins expressed in all nucleated cells in most vertebrates. Their main function is to present antigenic peptides derived from intracellular pathogens to cytotoxic T lymphocytes and thus elicit an immune response. The MHC class I region varies in size and complexity between mammalian species, most probably due to frequent expansions and contractions of that area of the genome (Delarbre et al. 1992; Vincek et al. 1987). In all examples studied to date (mostly primate and rodent), between one and three genes are expressed and have antigen presenting function—the classical class I, or class Ia, genes. All other genes in the region either are not expressed or have unknown or unrelated functions—the nonclassical class I, or class Ib, genes and pseudogenes (Trowsdale 1995).

Classical MHC class I genes have common characteristics relating to their function including high levels of polymorphism and expression, the ability to present peptides, and a conserved overall structure (Kaufman et al. 1994; Parham et al. 1995). The MHC is under evolutionary pressure to respond to pathogens which may themselves undergo relatively rapid changes due to immune selection pressure (Hedrick 1994; Hughes and Nei 1988). The genetic diversity which feeds this coevolutionary process is generated by both point mutation and also recombination, a process which may affect a few nucleotides to whole exons (Bontrop et al. 1995; Parham 1992; Parham and Ohta 1996). In humans there is little evidence of interlocus exchange, whereas in mice this happens frequently and may account for the lack of locus-specific characteristics in this case (Rada et al. 1990). Recombination among classical and nonclassical class I genes, as well as with pseudogenes, can also occur (Kaufman 1996).

Correspondence to: Dr. S.A. Ellis; *e-mail:* Shirley.Ellis@bbsrc.ac.uk



Oligonucleotide primers

	5' 3'
15:	TCCTGGACCGCTGCGGAC
5:	GTGAACAAATCTCGCATC
6:	CCTGCAGGGGACGGGACC
22:	GTGCTGCACATGGCACGTG
RHI6:	AGCTTGTGGAGACCAGG
RHI22:	GTATCTCTGCTCCTCTC
4:	CGTGCCGTGTGCAGCACG

Fig. 1. MHC Class I cDNA (coding region, approximately 1080 bp) showing the position of oligonucleotide primers used for PCR amplification and sequencing and the primer sequences used.

Nonclassical class I genes have a similar structure to classical genes although they are often truncated at the 3′ end, by either alternative splicing or the presence of early termination codons, which results in a short or missing cytoplasmic region (Fig. 1). Although most of these genes have not been assigned a function, some (such as *HLA-G, Qa-2*) have been shown to be capable of presenting peptides (Shawar et al. 1994). Nonclassical class I genes also exhibit much lower levels of polymorphism and often have a very restricted tissue distribution compared to their classical counterparts (Stroynowski 1990). It is possible that some of these less polymorphic genes may encode MHC molecules with binding motifs for particular conserved antigens (Bouwer et al. 1994), so that they are not under selection pressure to increase genetic diversity (Cheroutre et al. 1995).

It is clearly important to understand the evolutionary processes which have shaped MHC genetic diversity when considering the immune response of a species to a range of pathogens. A fruitful way this can be achieved is to study the phylogenetic relationships among MHC genes (and alleles) within and among different species. MHC alleles subject to some form of balancing selection survive for much longer periods of time than those which are fixed by genetic drift alone, and may even be maintained across species boundaries (Cooper et al. 1998; Figueroa et al. 1988; Takahata and Nei 1990). In contrast, positive selection for amino acid change is expected to produce genealogies with much more recent coalescent times (Hudson 1990), as can recent gene duplication events (Cadavid et al. 1997).

A previous study suggested that horses contain an unusually large number of apparently classical class I loci (Ellis et al. 1995). Phylogenetic analysis revealed that the majority of these sequences diverged within a relatively short time period from one another, such that the division between different loci and alleles was not detectable. To elucidate further the relationship among these genes and to place their evolution within a time frame, we have identified homologues of the horse MHC sequences in other members of the mammalian order Perissodactyla, including those from the distantly related Rhinocerotidae family. In the case of genes subject to such complex evolutionary processes as those in the MHC, examining the differing distributions of gene loci among species is perhaps the most profitable way to reconstruct the time scale of evolutionary change.

Materials and Methods

Animals. Peripheral blood was obtained from the following perissodactyl species: Przewalski's horse (*Equus ferus przewalskii*), donkey (*Equus asinus*), Persian onager (*Equus hemionus onager*), Grevy's zebra (*Equus grevyi*), Indian rhinoceros (*Rhinocerus unicornus*), Sumatran rhinoceros (*Dicerorhinus sumatrensis*), and Malayan tapir (*Tapi-*

Table 1. Details of animals and sequences used in the study

Species/breed	Name/No.	Sequence (GenBank accession No.)
Thoroughbred horse	Bear	8 and 9[a] (M95409), 1–29[a] (M9 5410)
Welsh pony	TFY	Al[a,c] (X71809), B1[a] (X79890), B2[a] (X79891), B3[a] (X80018), B4[a] (X79892), C[ac] (X79893), E1[ac] (X79894)
Przewalski's horse	L2052	PH1[b] (AJ133666)
	L1706	PH3[b] (AJ133667)
Donkey	BD	BD1[b] (AJ133663), BD2[b] (AJ133664), BD3[b] (AJ133665)
	WL	D1[b] (AJ133668)
Onager	L218	ON2[b] (AJ133669), ON3[b] (AJ133671), ON4[b] (AJ133672)
Grevy's zebra	L1671	GZ1[b] (AJ133673), GZ3[b] (AJ133674), GZ4[b] (AJ133675), GZ5[b] (AJ133676)
	L242	GZ6[b] (AJ133677), GZ7[b] (AJ133678)
Sumatran rhino	SR	SR1[b] (AJ133679), SR2[b] (AJ133680)
Indian rhino	Bardia	IR1[b] (AJ133681), IR2[b] (AJ133682), IR3[b] (AJ133683), IR4[b] (AJ133684), IR5[b] (AJ133670)
Malayan tapir		TAP1[b] (AJ133720)
Human		HLA-A1 (U07161), HLA-A2 (M24042), HLA-B7 (U04245), HLA-B38 (L22028), HLA-Cw2 (M24030)
Cow		BL3-7[a] (M21043), BL3-6[a] (M21044), KN104[a] (M69204), A10[a] (M69206)
Sheep		SHPMHCE (M34676)

[a] Full-length sequence derived by cDNA cloning.
[b] Partial sequence obtained by PCR amplification.
[c] Identical sequence found in several individuals.

rus indicus).[1] Peripheral blood mononuclear cells were isolated and stored as pellets at −70°C until required.

Polymerase Chain Reaction (PCR) and Sequencing. Messenger RNA (mRNA) was isolated from fresh or frozen lymphocyte pellets using a Dynal mRNA Direct kit (Dynal, Oslo, Norway), and first-strand cDNA was produced with a BRL cDNA synthesis system (BRL, Gaithersberg, MD, USA). Oligonucleotide primers were designed based on known horse MHC class I sequences (Ellis et al. 1995; Barbis et al. 1994), with the aim of amplifying the 3′ half (644 bp) of all class I sequences present. Figure 1 shows the position and details of the primers (5 and 15). PCR amplification was performed using a Programmable Thermal Controller (MJ Research, Boston, MA, USA) and was carried out in 100-μl reactions with 5 m*M* MgCl, a 0.5 mM concentration of each dNTP, a 10 μmol concentration of each primer, and 2.5 U of Taq DNA polymerase in 1× buffer (Promega, Southampton, UK). An annealing temperature of 50°C was used, followed by 30 cycles under the following conditions: 94°C (1 min), 50°C (1 min), and 72°C (1 min), with a last extension reaction at 72°C for 10 min. Amplified DNA was eluted from 1.5% low-melting point agarose, kinased, and blunt end ligated into *Sma*1-digested, phosphatased M13mp18. Complete sequencing of multiple clones was carried out using the chain termination method (Sanger et al. 1977) with a Sequenase 2.0 kit (U.S. Biochemicals, Cleveland, OH, USA). Sequence was obtained using the M13 primer, followed by internal primers 6 and 22 (depending on orientation of the insert), or rhino-specific RHI 6 and RHI 22 (Fig. 1). These primers and conditions were also used to amplify DNA from horses carrying previously sequenced class I genes, to ensure that the incidence of PCR-related artifacts was kept to a minimum. Tapir cDNA did not amplify with primers 5 and 15 and, in this case, a smaller region (206 bp) was amplified using primers 4 and 5.

All sequences generated in this study have been submitted to GenBank and assigned accession numbers AJ133663 to AJ133684 and AJ133720.

[1]Blood samples were taken from animals for medical diagnostic purposes in all cases, and excess (<10%) was used for experimentation. Sources as follows: Donkeys—TBA Equine Fertility Unit, Newmarket, Suffolk, UK; Sumatran rhinoceros and tapir—Howletts Zoo, Port Lympne, Kent, UK; and Onager, Grevy's zebra, Indian rhinoceros, and Przewalski's horse—Whipsnade Wild Animal Park, Dunstable, Bedfordshire, UK.

Data Composition. Because genetic diversity in the MHC is known to be maintained by natural selection (Hughes and Nei 1988; Takahata et al. 1992), we decided to avoid the α 1 and 2 domains of the class I genes, as this contains the peptide binding region where selection pressure is strongest (Fig. 1). The relatively conserved 3′ portion (α 3 to the 3′ UTR, exons 4 to 8) was therefore chosen as the most appropriate region to study the long-term evolution of the perissodactyl MHC. This region has previously been shown to be a good phylogenetic marker in studies of MHC loci from New World primates (Cadavid et al. 1997).

A total of 41 MHC class I sequences of aligned length 488 bp was used in the initial phylogenetic analysis. Of these, 31 came from perissodactyls (horse, Przewalski's horse, donkey, onager, zebra and two species of rhinoceros), 5 were from artiodactyls (cow and sheep), and 5 from humans. Twenty-two of the perissodactyl sequences were obtained by PCR as described, and nine by cDNA cloning (sequences previously published by Barbis et al. 1994; Ellis et al. 1995). The artiodactyl and human sequences were extracted from the GenBank database. A full list of the sequences used is given in Table 1 and alignments are available from the authors on request. To obtain a more precise estimate of the evolutionary relationships within the perissodactyl MHC, phylogenetic trees were also constructed on the 31 sequences (aligned length, 482 bp) from these species in isolation. Finally, a very short (aligned length, 206 bp) sequence was obtained from a Malayan tapir. This was included in a third phylogenetic analysis with the 31 perissodactyl sequences (trimmed to 206 bp). All sequence alignments were constructed by eye.

Sequence Analysis. MHC evolution is characterized by recombination between allelic and nonallelic sequences (Rada et al. 1990; Parham and Ohta 1996), so that it is incorrect to assume that genes always evolve in a strict tree-like manner. Consequently, an extensive search for possible recombinant events was undertaken prior to phylogenetic analysis. This was achieved first by reconstructing the evolutionary relationships between sequences using split decomposition (Bandelt and Dress 1992; Dopazo et al. 1992), a method which determines whether sequences are linked by multiple pathways (networks), as expected if recombination has taken place [program SplitsTree, version 2.4 (Huson 1998)] and then using a maximum-likelihood approach to detect regions with different evolutionary histories as might be produced under recombination. The latter analysis was performed with the PLATO program, version 2.0 (Grassly and Holmes 1997).

Phylogenetic trees between sequences were reconstructed using the maximum-likelihood method implemented in the 4.64d test version of

PAUP* kindly provided by David L. Swofford. The HKY85 model of DNA substitution was used with the transition:transversion ratio (Ts/Tv) and shape parameter (α) of a discrete approximation to a gamma distribution (with eight categories) of rate variation among sites estimated from the empirical data during tree reconstruction. The values of these parameters were 1.911 (Ts/Tv) and 0.635 (α) for the 41-sequence set, 2.024 (Ts/Tv) and 0.454 (α) for the 31-sequence set, and 1.683 (Ts/Tv) and 0.838 (α) for the analysis of the 31 sequences plus the short tapir sequence. To assess the robustness of the groupings obtained, a bootstrap analysis (1000 replications) was performed using the neighbor-joining method (with the same DNA substitution model as used in the maximum-likelihood analysis), again using the PAUP* package.

Results

Using both split decomposition (Bandelt and Dress 1992; Dopazo et al. 1992; Huson 1998) and a likelihood-based sliding window analysis (Grassly and Holmes 1997), we found no evidence for recombination in the perissodactyl MHC sequences presented here (data not shown, available from the authors on request). Any networked evolution observed was caused by scattered nucleotide substitutions, and all anomalously evolving regions (as detected by PLATO) were small in length, probably reflecting rate variation among sites. An analysis of nine full-length class I horse MHC sequences previously determined by us (Ellis et al. 1995) likewise provided no evidence of recombination.

Given that the sequences appeared to be evolving in a tree-like manner, we were then able to reconstruct their phylogenetic relationships. Six groups of perissodactyl MHC sequences, designated A to F, can be tentatively identified in both the 41 and the 31 sequence data sets (Figs. 2 and 3, respectively). These groups are defined as those which have strong bootstrap support (>70%) and are separated by long branches. Five of these groups (A to E) were previously found in horses (Ellis et al. 1995) and here we put their evolutionary history within an approximate time frame by examining their presence (or absence) in other perissodactyl species.

Group A contains five sequences from three equid species (Fig. 3). One of these sequences, A1 from horses, appears to represent a nonpolymorphic gene since the full-length sequence has been found in a number of unrelated individuals from different breeds which carry different serologically defined MHC class I haplotypes (data not shown). This gene also has a number of unusual features, including an early termination codon, resulting in a truncated cytoplasmic region of eight amino acids, and a number of amino acid substitutions at positions generally conserved in classical class I genes (Ellis et al. 1995). The predicted transmembrane (TM) length of the A1 gene is 40 amino acids. Interestingly, the only tapir class I sequence which amplified successfully (TAP1) also has a TM length of 40 amino acids, although a close relationship with group A was not confirmed in the phylogenetic analysis, where the tapir sequence appeared distinct, possibility reflecting the divergent position of the family Tapiridae (tree not shown, available from the authors on request). However, given the very short length of the tapir sequence it is difficult to assign its phylogenetic position with any certainty.

The remaining four group A sequences all have a high level of sequence similarity to A1—from 93.4 to 97.7%—but none have the early termination codon, indicating that this mutation was acquired only recently in A1. All three zebra sequences also have a three nucleotide deletion in exon 5, resulting in a predicted TM length of 39 amino acids, and sequences GZ4 and GZ5 were found in the same individual so that they may represent alleles of a single gene.

The largest group, designated B, contains 15 sequences from all 5 equid species available for analysis (Fig. 3). Aside from phylogenetic placing, all group B sequences have a TM length of 35 amino acids. The sequences in this group appear to be encoded by at least two genes since there are two sets of three sequences obtained from single animals (TFY and L218; see Table 1), and sequences 1–29 and 8–9 were cloned from a homozygous animal.

In contrast, the C and E groups each contain only a single horse sequence, C1 and E1, respectively, which have no clear phylogenetic relationship to any of the other perissodactyl sequences (as identified by the low bootstrap values and long branch lengths). These genes have subsequently been identified in a number of unrelated horses and appear to be nonpolymorphic (S.A. Ellis, unpublished results). C1 has a predicted TM length of 36 amino acids as well as a truncated cytoplasmic tail due to alternative splicing and hence lacks exon 7. E1 has a predicted TM length of 38 amino acids. The D group contains just two sequences (from horse and Grevy's zebra), but unlike C and E, these genes have not been identified in other individuals.

The final group, designated F, represents those sequences isolated from rhinos and has six or seven members, depending on the precise placement of sequence SR1 from a Sumatran rhinoceros. Although SR1 clusters with the other rhino sequences in the majority (59%) of the bootstrap replicates and has the same TM length, in does not fall in group F in either maximum-likelihood tree (Figs. 2 and 3). However, a (41 taxa) phylogeny in which SR1 is placed as a member of group F is not significantly different from the maximum-likelihood tree under the Kishino–Hasegawa test [difference in log likelihood = 1.658, $p = 0.899$ (Kishino and Hasegawa 1989)] so that the position of this sequence is unclear. Other, well supported, clusters of sequences are found within group F and sequences IR1 to IR5 were all found in a single animal, suggesting that there are at least three loci present which encode transcribed class I products.

While the six groups of class I sequences are clear-cut, the relationships among them are less certain, with

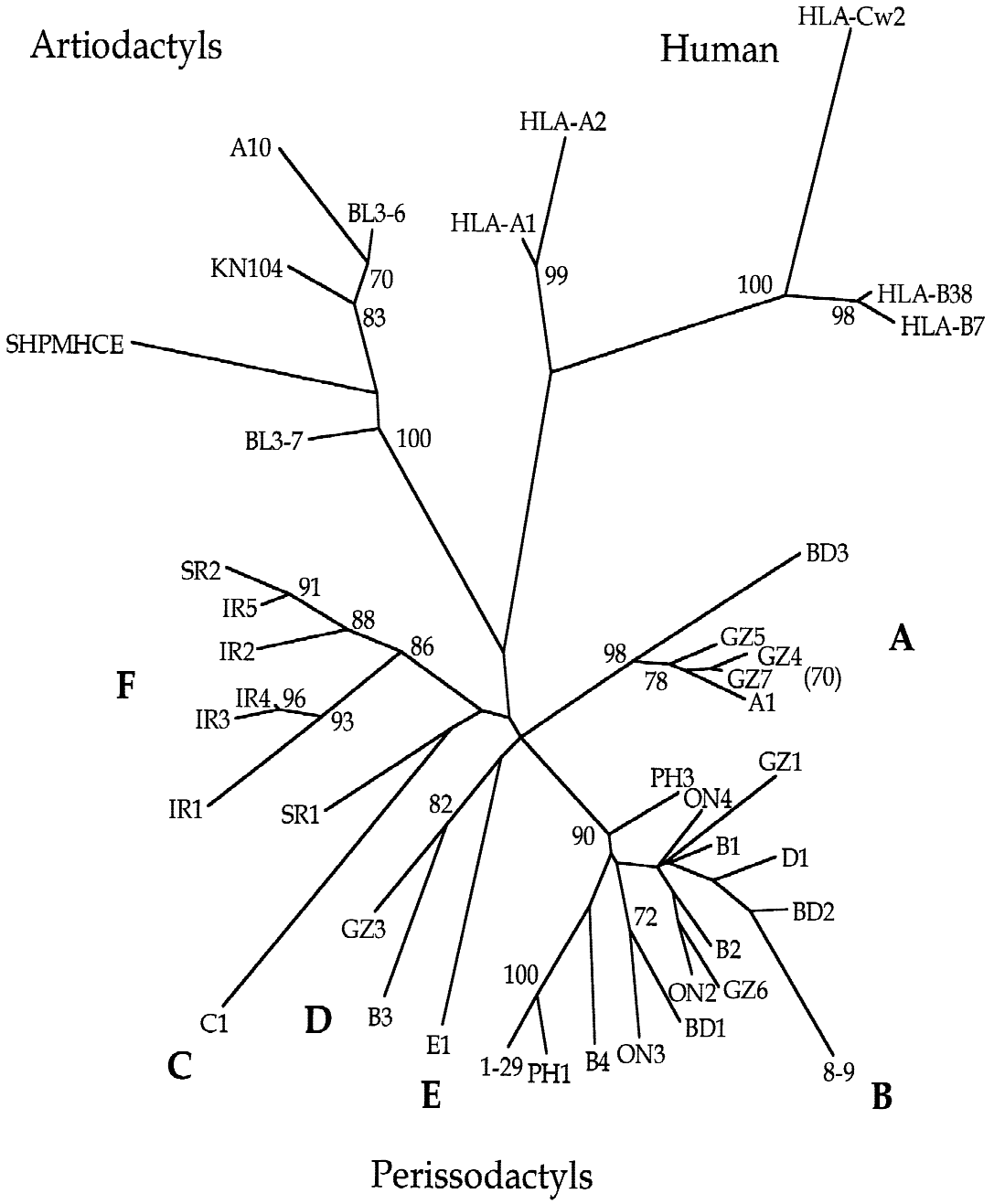


Fig. 2. Maximum-likelihood tree for 41 MHC sequences of artiodactyls, perissodactyls, and primates (humans). Numbers of supporting bootstrap replications (>70%), constructed using a neighbor-joining analysis with the maximum-likelihood substitution model, are shown next to the appropriate branches. All branch lengths drawn to scale.

all intergroup branch lengths short and only low levels of bootstrap support. Such a topology implies that the groups diverged within a relatively short time period and, because this occurs near the base of the tree, that this happened relatively early on in the evolutionary history of the perissodactyls. Interestingly, the monophyly of the perissodactyl sequences as a whole also does not receive particularly strong bootstrap support, meaning that these MHC genes either diverged shortly after the origin of the order, were polymorphic at the speciation event, or contain insufficient information to resolve phylogenetic relationships at this level. However, a lack of resolution seems unlikely given that a monophyletic origin of all perissodactyl MHC genes likewise received poor support in a previous analysis of full-length class I genes (Ellis et al. 1995).

Discussion

The MHC sequence data presented here show that the class I sequences from the Perissodactyla can be divided

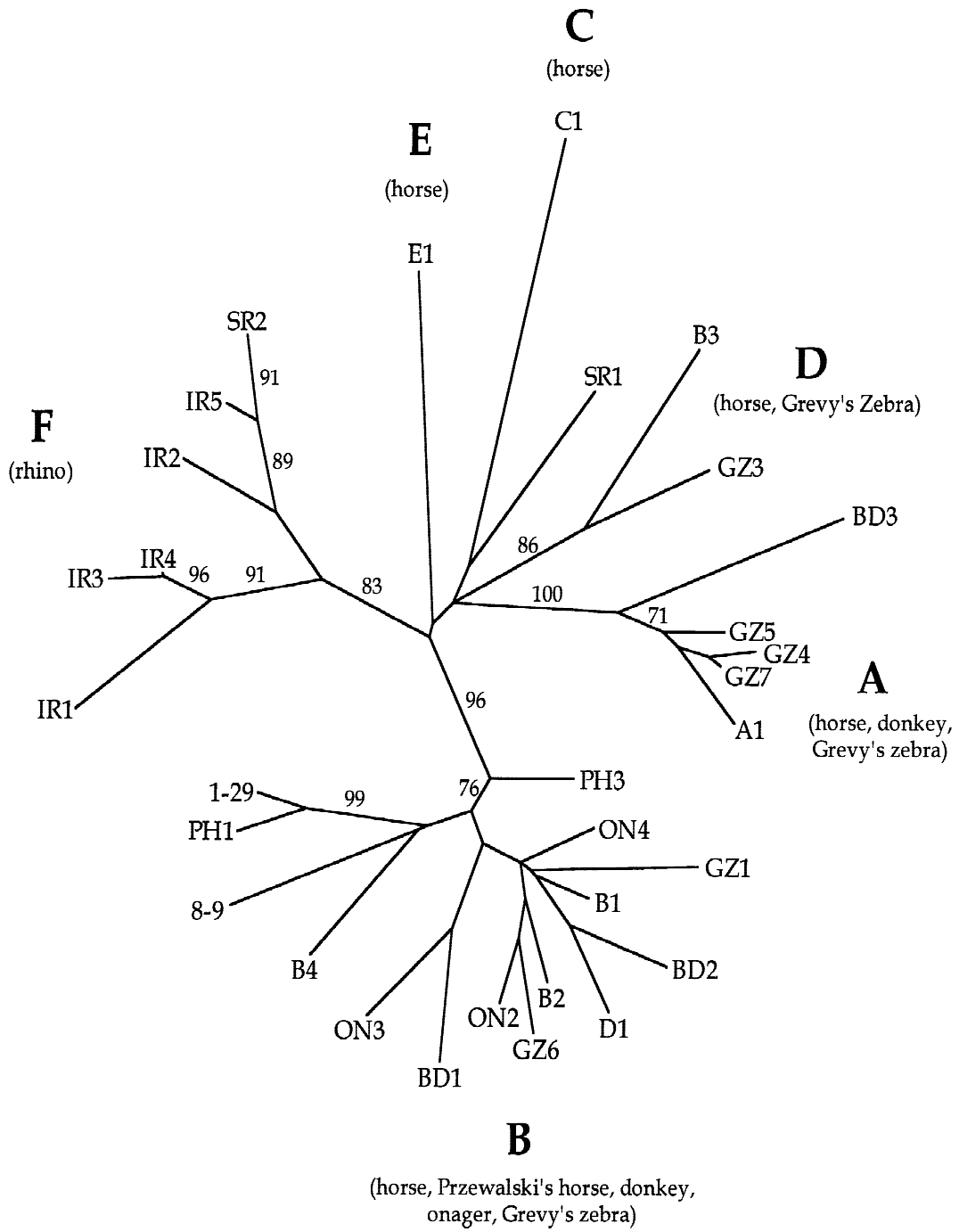


Fig. 3. Maximum-likelihood tree for the 31 MHC sequences from the perissodactyls. Numbers of supporting bootstrap replications (>70%), constructed using a neighbor-joining analysis with the maximum-likelihood substitution model, are shown next to the appropriate branches. All branch lengths drawn to scale.

into a number of distinct phylogenetic groups, some of which contain sequences found in a variety of equid species, but with one group specific to the rhinoceros. Since all genes included in the study were detectable in cDNA, it may be that most, if not all, are expressed and hence functional.

A previous study on a small number of horse class I sequences (Ellis et al. 1995) tentatively assigned them to at least four loci on the basis of transmembrane (TM) domain length, and these generally correspond to the groupings reported here. Group A sequences have 40 or 39 amino acid TM domains, group B sequences have 35 residue TM domains, as does group D, while groups C and E (with only one sequence each) have TM domain lengths of 36 and 38 amino acids, respectively, and group F sequences all have 38-amino acid-long TM domains. However, because some of the groups have equivalent TM lengths (groups B and D), and there is

some intragroup variation (group A) it is clear that this trait is not necessarily a good phylogenetic marker. Small variations in MHC class I gene TM length have been observed in other species, and it has been suggested that this area represents a hotspot for recombination and so is particularly susceptible to tandem duplications and deletions (Crew et al. 1991), although we found no evidence of these processes here.

The most ubiquitous genes in equids are those belonging to group B, with all individual animals studied containing two or more sequences from this group. Detailed analysis of both homozygous and heterozygous individuals reveals that the products of at least two loci are represented by this group (see Results), and it seems likely that these genes encode the functional, antigen presenting class I molecules. This is also supported by the observation that alloreactive T cells recognize the products of the 8–9 and 1–29 genes (D.F. Antczak, personal communication), and that a peptide binding motif has been identified for the gene B2 (data not shown), and because of the high levels of amino acid polymorphism seen in the peptide binding region of this group, indicative of immune-driven selection pressure (Ellis et al. 1995).

Despite their presumed functional importance, there is little phylogenetic structure within the group B sequences so that different loci and alleles at these loci cannot be easily distinguished. There are a number of possible explanations for this. It may be that a single functional gene has only recently duplicated, a process which has been used to explain the low levels of sequence variation observed in Callitrichid monkeys (Cadavid et al. 1997). Because group B sequences are found in all the five equid species studied here, such duplication events must have occurred prior to the origin of the extant members of this family (= genus) beginning 2–4 million years ago (MYA) (Ishida et al. 1995; Oakenfull and Clegg 1998). This may also apply, although to a lesser extent, to groups A and D, which likewise contain sequences from a variety of equid species, and always horse and zebra, and which therefore must have diverged before this speciation event.

Another possibility is that two (or more) much older genes are involved, whose alleles have become homogenized over time by recombination events. However, both split decomposition and a maximum-likelihood analysis failed to detect any evidence for recombination either in the partial MHC class I sequences presented here or in the full-length sequences determined previously. Whatever the explanation, it is clear that the Equidae contain a number of distinct groups of MHC sequences which must have existed for at least several million years.

In contrast to the diversity seen in group B, groups C and E each contain only one sequence, and only two sequences are found in group D. C1 is distinctive in lacking exon 7 and similarly truncated sequences have been detected in other equid species, but in each case there were problems in obtaining sufficient sequence so that they could not be included in the analysis presented here. C1 is also divergent throughout its coding region, with many amino acid substitutions appearing to be unique to this sequence. This implies that it has a long history of independent evolution, as also appears to be the case for E1. The small amount of data available further suggest that C1 and E1 may be nonpolymorphic or exhibit only limited polymorphism.

The A1 gene of horses (group A) also appears to be nonpolymorphic and this feature, along with a truncated cytoplasmic tail, suggests that it may be a nonclassical gene. However, because A1 is transcribed at unusually high levels in lymphocytes (Ellis et al. 1995), and synthesis of protein has been demonstrated (although no data are available regarding surface expression levels), it appears to be a gene of functional importance. Indeed, both A1 and C1 cDNAs have been transfected and shown to encode stable surface expressed molecules which bind peptide (data not shown). Furthermore, although A1 and C1 are nonpolymorphic both these genes, along with E1, B3 (from group D), and five group B sequences (B2, B3, B4, 8 and 9, 1–29) were previously shown to have higher rates of nonsynonymous than synonymous substitution in the peptide binding region, indicating that they may all function to present peptide to T cells (Ellis et al. 1995). Finally, while both A1 and C1 have the threonine-to-serine amino acid replacement at position 143 in the α 2 domain, thought to be a marker of nonclassical class I genes, this change is also present in a number of classical class I sequences from cattle (Ellis et al. 1996) and so may not be diagnostic of gene function or origin in all species. The status of A1 and C1 are therefore ambiguous, apparently being both nonpolymorphic yet with classical class I features and, so, do not appear to fulfill the standard criteria for either class Ia or class Ib genes (Hughes et al. 1999).

Perhaps most interesting of all are the group F sequences from the rhinoceros which show no clear relationship to any of the equid sequences. Since five of this group were obtained from a single animal, at least three loci must be represented, although more extensive analysis of individual rhinos is necessary to confirm the number of genes present and/or transcribed. Moreover, the finding that the rhino sequences, with the possible exception of SR1, form a single phylogenetic group suggests that the divergence of these genes coincides with the separation of the Equidae and the Rhinocerotidae at approximately 50 MYA (Prothero and Schoch 1989), but before the separation of the Indian and Sumatran rhinoceros species. Although such an inference is supported by the early divergence of the F group in the tree presented in Fig. 2, the precise position of the F group is uncertain so that it is formally possible that genes found in extant equids were also present in the rhinoceros (and vice

versa) but were later deleted or replaced in gene conversion events. This is clearly an area for future investigation.

To conclude, our study reveals that the period following the divergence of the Equidae and Rhinocerotidae, but before the speciation events within the genus *Equus*, saw a genetic diversification within the MHC, such that a number of distinct groups of genes and alleles (which are transcribed/expressed) were produced. Since this time some of these groups, most notably C, D, and E, have acquired little allelic variation, suggesting that they may represent nonclassical loci and/or have a specific function, whereas others, especially B, have accumulated a great deal of intragroup diversity, which presumably has a functional significance. Overall such a complex evolutionary history has left the equids rich in MHC class I genetic diversity.

Acknowledgments. This work was supported by the Biological Sciences and Biotechnology Research Council, UK, the Equine Virology Research Foundation, the Royal Society, and the Wellcome Trust. We also thank A.J. Martin for excellent technical assistance, two anonymous referees for helpful comments, and all those involved in acquiring blood samples, in particular, Edmund Flach at Whipsnade Wild Animal Park.

References

Bandelt H-J, Dress A (1992) Split decomposition: A new and useful approach for the phylogenetic analysis of distance data. Mol Phylogenet Evol 1:242–252

Barbis D, Maher JK, Stanek J, Klaunberg BA, Antczak DF (1994) Horse cDNA clones encoding 2 MHC class I genes. Immunogenetics 40:163–167

Bontrop RE, Otting N, Slierendregt BL, Lanchbury JS (1995) Evolution of MHC polymorphism and T-cell receptor diversity in primates. Immunol Rev 143:33–62

Bouwer HGA, Lindahl KF, Baldridge JR, Wagner CR, Barry RA, Hinrichs DJ (1994) An H2-T MHC class Ib molecule presents *Listeria monocytogenes*-derived antigen to immune CD8+ CTL. J Immunol 152:5352–5360

Cadavid LF, Shufflebotham C, Ruiz FJ, Yeager M, Hughes AL, Watkins DI (1997) Evolutionary instability of the major histocompatibility complex class I loci in New World primates. Proc Natl Acad Sci USA 94:14536–14541

Cheroutre H, Holcombe HR, Tangri S, et al. (1995) Antigen-presenting function of the TL antigen and mouse CD1 molecules. Immunol Rev 147:31–52

Cooper S, Adams EJ, Spencer Wells R, Walher CM, Parham P (1998) A MHC class I allele shared by two species of chimpanzee. Immunogenetics 47:212–217

Crew MD, Filiposwky ME, Neshat MS, Smith S, Walford RL (1991) Transmembrane domain length variation in the evolution of MHC class I genes. Proc Natl Acad Sci USA 88:4666–4670

Delarbre C, Jaulin C, Kourilsky P, Gachelin G (1992) Evolution of the MHC: A hundred-fold amplification of the MHC genes in the African pygmy mouse *Nannomys setulosis.* Immunogenetics 37:29–38

Dopazo J, Dress A, von Haeseler A (1993) Split decomposition: A technique to analyse viral evolution. Proc Natl Acad Sci USA 90: 10320–10324

Ellis SA, Martin AJ, Holmes EC, Morrison WI (1995) At least four MHC class I genes are transcribed in the horse. Eur J Immunogenet 22:249–260

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

Figueroa F, Günther E, Klein J (1988) MHC polymorphism pre-dating speciation. Nature 355:265–267

Grassly NC, Holmes EC (1997) A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol Biol Evol 14:239–247

Hedrick PW (1994) Evolutionary genetics of the major histcompatibility complex. Am Nat 143:945–964

Hudson RR (1990) Gene genealogies and the coalescent process. In: Futuyma D, Antonovics J (eds) *Oxford surveys in evolutionary biology,* Vol 7. Oxford University Press, New York, pp 1–44

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Yeager M, Ten Elshof AE, Chorney MJ (1999) A new taxonomy of mammalian class I molecules. Immunol Today 20: 22–26

Huson DH (1998) Splits Tree: A program for analyzing and visualizing evolutionary data. Bioinformatics 14:68–73

Ishida N, Oyunsuren T, Mashima S, Mukoyama H, Saitou N (1995) Mitochondrial DNA sequences of various species of the genus *Equus* with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. J Mol Evol 41:180–188

Kaufman J (1996) Evolution of the major histocompatibility complex and MHC-like molecules. In: Browning M, McMichael AJ (eds) HLA and MHC: Genes, molecules and function. BIOS Scientific, Oxford, pp 1–21

Kaufman J, Salomonsen J, Flajnik M (1994) Evolutionary conservation of MHC class I and class II molecules—Different yet the same. Semin Immunol 6:411–424

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of the hominoidea. J Mol Evol 29: 170–179

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetic analysis, version 1.0. Pennsylvania State University, University Park

Oakenfull EA, Clegg JB (1998) Phylogenetic relationships within the genus *Equus* and the evolution of α and θ globin genes. J Mol Evol 47:772–783

Parham P (1992) Typing for class I HLA polymorphism: Past, present and future. Eur J Immunogenet 19:347–359

Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67–74

Parham P, Adams EJ, Arnett KL (1995) The origins of HLA-A, B, C polymorphisms. Immunol Rev 143:141–180

Prothero DR, Schoch RM (1989) Classification of the Perissodactyla. In: Prothero DR, Schoch RM (eds) The evolution of the Perissodactyla. Oxford University Press, New York, pp 530–537

Rada C, Lorenzi R, Powis SJ, van den Bogaerde J, Parham P, Harvard JC (1990) Concerted evolution of class I genes in the MHC of murine rodents. Proc Natl Acad Sci USA 74:5463–5469

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 87:2167–2171

Shawar SM, Vyas JM, Rodgers JR, Rich RR (1994) Antigen presentation by MHC class 1-B molecules. Annu Rev Immunol 12:839–880

Stroynowski I (1990) Molecules related to class I like MHC antigens. Annu Rev Immunol 8:501–530

Takahata N, Nei N (1990) Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124:967–978

Takahata N, Satta Y, Klein J (1992) Polymorphism and balancing selection at major histocompatibility complex loci. Genetics 130: 925–938

Trowsdale J (1995) Both man and beast: comparative organisation of MHC genes. Immunogenetics 41:1–17

Vincek V, Nizetic D, Golubic M, Figueroa F, Nevo E, Klein J (1987) Evolutionary expansion of the MHC class I loci in the mole rat, Spalax ehrenbergi. Mol Biol Evol 4:483–491