Published in final edited form as:

Immunogenetics. 2012 December ; 64(12): 915–933. doi:10.1007/s00251-012-0651-z.

Natural selection on marine carnivores elaborated a diverse family of classical *MHC class I* genes exhibiting haplotypic gene content variation and allelic polymorphism

John A. Hammond, Lisbeth A. Guethlein, Paul J. Norman, and Peter Parham

Department of Structural Biology, Stanford University School of Medicine, Fairchild D-159 299
Campus Drive West, Stanford, CA 94305, USA

John A. Hammond: john.hammond@iah.ac.uk; Lisbeth A. Guethlein: lisbeth.guethlein@stanford.edu

Abstract

Pinnipeds, marine carnivores, diverged from terrestrial carnivores ~45 million years ago, before their adaptation to marine environments. This lifestyle change exposed pinnipeds to different microbiota and pathogens, with probable impact on their *MHC class I* genes. Investigating this question, genomic sequences were determined for 71 *MHC class I* variants: 27 from harbor seal and 44 from gray seal. These variants form three *MHC class I* gene lineages, one comprising a pseudogene. The second, a candidate nonclassical *MHC class I* gene, comprises a nonpolymorphic transcribed gene related to dog *DLA-79* and giant panda *Aime-1906*. The third is the diversity lineage, which includes 62 of the 71 seal *MHC class I* variants. All are transcribed, and they minimally represent six harbor and 12 gray seal *MHC class I* genes. Besides species-specific differences in gene number, seal *MHC class I* haplotypes exhibit gene content variation and allelic polymorphism. Patterns of sequence variation, and of positions for positively selected sites, indicate the diversity lineage genes are the seals' classical *MHC class I* genes. Evidence that expansion of diversity lineage genes began before gray and harbor seals diverged is the presence in both species of two distinctive sublineages of diversity lineage genes. Pointing to further expansion following the divergence are the presence of species-specific genes and greater *MHC class I* diversity in gray seals than harbor seals. The elaboration of a complex variable family of classical *MHC class I* genes in pinnipeds contrasts with the single, highly polymorphic classical *MHC class I* gene of dog and giant panda, terrestrial carnivores.

Keywords

Pinniped; MHCclass I; Evolution; Polymorphism

Introduction

Major histocompatibility complex (MHC) class I molecules are central components of the immune systems in all jawed vertebrates (Flajnik and Kasahara 2010). Their principal function is to bind peptide antigens inside cells and present them on the cell surface to natural killer (NK) and CD8 T cell receptors (Marrack et al. 2008; Neefjes et al. 2011; Parham and Ohta 1996). Such interactions between lymphocyte receptors and *MHC class I*

ligands regulate aspects of both the innate and adaptive immune response to intracellular infections, particularly viral infections (Bashirova et al. 2011; Cheent and Khakoo 2009; Orange and Ballas 2006). The evolution of pathogens to evade, escape, and subvert host immunity imposes strong selection pressures on *MHC class I* genes. These combine with other evolutionary forces, notably genetic drift, to drive the rapid evolution of *MHC class I* alleles, genes, and haplotypes. As a consequence, *MHC class I* diversity, which focuses on exons 2 and 3 that encode sites of peptide-binding and lymphocyte receptor contact, can evolve to become species-specific (Adams and Parham 2001) or population-specific (Kelley et al. 2005).

At the levels of alleles, genes, haplotypes, and populations, the only species in which *MHC class I* variation has been studied to the depth and breadth where generalizations can be made with confidence is arguably the human species (<http://www.ebi.ac.uk/imgt/hla/> (Robinson et al. 2011), <http://www.allelefrequencies.net/> (Gonzalez-Galarza et al. 2011)). In the human *MHC* (the *HLA* complex), there are six functional *MHC class I* genes, all of which are fixed in the human genome. Three of these, *HLA-A*, *HLA-B*, and *HLA-C*, are highly polymorphic genes, for which thousands of allelic variants have been defined. Consequently, most individuals are *HLA-A*, *HLA-B*, and *HLA-C* heterozygotes, and they differ from each other in *HLA class I* type. This diversifies immune systems within human populations, spreading both risks and benefits and increasing the likelihood that any new, emerging pathogen will encounter effective immune defense from some fraction of the population (Trowsdale 2011). Contrasting with *HLA-A*, *HLA-B*, and *HLA-C*, the *HLA-E*, *HLA-F*, and *HLA-G* genes exhibit little variation, being essentially conserved in both structure and function. Because of the qualitative difference in their polymorphism, *HLA-A*, *HLA-B*, and *HLA-C* are collectively referred to as classical *MHC class I* genes (or *MHC class Ia* genes), whereas *HLA-E*, *HLA-F*, and *HLA-G* are referred to as nonclassical *MHC class I* genes (or *MHC class Ib* genes). In addition to their low polymorphism, nonclassical class I genes tend to be older and more divergent than their classical counterparts; they also embrace a wider range of more specialized functions and can be expressed by fewer types of cells (Hofstetter et al. 2011; Rodgers and Cook 2005). However, these tendencies vary with the species and the gene. For example, human *HLA-G* is expressed only by extra-villous trophoblast, whereas *HLA-E* has a ubiquitous tissue distribution like *HLA-A*, *HLA-B*, and *HLA-C* (Moffett and Loke 2006).

Although the classical/nonclassical distinction is clearly defined for the human species, where the functional *MHC class I* genes are all fixed, that is not the situation in many other mammalian species, where variation in *MHC class I* gene content makes a significant contribution to the overall genetic diversity. Such species include other primates (Boyson et al. 1996; Doxiadis et al. 2011; Kulski et al. 2004; Saito et al. 2012), rodents (Rada et al. 1990), equids (Tallmadge et al. 2010), swine (Tanaka-Matsuda et al. 2009), and ruminants (Ballingall et al. 2008; Codner et al. 2012). For *MHC class I* genes that are present on only a subset of *MHC* haplotypes, the functional effect of the difference between the presence and absence of a gene will likely be greater than that achieved by alternative combinations of alleles of a fixed gene. Thus, gene content variation and allelic polymorphism can be seen to represent alternative, and even complementary, strategies for diversifying *MHC class I* haplotypes.

Having some similarity to their human counterparts, *MHC class I* haplotypes of domestic dog (a terrestrial carnivore) have a relatively fixed content of functional *MHC class I* genes (Kennedy et al. 2012; Wagner 2003; Wagner et al. 2005). A major difference in the dog is a translocation that split the *MHC class I* region between chromosomes 12 and 18 (Yuhki et al. 2007). The region on chromosome 12 that is syntenic to the *HLA* region (Breen et al. 1999; Mellersh et al. 2000) contains a highly polymorphic *MHC class I* gene (*DLA-88*) and

two nonpolymorphic *MHC class I* genes (*DLA-12* and *DLA-64*) (Burnett et al. 1997). On chromosome 18 is a moderately polymorphic *MHC class I* gene (*DLA-79*) (Burnett and Geraghty 1995; Graumann et al. 1998). The only known example of gene content variation is a duplication of the *DLA-88* gene, which is estimated to be a feature of ~10 % of *DLA* haplotypes, but can vary between breeds of dog (Kennedy et al. 2012; Ross et al. 2012). The three functional *MHC class I* genes of the giant panda (*Ailuropoda melanoleuca*), a dog-like carnivore, are related to the dog *MHC class I* genes. *Aime-1906* is moderately polymorphic and related to *DLA-79*. *Aime-128* and *Aime-152* are related to each other and to *DLA-12*, *DLA-64*, and *DLA-88*. *Aime-128* appears highly polymorphic and may be present in two copies on some haplotypes, whereas *Aime-152* appears conserved (Pan et al. 2008). By way of contrast, the domestic cat, representing the more distantly related cat-like carnivores, has a distinctive set of 12 *MHC class I* genes (Yuhki et al. 2008).

Pinnipeds are marine, dog-like carnivores that diverged from their terrestrial relatives ~45 million years ago (mya) in North America. Following the split, pinniped ancestors adapted to living in the marine environment (Arnason et al. 2006; Cao et al. 2000; Nery et al. 2012). Such a radical change in lifestyle involved adaptation to a different microbial universe and its pathogens, with the potential for consequent changes in pinniped *MHC class I and II* genes. Because *MHC class II* genes are phylogenetically more conserved, and their regions of variability are limited to a single exon, they have been more tractable to analysis than *MHC class I*. Indeed, *MHC class II* genes have been studied in pinniped populations for more than two decades, initially by analysis of restriction fragment length polymorphisms (Slade 1992) and subsequently by sequencing and high-resolution typing (Bowen et al. 2002, 2004, 2006; Cammen et al. 2011; Decker et al. 2002; Hoelzel et al. 1999; Weber et al. 2004).

In contrast, very little is known of pinniped *MHC class I*. First described was a cDNA sequence encoding an *MHC class I* polypeptide from a harbor seal of the Eastern Pacific subspecies (*Phoca vitulina richardii*) (Zhong et al. 1998). More recently, analysis of six Hawaiian monk seals (*Monachus schauinslandi*) identified six closely related *MHC class I* variants, each one coming from a different seal (Aldridge et al. 2006). Here we examine *MHC class I* gene diversity in two seal species from the North Sea: the Eastern Atlantic subspecies of the harbor seal (*Phoca vitulina vitulina*) and the gray seal (*Halichoerus grypus*). Monk seals, harbor seals, and the gray seal are all species of phocid pinniped (also known as true seals or earless seals), to be distinguished from the otarid seals (sea lions, also called fur seals, and the walrus). Harbor seal and gray seal are closely related and sympatric species that diverged recently from a common ancestor. Two estimates for the divergence time are ~5–7 mya (Arnason et al. 2006) and ~1–4 mya (Higdon et al. 2007).

Our purpose with this study was to perform a systematic analysis of the gray and harbor seal *MHC class I* gene families. Because each seal was to be investigated in considerable detail, the study was focused on relatively small panels of animals, consisting of eight harbor seals and four gray seals. However, the information obtained should provide sufficient knowledge and insight to facilitate future, broad-based population and epidemiological studies of *MHC class I* variation in pinniped populations.

Unlike *MHC class II*, which controls CD4 T cell immune responses, *MHC class I* is dedicated to immunity mediated by NK cells and CD8 T cells (Neefjes et al. 2011). These killer lymphocytes of innate and adaptive immunity provide crucial defense against viral infections, which are heavily influenced by *MHC class I* type as illustrated by the striking correlations of HLA class I polymorphism with the progress and outcome of human HIV infections (Bashirova et al. 2011). That recent epidemics of phocine distemper virus (PDV) have severely reduced European harbor seal populations, but not gray seal populations,

suggests that investigation of *MHC class I* variation in seal populations could lead to the identification of genetic factors that provide resistance or susceptibility to PDV (Dietz et al. 1989; Harkonen et al. 2006, 2007; Harris et al. 2008).

Materials and methods

Seal sample collection

Tissue from harbor seals was sampled at two sites: in 2003 at Abertay Sands (AB) on the east coast of Scotland in St. Andrews Bay (latitude 56° 23 00', longitude -2° 45 00') and in 2002 at The Wash (WSH) on the east coast of England in East Anglia (latitude 52° 92 00', longitude 0° 05 00'). Blood samples were obtained from four free-living harbor seals at Abertay Sands, whereas spleen samples were taken from four dead harbor seals discovered at The Wash. Blood samples were obtained from four gray seals, housed temporarily at the Sea Mammal Research Unit (University of St. Andrews, Scotland, UK), following capture at Abertay Sands: two individuals in 2003 and the other two individuals in 2004. All United Kingdom Home Office Procedures performed on the seals were conducted under license no. 60/3303, following the ethical review process at St. Andrews University.

Using microsatellite DNA polymorphisms as markers (Goodman 1998), it was shown that harbor seal populations are outbred and that there is movement of seals between them, resulting in admixture. The seals studied here have not been genotyped for markers other than *MHC class I*, so we have no independent assessment of their relatedness. For several reasons, including the *MHC class I* genotypes, it seems unlikely that any of the seals studied were closely related. The WSH harbor seals were randomly discovered animals that died of PDV infection. The AB harbor seals were randomly selected and caught. The gray seals all had different mothers, but we do not know if their parents were in any way related. The populations from which the seals derived are large and outbred; there is no evidence of family structure and the seals were caught or discovered outside of the breeding season, a time when family structure is not maintained.

RNA and DNA extraction

Seals were manually restrained, while blood was withdrawn from the extradural vein into EDTA vacutainers using the Vacutainer system (BD Biosciences, San Jose, CA, USA). Peripheral blood was decanted from the vacutainers, mixed 1:1 with RPMI 1640 medium, and overlaid onto Histopaque-1077 (Sigma-Aldrich, St. Louis, MO, USA) for separation of the mononuclear cells (PBMC) according to the manufacturer's guidelines. The layer of PBMCs was removed, washed twice in RPMI 1640, and used immediately for RNA extraction. PBMCs were homogenized in TriReagent™ (Sigma-Aldrich) and RNA extracted according to the manufacturer's guidelines. After assessing the quality of the RNA by agarose gel electrophoresis, cDNA was synthesized using Superscript III (Invitrogen, Carlsbad, CA, USA).

With removal of the PBMC layer from the Histopaque-1007, the pellet consisting of red blood cells (RBC) and granulocytes (polymorphonuclear leukocytes) was used as the source for gDNA. The pellets were washed twice with 70 ml of RPMI 1640 (700 g for 10 min at room temperature). The pellet was resuspended, by swirling, in 40 ml of phosphate-buffered saline (PBS) containing 1 % saponin for 10 min to lyse RBC. After centrifugation at 2,000×g for 5 min at 4 °C, the granulocyte pellet was washed twice in PBS with centrifugation at 2,000×g for 5 min at 4 °C. The washed granulocytes were then resuspended in 10 ml of 100 mM NaCl, 10 mM Tris pH 7.5, 25 mM EDTA, containing 0.5 % SDS and 0.5 mg/ml proteinase K, and incubated at 50 °C for 6 h. The same solution was used to digest samples of spleen (10 mm²) obtained from the cadaveric, Wash harbor seals, but with

the time of digestion being overnight. From the proteinase K digest, gDNA was extracted with phenol/chloroform/isoamyl alcohol in a standard two-step manner. Extracted gDNA was then precipitated with ethanol and suspended in Tris-EDTA buffer.

PCR amplification and sequencing of seal *MHC class I* genes

From comparison of *MHC class I* cDNA sequences from a range of mammalian species, conserved regions were identified. PCR primers based on these conserved regions were used to amplify multiple gene fragments from the 5'- and 3'-ends of *MHC class I* genes from one harbor seal and one gray seal. Sequences determined for 48 of these amplification products were used to identify conserved sequences as priming sites that could potentially amplify all phocid seal *MHC class I* genes. Of the two primers chosen, one (5'-CTACGTGGACGACACGCAGTTC-3') corresponded to sequence at the 5'-end of exon 2, encoding the α_1 domain, and the other (5'-GGKCAKGGTGGGCMATCTC-3') corresponded to sequence at the 3'-end of exon 8, specifying the 3' UTR. When used together in PCR on gDNA, these two primers produced an amplicon of ~2.65 kb. Both primers corresponded to conserved regions of mammalian *MHC class I* and were shown to amplify all the phocid *MHC class I* genes recovered from the initial primer screen. The primer pair was also tested for its capacity to amplify cDNA from a second harbor seal and a second gray seal: this PCR produced an amplicon of 1,064 or 1,062 bp (only *Phvi-N*04*), results confirming the general applicability of these priming sites for amplification from both gDNA and cDNA.

A limitation of the above-described primer pair is placement of the 5'-primer in exon 2, which produces incomplete coverage of that exon and thus lack of sequence for the 5'-end of exon 2, in and beyond the priming site. To rectify this deficiency, we applied 5'-RACE to the genomic DNA of one harbor seal and one gray seal, for the purpose of identifying a conserved sequence for primer design in exon 1 or intron 1. This analysis used the SMART RACE system according to the manufacturer's guidelines (BD Biosciences). The results enabled the design of a degenerate sense primer (5'-GGCTCCCACTCCMTGARGT-3') that corresponded to sequence in intron 1 just upstream of exon 2. This primer was used in PCR with an antisense primer (5'-GTGGGGATGGGGGTCGTGAC-3') derived from a conserved sequence in intron 2. From gDNA, this second pair of primers amplified a segment from intron 1 to intron 2 that included all of exon 2. The 3'-end of this ~300-bp amplification product was designed to overlap the 5'-end of the 2.65-kb amplification product by 200 bp. This overlap allowed sequences obtained from the two amplifications to be matched, thus giving complete coding region sequences (exons 2–8) for mature seal *MHC class I* proteins.

PCR amplification of the 2.65-kb segment was performed on 200 ng of gDNA, using the Advantage 2 genomic polymerase mix (BD Biosciences). Following initial denaturation of 2 min duration, 28 cycles of amplification (94 °C for 25 s, 61 °C for 25 s, and 68 °C for 3 min) were performed, with a final extension of 15 min at 68 °C. PCR amplification of the 300-bp segment used the same conditions except the initial denaturation time was 3 min and the cycling denaturation time was 40 s.

PCR products from the 2.65-kb and 300-bp amplifications were cloned using the TOPO TA cloning kit (Invitrogen). For the 2.65-kb amplicon, 72 clones from each of two or three (when necessary to confirm sequences) independent PCR reactions were sequenced in their entirety from both directions by end-sequencing using vector primers followed by primer-walking. The clones from the 300-bp amplicon were sequenced using the M13F site in the vector. To extend the 5'-end sequence of individual 2.65 kb clones, the sequences of clones from the 300-bp amplicon were compared to the 2.65-kb sequences in the 200-bp region of overlap. Clones containing an identical overlapping sequence with a cloned 2.65-kb

sequence were then sequenced from the other direction using the M13R primer to complete and confirm the sequence.

Nucleotide sequencing was performed on a Beckman Coulter CEQ 2000 instrument and the data assembled using the STADEN PACKAGE (Staden et al. 2000). Sequences were only used for further analysis once their validity was confirmed, either by them being identified in two independent experiments, or in two or more seals. All confirmed seal *MHC class I* sequences are deposited in GenBank, their accession numbers being given in Table 1.

Recombination and phylogenetic analysis

Sequences were aligned using CLUSTAL X (Thompson et al. 1997) and manually edited, when necessary, using Bioedit version 7.0.5.3 (Hall 1999). Detection of recombinant sequences was performed using the Recombination Detection Program (RDP) 3.41 (Martin et al. 2010). Recombination signals were analyzed in aligned sequence datasets, for each species individually and the two species together, with a maximum accepted probability value of $P=0.05$. Neighbor-joining (NJ) and maximum likelihood (ML) phylogenetic analyses, the average p distance, and the codon-based Z test of selection were all calculated with MEGA version 4.1 (Tamura et al. 2007). The phylogenetic trees were constructed using either the Tamura–Nei method or the Tamura three-parameter model with pairwise/partial deletion comparison and 1,000 replicates. Construction of the ML trees also used the nearest neighbor interchange optimization method.

In the statistical analyses of the gray and harbor seal *MHC class I* genes, we made comparisons with *MHC class I* genes from other carnivore species, for which the sequences were determined by other investigators. These *MHC class I* sequences are listed here with their GenBank accession numbers, where available, and references to relevant publications: Eastern Pacific harbor seal (PLA-1: U888874) (Zhong et al. 1998); Six Hawaiian monk seal *MHC class I* sequences (Mosc1.1 through to Mosc1.6) were not available from GenBank, so we copied them from the publication (Aldridge et al. 2006); domestic dog (DLA-12:U55030, DLA-64:U55027, DLA-79:Z25418 and DLA-88:U55028) (Burnett et al. 1997; Burnett and Geraghty 1995; Mellersh et al. 2000; Wagner 2003; Wagner et al. 2005); domestic cat (FLA-A1:U07667, FLA-A10:U07668, FLA-A23:U07669, FLA-A24:U07670, FLA-B2:U0767, FLA-B9:U07672, FLA-X8:U07674, and FLA-X10:U07668) (Yuhki et al. 1989, 2007, 2008; Yuhki and O'Brien 1988, 1994); and giant panda (Aime-126:EU162661, Aime-128:EU162656, and Aime-152:EU162657) (Pan et al. 2008).

Nomenclature for seal *MHC class I* variants

Because mammalian *MHCs* can contain multiple polymorphic genes with divergent allelic lineages, it can be difficult to distinguish genes from alleles on the basis of sequence comparisons alone, and such assignments remain uncertain until confirmed, either by segregation patterns, whole-haplotype sequencing, or other independent forms of data. We therefore adopted a provisional nomenclature for the groups of seal *MHC class I* sequences, following precedent set by the Immuno Polymorphism Database (IPD-MHC; <http://www.ebi.ac.uk/ipd/>) (Robinson et al. 2010)).

For each seal species, the genomic *MHC class I* sequences were assigned to different groups according to their sequence similarity: 12 groups were defined for the harbor seal and 15 groups for the gray seal. Sequences assigned to the same group had >99.5 % sequence similarity in the combined intron sequences and >98.5 % sequence similarity in the combined intron and exon sequences. These conservative criteria were designed to prevent alleles from different genes being assigned to the same group. Consistent with the achievement of this goal, no seal had more than two variants in any of the groups. Thus,

each group is a candidate for being a discrete *MHC class I* gene. The harbor seal groups were named *Phvi-N*01-N*12* and the gray seal groups *Hagr-N*01-N*15*, in which the prefix N indicates they have not been officially designated as a locus, but are candidates for that status. For the three harbor seal groups (*Phvi-N*01*, *N*02*, and *N*03*) for which some seals have two variants, this provides corroborating evidence that they do represent discrete genes, likewise for eight gray seal groups (*Hagr-N*03*, *N*05*, *N*06*, *N*07*, *N*08*, *N*09*, *N*14*, and *N*15*). In addition, *Hagr-N*01* and *N*02* are almost certainly discrete genes because of their highly divergent sequences. Because of the conservative criteria used to define the groups, it is possible that for genes with divergent lineages of alleles (like human *HLA-B*), such lineages would be assigned to different groups. In that case, two or more groups would define a single gene. Groups of seal *MHC class I* variants for which their status as genes or allelic lineages is uncertain are *Phvi-N*04-N*12*, *Hagr-N*04*, and *N*10-N*13*.

Because the variants within a group are almost certainly alleles of the same gene, they have been distinguished with a nomenclature similar to that used for *HLA class I and II alleles* (Marsh et al. 2010). For example, the *Phvi-N*03:01* and *Phvi-N*03:02* alleles differ by nonsynonymous substitution in the coding region, whereas the *Phvi-N*03:01:01* and *Phvi-N*03:01:02* alleles differ by synonymous substitution in the coding region.

Selection analysis

Estimation of the ratio between the average rate of non-synonymous substitution and the average rate of synonymous substitution (dN/dS or alternatively ω) was performed by maximum likelihood (ML) using the F3 X 4 model of codon frequencies and implemented in PAML, v3.14 (Yang 1997). For the site analysis, the likelihood of tree topology was estimated using four site-specific models that allow the ω ratio to vary among codons, but for which the site-specific patterns are identical across all lineages. These four models, M1a (nearly neutral), M2a (positive selection), M7 (β), and M8 (β and ω) were implemented using the program codeml in the PAML package. To test for positive selection, a likelihood ratio test (LRT) was used to compare a model that does not allow $\omega > 1$ with a model that does. LRTs were performed on M1a vs M2a and M7 vs M8. The Bayes empirical Bayes approach was then used to identify the codons with $\omega > 1$ (Yang et al. 2005).

Calculation of pairwise differences

Two datasets were assembled: the first consisted of coding region sequences for the harbor and gray seal *MHC class I* sequences reported here; the second consisted of previously reported coding region sequences of human and chimpanzee *MHC class I* genes, deposited in curated, specialist databases. *HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, *HLA-F*, *HLA-G*, and *HLA-H* were obtained from IMGT-HLA (<http://www.ebi.ac.uk/imgt/hla/> (Robinson et al. 2011)). For the highly polymorphic *HLA-A*, *HLA-B*, and *HLA-C*, genes only one allele for each allotype group was included in the dataset, e.g., *HLA-A*01:01*, *HLA-A*02:01*, and *HLA-A*03:01*. Sequences for the chimpanzee *MHC class I* genes (*Patr-A*, *Patr-AL*, *Patr-B*, *Patr-C*, *Patr-E*, *Patr-F*, *Patr-G*, and *Patr-H*) were obtained from IPD-MHC (<http://www.ebi.ac.uk/ipd/> (Robinson et al. 2010)).

Pairwise distances for the sequences in each dataset were calculated in MEGA using pairwise deletion and p-dist (Tamura et al. 2007). Histograms were made by sorting the *p* distance values into bins corresponding to increments of 0.005 substitutions/site. The number of pairwise comparisons in each bin was calculated as a percentage of the total number of pairwise comparisons and plotted as a histogram.

Results

MHC class I diversity was studied in harbor seals and gray seals. The analysis of harbor seal will be described first, followed by analysis of the gray seal and then comparison of the two species.

Eleven groups of transcribed *MHC class I* variants in the harbor seal

MHC class I genes were isolated and sequenced from the gDNA of eight harbor seals. From individual seals, between 8 and 13 different *MHC class I* variants were defined. On the basis of nucleotide sequence similarity, each *MHC class I* variant was assigned to 1 of 12 groups: *Phvi-N*01–N*12*, each containing variants with >99.5 % sequence similarity in the intron sequences and >98.5 % sequence similarity for the combination of introns and exons (Fig. 1a).

Each of these groups represents a different candidate *MHC class I* gene. *Phvi-N*02P* was judged to be a pseudogene, because its sequence contains several deletions in exons 2 and 3 and lacks a donor splice site in intron 2 that links these two exons. The sequences of the other 11 candidate genes indicate that they are transcribed and translated to produce functional *MHC class I* molecules. To assess gene transcription, mRNA was isolated from the cells of Pv149 and Pv150, seals whose genotypes cover all 12 groups (Fig. 1a). Complementary DNA was made from the RNA, subjected to PCR, and then cloned and sequenced. No cDNA clones corresponded to *Phvi-N*02P*, consistent with it being a nonfunctional gene. In contrast, cDNA clones corresponding to all the other 11 *MHC class I* genes were obtained. Moreover, for both Pv149 and Pv150, cDNA clones corresponding to each *MHC class I* gene in the seal's genotype were identified. Thus, 11 of the 12 groups of harbor seal *MHC class I* genes were shown to be transcribed. By extrapolation from the patterns of transcription in Pv149 and Pv150, the numbers of transcribed *MHC class I* genes for the other six harbor seals were estimated (Fig. 1b).

Allelic polymorphism and variable *MHC class I* gene content in the harbor seal *MHC*

That individual seals can carry two variants of *Phvi-N*01*, *Phvi-N*03*, and *Phvi-N*02P* demonstrates that each of these three groups corresponds to a different *MHC class I* gene, and their variants represent different alleles. Because all 12 groups are represented in seal Pv149 (Fig. 1a), the other nine groups must represent a minimum of five and a maximum of nine transcribed *MHC class I* genes. Thus in total, there is a minimum of seven and a maximum of 11 transcribed harbor seal *MHC class I* genes, plus the pseudogene. Contrasting with Pv149, seal Pv1912 could have as few as five transcribed *class I* genes. Six groups (*Phvi-N*01*, *Phvi-N*02P*, *Phvi-N*03*, *Phvi-N*04*, *Phvi-N*07*, and *Phvi-N*10*) are present in all individuals, whereas the other six groups are variably represented. Thus, harbor seal *MHC* haplotypes are demonstrated to vary in *MHC class I* gene content.

Eight of the 11 groups of transcribed *MHC class I* genes exhibited allelic polymorphism, but for three of them, the substitutions are synonymous. Of the remaining five groups, three have two allotypes, one has three allotypes, and one has four allotypes (Fig. 1a). The four *Phvi-N*03* allotypes differ by between one and eight substitutions (Fig. 1c). The substitutions are at nine positions: two in the α_1 domain (positions 62 and 83) and seven in α_2 (positions 108, 114, 116, 147, 152, 155, and 156). All but one of these positions are predicted to influence the recognition of *MHC class I* by T cell and NK cell receptors (Bjorkman et al. 1987; Boyington and Sun 2002; Marrack et al. 2008; Saper et al. 1991; Vivian et al. 2011; Wang et al. 2002). The extent of the allelic differences and their location within the class I molecule are typical of classical *MHC class I* molecules.

The three *Phvi-N*04* allotypes differ by substitutions at one or two positions. The first is at TCR-contacting position 62 in the α_1 domain and the second at peptide-contacting position 147 in α_2 (Fig. 1b). Despite the presence of three alleles in the panel, each of the eight seals has only one *Phvi-N*04* allele (Fig. 1a). This is an unusual phenomenon, for which several explanations could be considered. Given the stringent criteria (>98.5 % sequence similarity) used for assigning variants to the same *N** group, it is possible that *Phvi-N*04* is one of two divergent allelic lineages of the same gene, which we assigned erroneously to different *N* groups. Because all eight seals have only one form of *Phvi-N*07* and *Phvi-N*10*, these two groups are both candidates for such a second lineage. Alternatively, the second lineage could have been missed completely by our method for capturing *MHC class I* genes. A second possible explanation for the phenomenon, is that each of the eight harbor seals has one *MHC* haplotype containing the *Phvi-N*04* gene, combined with a second *MHC* haplotype that lacks *Phvi-N*04*. A third possibility is that all eight seals are homozygous for *Phvi-N*04* alleles, an improbable occurrence by chance, but a possible consequence of selection. A final and fourth possibility is that the *Phvi-N*04* gene was translocated to the X chromosome, so the six male seals in the panel (Fig. 1b) could only express one *Phvi-N*04* allele, and for the two females to be homozygous for *Phvi-N*04* alleles. Precedent for this latter mechanism is translocation of the mouse killer cell immunoglobulin-like receptors from the leukocyte receptor complex on chromosome 7 to the X chromosome (Welch et al. 2003).

Phvi-N*01 is a recombinant related to a dog and giant panda *MHC class I* gene

For phylogenetic analysis, the harbor seal *MHC class I* sequences were split into a 5'-segment that included exons 2 and 3, encoding the peptide-binding site, and a 3'-segment that includes exons encoding the α_3 , transmembrane, and cytoplasmic domains (Fig. 2). In trees obtained from the 5'-genomic segment (Fig. 2a), harbor seal *MHC class I* genes form three branches: the first containing the *Phvi-N*02P* pseudogene, the second containing the divergent transcribed gene *Phvi-N*01*, and the third containing the other ten transcribed genes, *Phvi-N*03-12*. In this 5'-segment, *Phvi-N*01* clusters strongly with dog *DLA-79* and giant panda *Aime-1906*. In contrast, this cluster is not reiterated in trees constructed from the 3'-segment sequences (Fig. 2b). Here, *DLA-79* and *Aime-1906* are at the base of the tree, whereas *Phvi-N*01* is situated well within the cluster formed by the transcribed harbor seal *MHC class I* genes and closest to *Phvi-N*08*. Similar analysis of coding region sequences, showed that *Phvi-N*01* again grouped with *DLA-79* and *Aime-1906* in the 5'-segment (Fig. 2c), but with the other transcribed harbor seal *MHC class I* genes in the 3'-segment (Fig. 2d). These results pointed to *Phvi-N*01* being a recombinant *MHC class I* gene, a conclusion supported by investigation of the carnivore *MHC class I* gene sequences using the recombination detection program RDP3 (Martin et al. 2010). Emerging from this analysis was strong statistical support for only one recombination: the event that formed *Phvi-N*01* ($P < 0.001$) from one gene resembling *DLA-79/Aime-1906* and a second resembling *Phvi-N*08* (Fig. 2e).

With the exception of the 5'-part of *Phvi-N*01*, the transcribed harbor seal *MHC class I* genes form a seal-specific gene family that has no single equivalent in dog, giant panda, or cat (Fig. 2). The relationship of the 5'-part of *Phvi-N*01* with dog *DLA-79* and giant panda *Aime-1906* indicates that the last common ancestor of seal, dog, and panda (~45 mya) had an ancestral *MHC class I* gene with a 5'-segment resembling that of modern *Phvi-N*01*, *DLA-79*, and *Aime-1906*. This gene was passed to the pinniped line, where it recombined with a member of the seal-specific family to produce the recombinant structure seen in *Phvi-N*01*. That no cat *MHC (FLA) class I* gene clusters with *DLA-79*, *Aime-1906*, and *Phvi-N*01* is consistent with the divergence of cat-like carnivores from dog-like carnivores ~52

mya, at a time ~7 mya before the separation of dog, giant panda, and seals (Arnason et al. 2006).

Fourteen groups of transcribed *MHC class I* variants in the gray seal

MHC class I genes were isolated and sequenced from the genomic DNA of four gray seals. From individual seals, between 11 and 18 different *MHC class I* variants were obtained. By applying the same criteria used to group harbor seal variants, each gray seal *MHC class I* variant was assigned to 1 of 15 groups: *Hagr-N*1-15* (Fig. 3a). Of these candidate *MHC class I* genes, the *Hagr-N*02P* pseudogene has the same deleterious exon 2 and 3 mutations as harbor seal *Phvi-N*02P*, as well as lacking the donor splice site in intron 2. *Hagr-N*02P* is clearly orthologous to *Phvi-N*02P*: they have 98.6 % sequence similarity compared to <84% sequence similarity with other seal *MHC class I* genes.

We were able to isolate and study only the 5'-segment of *Hagr-N*01*, indicating that its 3'-segment differs from that of *Phvi-N*01*. In the 5'-segment, *Hagr-N*01* exhibits >97.8 % sequence similarity with *Phvi-N*01*, showing the two 5'-segments are orthologous. Our failure to amplify the 3'-segment of *Hagr-N*01* by PCR could have a trivial explanation, such as localized variation within the priming sites. Alternatively, the 3'-segments of *Hagr-N*01* and *Phvi-N*01* could have substantially divergent sequences. The recombinant nature of *Phvi-N*01* (Fig. 2) raises the intriguing possibility that *Hagr-N*01* is not a recombinant and that both its 3'- and 5-segments resemble *DLA-79/Aime1906*. In that case *Hagr-N*01* would represent the ancestor of *Phvi-N*01*.

The sequences of the other 13 genes pointed to them being functional. To assess gene transcription, an analysis of *MHC class I* cDNA was made for gray seals HgQ and HgU1. The genotypes of these two individuals cover almost all the groups, exceptions being *Hagr-N*08* and *Hagr-N*10*. Complementary DNA clones corresponding to *Hagr-N*03*, *Hagr-N*04*, *Hagr-N*05*, *Hagr-N*06*, *Hagr-N*07*, *Hagr-N*09*, *Hagr-N*11*, *Hagr-N*12*, *Hagr-N*13*, *Hagr-N*14*, and *Hagr-N*15* were obtained and were those expected from each seal's genotype. In the absence of a complete genomic sequence, the functionality of *Hagr-N*01* was not tested. No cDNA corresponding to *Hagr-N*02P* was obtained, consistent with it being nonfunctional. By extrapolating from these results, we estimated the number of transcribed *MHC class I* genes in the other two gray seals (Fig. 3b).

Allelic polymorphism and variable *MHC class I* gene content in the gray seal *MHC*

That individual gray seals can carry two variants of the *Hagr-N*03*, *Hagr-N*05*, *Hagr-N*06*, *Hagr-N*07*, *Hagr-N*08*, *Hagr-N*09*, *Hagr-N*14*, and *Hagr-N*15* genes shows these eight groups represent different *MHC class I* genes (Fig. 3a). Because *Hagr-N*01* and *Hagr-N*02P* are divergent and orthologous to *Phvi-N*01* and *Phvi-N*02P*, respectively, it is very likely that these two also represent discrete genes. This leaves only *Hagr-N*04*, and *N*10-N*13*, as the remaining candidate genes. Minimally, this panel of gray seals has 13 *MHC class I* genes and pseudogenes, maximally it has 15. Six genes were represented in all four individuals, whereas nine groups exhibited presence/absence polymorphism. Thus, gray seal *MHC* haplotypes are demonstrated to vary in *MHC class I* gene content.

Polymorphism is particularly evident for the *Hagr-N*05*, *Hagr-N*06*, *Hagr-N*07*, and *Hagr-N*15* genes, each represented by three to five alleles. Contrasting with *Phvi-N*03* and *Phvi-N*04* (Fig. 1c), the variable residues of these four polymorphic gray seal *MHC class I* molecules do not concentrate at the peptide-binding site or known receptor-contacting positions of the α_1 and α_1 domains (Fig. 3c). This raises the possibility that the polymorphic *MHC class I* genes of gray and harbor seals have been subject to different selection

pressures. Alternatively, they may affect peptide binding and presentation in an indirect manner.

Gray and harbor seals have the same lineages and sublineages of *MHC class I* genes

Overall, the systems of *MHC class I* genes in harbor seal and gray seal are seen to be similar. They comprise the same three types of genes: a diverse seal-specific family of transcribed and likely functional *MHC class I* genes; a single and divergent transcribed gene related to a gene present in dog and giant panda, but not cat; and a divergent pseudogene (Fig. 4). These three groups represent three lineages of seal *MHC class I* genes. The lineage corresponding to the seal-specific expansion of transcribed *MHC class I* genes, we will call the diversity lineage.

Within the diversity lineage, some orthologous relationships between gray and harbor seal *MHC class I* genes could be discerned (Fig. 4). *Phvi-N*07* appears orthologous to *Hagr-N*07*, *Phvi-N*12* to *Hagr-N*12*, and *Phvi-N*06* to *Hagr-N*09*. Conversely, some genes appear specific to one of the species, harbor seal *Phvi-N*08*, and gray seal *Hagr-N*03*, *Hagr-N*04*, *Hagr-N*11*, and *Hagr-N*15*. Such assignments, however, can only be provisional because of the small sample sizes. Forming part of the diversity lineage are the Pacific harbor seal *MHC class I* variant (*PLA-1*) (Zhong et al. 1998) and the six Hawaiian monk seal *MHC class I* variants (*Mosc1.1–6*) (Aldridge et al. 2006). *PLA-1* is a distinctive member of the cluster formed by *Phvi-N*07* and *Hagr-N*07*, suggesting that it is orthologous to these two genes. The monk seal *MHC class I* sequences add two “new” branches to the tree, consistent with monk seals having diverged from harbor and gray seals 14–18 mya (Higdon et al. 2007).

Gray and harbor seal *MHC class I* genes of the diversity lineage further divide into two sublineages. Sublineage 1 comprises harbor seal *Phvi-N*05*, *Phvi-N*10*, and *Phvi-N*11* and gray seal *Hagr-N*05* and *Hagr-N*14*; sublineage 2 comprises all the other *MHC class I* groups in the diversity lineage (Fig. 4). Distinguishing sublineage 1 from sublineage 2 are substitutions at four positions in the α_1 domain and seven in the α_2 domain (Fig. 5a). These are sites that influence peptide binding (Bjorkman et al. 1987; Saper et al. 1991), TCR interaction (Marrack et al. 2008), and killer cell immunoglobulin-like receptor (KIR) interaction (Boyington and Sun 2002; Vivian et al. 2011), whereas the amino acid substitutions that distinguish the members of sublineage 1 are predominantly not at such sites (Fig. 5b). For sublineage 1, the surface contacting the T cell receptor has more basic and polar residues, compared to more acidic residues for sublineage 2. In addition, the variation at peptide-binding positions is more limited for sublineage 1 than sublineage 2. Monk seal *MHC class I* sequences also partition between the two sublineages: *Mosc1.2* and *1.3* in sublineage 1 and *Mosc1.1*, *1.4*, *1.5*, and *1.6* in sublineage 2 (Fig. 4). That both sublineages are represented in gray, harbor, and monk seals shows they existed prior to separation of monk seals, from harbor and gray seals, 14–18 mya (Higdon et al. 2007).

That harbor seals Pv149 and Pv150 have three members of sublineage 1 (*Phvi-N*05*, *Phvi-N*10*, and *Phvi-N*11*) shows that at least two different *MHC class I* genes contributed to the evolution of sublineage 1 in this species. At least two *MHC class I* genes also contributed to the evolution of sublineage 1 in gray seals because *Hagr-N*05* and *Hagr-N*14* represent different genes. The presence of this sublineage 1 in three seal species, with its preservation of a distinctive site for binding peptides and interacting with T cell receptors, points to sublineage 1 *MHC class I* molecules having a shared, distinctive, and beneficial function.

Application of the RDP3 program to the gray seal *MHC class I* sequences detected only one well-supported ($P < 0.01$) recombination event, from which the product was *Hagr-N*14:01*.

In exons 1–6 and introns 1–5, *Hagr-N*14:01* is almost identical to *Hagr-N*14:02*, but their sequences diverge at the 3′-end of the gene, in exon 7, intron 7, and exon 8. Here, *Hagr-N*14:01* is identical in sequence to *Hagr-N*04*. Thus, *Hagr-N*14:01* was formed by recombination between *Hagr-N*14* of sublineage 1 and *Hagr-N*04* of sublineage 2 (Fig. 4).

Positive selection on seal MHC class I has diversified peptide presentation

To investigate the effects of natural selection on the diversity lineage of seal *MHC class I*, we used pairwise comparisons to estimate the synonymous (dS) and nonsynonymous (dN) substitution rates. Both harbor and gray seal *MHC class I* have dN/dS ratios >1, evidence for positive diversifying selection. Comparison of the 5′-segment, encoding α_1 and α_2 domains, with the 3′-segment, encoding α_3 and the transmembrane domain, showed selection had focused on α_1 and α_2 (Fig. 6a). Correspondingly, a codon-based *Z* test significantly rejected ($P < 0.05$) the null hypothesis H_0 (dN = dS) in favor of H_1 (dN > dS) for full-length harbor seal *MHC class I* sequences. Significance in this analysis was due to the α_1 and α_2 domains ($P < 0.05$) because selection was not detected in the α_3 and transmembrane domains. For gray seal *MHC class I*, the *Z* test approaches significance ($P = 0.065$), but only when analyzing the α_1 and α_2 domains in isolation, a difference that could reflect the smaller gray seal dataset (Fig. 6a).

To identify sites under positive selection, ML methods using a substitution model were used (Yang 1997). The two models that allow $\omega > 1$ (M2a and M8) were shown to be significantly more likely than the corresponding models that do not, confirming the results from the *Z* test that positive diversifying selection has driven the variation in harbor seal and gray seal *MHC class I* (Fig. 6b). The selection tests, models 8 and 2a, identified 24 sites in the harbor seal and 28 sites in the gray seal to have been subject to positive selection ($P < 0.05$) (Fig. 6c, d). Within the α_1 and α_2 domains, almost all the residues identified are predicted to contact bound peptides, the $\alpha\beta$ T cell receptors, the KIR of NK cells, or the Ly49 NK cell receptors (Fig. 6c, d).

The four *Phvi-N*03* variants with nonsynonymous substitutions have a dN/dS ratio of 1.4, and ML methods showed that the models permitting $\omega > 1$ are significantly more likely than their equivalents that do not ($\alpha = 0.001$) (Fig. 6b). Of the eight positions that distinguish the *Phvi-N*03* variants, positions 114 and 152 have been subject to positive selection ($P > 0.95$), and both are predicted to contact peptide (Fig. 6c). In contrast, no significant evidence for selection was found for the three *Phvi-N*04* variants.

The diversity lineage of seal MHC class I genes appears equivalent to classical MHC class I genes of humans and chimpanzees

The distribution of pairwise differences between *MHC class I* genes is very similar for gray seal (Fig. 7a) harbor seal (Fig. 7b) and combination of the two (Fig. 7c). The four components to this distribution, with increasing distance, are comparison between alleles, orthologs, and within sublineage 1 of the diversity lineage; other comparisons within the diversity lineage; comparisons with *N*01*; and comparisons with *N*02P*. We then compared the seal distributions to those of *MHC class I* gene sequences from human and chimpanzee (Fig. 7d), hominid species that have been extensively characterized for *MHC class I* (IMGT-HLA; <http://www.ebi.ac.uk/imgt/hla/> (Robinson et al. 2011) and IPD-MHC; <http://www.ebi.ac.uk/ipd/> (Robinson et al. 2010)), and which have a divergence time of 7–10 mya (Chatterjee et al. 2009) comparable to the higher estimate for that for gray and harbor seals (Arnason et al. 2006).

The comparison of alleles within the same hominid classical *class I* gene (*A*, *B*, or *C*) gives a distribution (Fig. 7e) with comparable range (0.01–0.07) to that observed for genes in the seal

diversity lineage (Fig. 7c). In contrast, comparison of alleles from different hominid classical *class I* genes gives a distribution of differences (Fig. 7f) with a range (0.07–0.125) that is greater, and non-overlapping, than that for seal diversity lineage genes. The distribution of comparisons between a hominid classical gene and a hominid nonclassical gene (Fig. 7g) also has a greater (0.65–0.16), and non-overlapping, range compared to that for the seal diversity lineage. Within this range are the comparisons with seal *N*01*, pointing to this nonpolymorphic gene (Figs. 1 and 3) being a nonclassical *MHC class I* gene. The distribution of comparisons with the seal *N*02P* pseudogene has a range (0.185–0.21) that is higher, and non-overlapping, than that for all functional hominid *MHC class I* genes (Fig. 7d). This result shows that seal *MHC class I* genes are not inherently constrained to be closer in sequence similarity than hominid *MHC class I* genes.

From these comparisons, we see that the diversity generated by *MHC class I* genes of the seal diversity lineage is comparable to that of hominid classical *MHC class I* genes. That the seal diversity lineage consists of classical *MHC class I* genes is supported by the distribution of positively selected sites (Fig. 6), which strongly argues for the variation having been selected to change the binding and presentation of peptides to T cell and NK cell receptors. In seals, the evolution of a family of *MHC class I* genes that exhibits gene sequence and gene content variation appears to have achieved a similar functional endpoint as the smaller number of fixed and highly polymorphic *MHC class I* genes in humans and chimpanzees.

Discussion

We find the *MHC class I* genes of harbor and gray seals to have much in common, consistent with their recent divergence ~5–7 mya. Both species have the same three lineages of *MHC class I* genes. One lineage comprises a pseudogene, and the other two lineages comprise expressed, and likely functional, genes. Divergent *N*01*, which shares ancestry with dog *DLA-79* and giant panda *Aime-1906*, is the only representative of the first lineage of transcribed genes. In contrast, the second lineage of transcribed genes, the diversity lineage of seal *MHC class I* genes, is a gene family that has no counterpart in terrestrial carnivores (dog, giant panda, and cat) and consists of at least six genes in the harbor seal and 12 genes in the gray seal.

The divergent sequence of *N*01* and its lack of functional polymorphism are consistent with it being a nonclassical *MHC class I* gene. In contrast, the collective properties of the diversity lineage genes are characteristically the ones associated with classical *MHC class I* genes. They contribute to *MHC* haplotype variability in two ways: gene content variation and allelic polymorphism. This variation is founded on two sublineages of *MHC class I* genes, present in both harbor and gray seals, and distinguished by functional motifs in residues of the α_1 and α_2 domains that determine the presentation of peptide antigens to the antigen receptors of CD8 T cells and KIR NK cell receptors. The evolutionary process of positive selection that drove this bifurcation occurred prior to the divergence of harbor and gray seals and has continued following the separation of harbor and gray seal ancestors. As a consequence, individual seals vary in the number of *MHC class I* variants they express: six to ten variants among the harbor seals we studied and 9–16 variants among four gray seals studied. Such use of a combination of gene content variation and allelic polymorphism to generate *MHC* diversity appears to be a common feature of mammalian species (Ballingall et al. 2008; Boyson et al. 1996; Codner et al. 2012; Doxiadis et al. 2011; Kulski et al. 2004; Rada et al. 1990; Saito et al. 2012; Tallmadge et al. 2010; Tanaka-Matsuda et al. 2009), but is not the case in the human species, where gene content variability is absent and diversity arises purely from the allelic polymorphism of three fixed genes.

An obvious limitation to this exploratory investigation of *MHC class I* gene diversity in seals is the small number of seals studied. Consequently, all allele and gene frequency estimates can only be considered provisional, and they will undoubtedly be modified by future studies of larger populations. In addition, there is uncertainty for some of the seal *MHC class I* groups, whether they uniquely define a gene or just one of two constituent sets of alleles. Despite these limitations, we have obtained unambiguous evidence that allelic polymorphism is a mechanism that contributes to the diversity of seal *MHC class I* haplotypes. Thus, two or more alleles, encoding different protein variants, were characterized for three harbor seal and ten gray seal *MHC class I* groups in the diversity lineage. The allelic polymorphism of *Phvi-N*03* and *Phvi-N*04* has characteristics associated with selection for changes in the specificity of peptide presentation to CD8 T cells and in the interaction of MHC class I molecules with KIR NK cell receptors, as does the extended motif that distinguishes sublineages 1 and 2 of the diversity lineage.

Other aspects to the polymorphism, particularly for gray seal *MHC class I*, do not conform to this paradigm. Such differences may have arisen from other types of selection upon the NK cell response. For example, residue 134, which exhibits positive selection in the gray seal, is known to be a contact site for mouse Ly49 receptors (Wang et al. 2002). Three types of NK cell receptor (CD94:KKG2, Ly49, and KIR) recognize MHC class I and have alternatively been diversified in different mammalian species (Parham et al. 2010). In gray and harbor seals, Ly49 and KIR are each represented by a single functional gene that is highly conserved (Hammond et al. 2009). This raises the possibility that in these species, variation in MHC class I is exclusively used to diversify the interactions with these conserved NK cell receptors. Such a situation could produce patterns of MHC class I polymorphism that differ from those in species like humans and mice, where both the NK cell receptors and their MHC class I ligands are highly variable (Carlyle et al. 2008; Rajalingam 2012).

There are hints in our data for *MHC class I* differences between the Abertay Sands and Wash populations of harbor seals (Fig. 1a, b). Groups *Phvi-N*05* and *Phvi-N*06* *MHC class I* were present in all Abertay Sands seals, but absent from the Wash seals. In contrast, the Abertay Sands seals lacked *Phvi-N*03* polymorphism, this group being represented only by *Phvi-N*03:01*, whereas the Wash seals lacked *Phvi-N*03:01*, but had *Phvi-N*03:02*, *Phvi-N*03:03*, and *Phvi-N*03:04*.

The 1988 outbreak of PDV is estimated to have killed ~50 % of European harbor seals (Dietz et al. 1989). Characterizing this epidemic was a lower mortality around Scotland (~50 %) compared to eastern and southern England (~60 %), which could have a genetic basis (Goodman 1998). In 2002, a further epidemic of PDV in 2002 also killed ~50 % of the European harbor seals (Harkonen et al. 2006). Thus in recent time, European harbor seal populations have undergone strong selection from PDV. Survivors of the 2002 epidemic were the Abertay Sands seals we studied, whereas the Wash seals did not survive. It is therefore possible that the *Phvi-N*03*, *Phvi-N*05*, and *Phvi-N*06* differences between these two groups of harbor seals are a reflection of their relative resistance and susceptibility to PDV.

Gray seal populations were relatively unaffected by the PDV epidemics. Although being susceptible to PDV infection, gray seals remain asymptomatic and clear the virus rapidly. Circumstantial evidence suggests that healthy, PDV-infected gray seals actually contributed to the spread of PDV between distant populations of harbor seal (Hall et al. 2006; Hammond et al. 2005; Harkonen et al. 2006). That gray seals have a greater number of *MHC class I* genes, and more polymorphic *MHC class I* genes than harbor seals, could in general contribute to the resilience of their immune defense against PDV infection. In particular,

*Hagr-N*09*, which is polymorphic and present in all four gray seals, appears orthologous to *Phvi-N*06*, a characteristic of the four harbor seals in our study that survived the 2002 PDV epidemic. These speculations provide testable hypotheses for future population and epidemiological studies.

Acknowledgments

We thank the Sea Mammal Research Unit at the Scottish Oceans Institute for their generous help in collecting the samples. This work was supported by NIH grant AI24258 to P.P. We thank three anonymous reviewers of this paper for their constructive and insightful advice.

References

- Adams EJ, Parham P. Species-specific evolution of MHC class I genes in the higher primates. *Immunol Rev*. 2001; 183:41–64. [PubMed: 11782246]
- Aldridge BM, Bowen L, Smith BR, Antonelis GA, Gulland F, Stott JL. Paucity of class I MHC gene heterogeneity between individuals in the endangered Hawaiian monk seal population. *Immunogenetics*. 2006; 58:203–215. [PubMed: 16528500]
- Arnason A, Gullberg A, Janke A, Kullberg M, Lehman N, Petrov EA, Vainola R. Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Mol Phylogenet Evol*. 2006; 46:345–354. [PubMed: 16815048]
- Ballingall KT, Miltiadou D, Chai ZW, McLean K, Rocchi M, Yaga R, McKeever DJ. Genetic and proteomic analysis of the MHC class I repertoire from four ovine haplotypes. *Immunogenetics*. 2008; 60:177–184. [PubMed: 18270697]
- Bashirova AA, Thomas R, Carrington M. HLA/KIR restraint of HIV: surviving the fittest. *Annu Rev Immunol*. 2011; 29:295–317. [PubMed: 21219175]
- Beck TW, Menninger J, Murphy WJ, Nash WG, O'Brien SJ, Yuhki N. The feline major histocompatibility complex is rearranged by an inversion with a breakpoint in the distal class I region. *Immunogenetics*. 2005; 56:702–709. [PubMed: 15592824]
- Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*. 1987; 329:506–512. [PubMed: 3309677]
- Bowen L, Aldridge BM, Gulland F, Woo J, Van Bonn W, DeLong R, Stott JL, Johnson ML. Molecular characterization of expressed DQA and DQB genes in the California sea lion (*Zalophus californianus*). *Immunogenetics*. 2002; 54:332–347. [PubMed: 12185537]
- Bowen L, Aldridge BM, Gulland F, Van Bonn W, DeLong R, Melin S, Lowenstine LJ, Stott JL, Johnson ML. Class II multifunctionality generated by variable MHC-DRB region configurations in the California sea lion (*Zalophus californianus*). *Immunogenetics*. 2004; 56:12–27. [PubMed: 14997355]
- Bowen L, Aldridge BM, DeLong R, Melin S, Godinez C, Zavala A, Gulland F, Lowenstine L, Stott JL, Johnson ML. MHC gene configuration variation in geographically disparate populations of California sea lions (*Zalophus californianus*). *Mol Ecol*. 2006; 15:529–533. [PubMed: 16448418]
- Boyington JC, Sun PD. A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. *Mol Immunol*. 2002; 38:1007–1021. [PubMed: 11955593]
- Boyington JC, Motyka SA, Schuck P, Brooks AG, Sun PD. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature*. 2000; 405:537–543. [PubMed: 10850706]
- Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI. The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. *J Immunol*. 1996; 156:4656–4665. [PubMed: 8648109]
- Breen M, Thomas R, Binns MM, Carter NP, Langford CF. Reciprocal chromosome painting reveals detailed regions of conserved synteny between the karyotypes of the domestic dog (*Canis familiaris*) and human. *Genomics*. 1999; 61:145–155. [PubMed: 10534400]
- Burnett RC, Geraghty DE. Structure and expression of a divergent canine class I gene. *J Immunol*. 1995; 155:4278–4285. [PubMed: 7594586]

- Burnett RC, DeRose SA, Wagner JL, Storb R. Molecular analysis of six dog leukocyte antigen class I sequences including three complete genes, two truncated genes and one full-length processed gene. *Tissue Antigens*. 1997; 49:484–495. [PubMed: 9174141]
- Cammen K, Hoffman JI, Knapp LA, Harwood J, Amos W. Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*). *Mol Ecol*. 2011; 20:740–752. [PubMed: 21199032]
- Cao Y, Fujiwara M, Nikaido M, Okada N, Hasegawa M. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene*. 2000; 259:149–158. [PubMed: 11163972]
- Carlyle JR, Mesci A, Fine JH, Chen P, Belanger S, Tai LH, Makrigiannis AP. Evolution of the Ly49 and Nkrp1 recognition systems. *Semin Immunol*. 2008; 20:321–330. [PubMed: 18595730]
- Chatterjee HJ, Ho SY, Barnes I, Groves C. Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evol Biol*. 2009; 9:259. [PubMed: 19860891]
- Cheent K, Khakoo SI. Natural killer cells: integrating diversity with function. *Immunology*. 2009; 126:449–457. [PubMed: 19278418]
- Codner GF, Birch J, Hammond JA, Ellis SA. Constraints on haplotype structure and variable gene frequencies suggest a functional hierarchy within cattle MHC class I. *Immunogenetics*. 2012; 64:435–445. [PubMed: 22454046]
- Decker DJ, Stewart BS, Lehman N. Major histocompatibility complex class II DOA sequences from three Antarctic seal species verify stabilizing selection on the DO locus. *Tissue Antigens*. 2002; 60:534–538. [PubMed: 12542748]
- Dietz R, Heide-Jørgensen MP, Harkonen T. Mass deaths of harbor seals (*Phoca vitulina*) in Europe. *Ambio*. 1989; 18:258–264.
- Doxiadis GG, de Groot N, Otting N, Blokhuis JH, Bontrop RE. Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. *Immunogenetics*. 2011; 63:73–83. [PubMed: 20949353]
- Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet*. 2010; 11:47–59. [PubMed: 19997068]
- Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res*. 2011; 39:D913–D919. [PubMed: 21062830]
- Goodman SJ. Patterns of extensive genetic differentiation and variation among European harbor seals (*Phoca vitulina vitulina*) revealed using microsatellite DNA polymorphisms. *Mol Biol Evol*. 1998; 15:104–118. [PubMed: 9491609]
- Graumann MB, DeRose SA, Ostrander EA, Storb R. Polymorphism analysis of four canine MHC class I genes. *Tissue Antigens*. 1998; 51:374–381. [PubMed: 9583809]
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999; 41:95–98.
- Hall AJ, Jepson PD, Goodman SJ, Harkonen T. Phocine distemper virus in the North and European Seas—data and models, nature and nurture. *Biol Conserv*. 2006; 131:221–229.
- Hammond JA, Pomeroy PP, Hall AJ, Smith VJ. Identification and real-time PCR quantification of Phocine distemper virus from two colonies of Scottish grey seals in 2002. *J Gen Virol*. 2005; 86:2563–2567. [PubMed: 16099915]
- Hammond JA, Guethlein LA, Abi-Rached L, Moesta AK, Parham P. Evolution and survival of marine carnivores did not require a diversity of killer cell Ig-like receptors or Ly49 NK cell receptors. *J Immunol*. 2009; 182:3618–3627. [PubMed: 19265140]
- Harkonen T, Dietz R, Reijnders P, Teilmann J, Harding K, Hall A, Brasseur S, Siebert U, Goodman SJ, Jepson PD, Dau Rasmussen T, Thompson P. The 1988 and 2002 phocine distemper virus epidemics in European harbour seals. *Dis Aquat Organ*. 2006; 68:115–130. [PubMed: 16532603]
- Harkonen T, Harding K, Rasmussen TD, Teilmann J, Dietz R. Age- and sex-specific mortality patterns in an emerging wildlife epidemic: the phocine distemper in European harbour seals. *PLoS One*. 2007; 2:e887. [PubMed: 17849016]

- Harris CM, Travis JM, Harwood J. Evaluating the influence of epidemiological parameters and host ecology on the spread of phocine distemper virus through populations of harbour seals. *PLoS One*. 2008; 3:e2710. [PubMed: 18628992]
- Higdon JW, Bininda-Emonds OR, Beck RM, Ferguson SH. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol Biol*. 2007; 7:216. [PubMed: 17996107]
- Hoelzel AR, Stephens JC, O'Brien SJ. Molecular genetic diversity and evolution at the MHC DQB locus in four species of pinnipeds. *Mol Biol Evol*. 1999; 16:611–618. [PubMed: 10335654]
- Hofstetter AR, Sullivan LC, Lukacher AE, Brooks AG. Diverse roles of non-diverse molecules: MHC class Ib molecules in host defense and control of autoimmunity. *Curr Opin Immunol*. 2011; 23:104–110. [PubMed: 20970974]
- Kelley J, Walter L, Trowsdale J. Comparative genomics of major histocompatibility complexes. *Immunogenetics*. 2005; 56:683–695. [PubMed: 15605248]
- Kennedy, LJ.; Ollier, WER.; Marti, E.; Wagner, JL.; Storb, R. Canine immunogenetics. In: Ostrander, EA.; Ruvinsky, A., editors. *The genetics of the dog*. 2nd edn.. Oxon: CAB International; 2012.
- Kulski JK, Anzai T, Shiina T, Inoko H. Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. *Mol Biol Evol*. 2004; 21:2079–2091. [PubMed: 15269276]
- Marrack P, Scott-Browne JP, Dai S, Gapin L, Kappler JW. Evolutionarily conserved amino acids that control TCR–MHC interaction. *Annu Rev Immunol*. 2008; 26:171–203. [PubMed: 18304006]
- Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Fernandez-Vina M, Geraghty DE, Holdsworth R, Hurley CK, Lau M, Lee KW, Mach B, Maiers M, Mayr WR, Muller CR, Parham P, Petersdorf EW, Sasazuki T, Strominger JL, Svejgaard A, Terasaki PI, Tiercy JM, Trowsdale J. Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*. 2010; 75:291–455. [PubMed: 20356336]
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*. 2010; 26:2462–2463. [PubMed: 20798170]
- Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C, Jouquand S, Werner P, Andre C, DeRose S, Patterson DF, Ostrander EA, Galibert F. An integrated linkage-radiation hybrid map of the canine genome. *Mamm Genome*. 2000; 11:120–130. [PubMed: 10656926]
- Moffett A, Loke C. Immunology of placentation in eutherian mammals. *Nat Rev Immunol*. 2006; 6:584–594. [PubMed: 16868549]
- Neefjes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol*. 2011; 11:823–836. [PubMed: 22076556]
- Nery MF, Gonzalez DJ, Hoffmann FG, Opazo JC. Resolution of the laurasiatherian phylogeny: evidence from genomic data. *Mol Phylogenet Evol*. 2012; 64:685–689. [PubMed: 22560954]
- Orange JS, Ballas ZK. Natural killer cells in human health and disease. *Clin Immunol*. 2006; 118:1–10. [PubMed: 16337194]
- Pan H-J, Wan Q-H, Fang S-G. Molecular characterization of major histocompatibility complex class I genes from the giant panda (*Ailuropoda melanoleuca*). *Immunogenetics*. 2008; 60:185–193. [PubMed: 18292994]
- Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science*. 1996; 272:67–74. [PubMed: 8600539]
- Parham P, Abi-Rached L, Matevosyan L, Moesta AK, Norman PJ, Older Aguilar AM, Guethlein LA. Primate-specific regulation of natural killer cells. *J Med Primatol*. 2010; 39:194–212. [PubMed: 20618586]
- Rada C, Lorenzi R, Powis SJ, van den Bogaerde J, Parham P, Howard JC. Concerted evolution of class I genes in the major histocompatibility complex of murine rodents. *Proc Natl Acad Sci U S A*. 1990; 87:2167–2171. [PubMed: 2315309]
- Rajalingam R. Overview of the killer cell immunoglobulin-like receptor system. *Methods Mol Biol*. 2012; 882:391–414. [PubMed: 22665247]
- Robinson J, Mistry K, McWilliam H, Lopez R, Marsh SG. IPD —the Immuno Polymorphism Database. *Nucleic Acids Res*. 2010; 38:D863–D869. [PubMed: 19875415]

- Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. *Nucleic Acids Res.* 2011; 39:D1171–D1176. [PubMed: 21071412]
- Rodgers JR, Cook RG. MHC class Ib molecules bridge innate and acquired immunity. *Nat Rev Immunol.* 2005; 5:459–471. [PubMed: 15928678]
- Ross P, Buntzman AS, Vincent BG, Grover EN, Gojanovich GS, Collins EJ, Frelinger JA, Hess PR. Allelic diversity at the DLA-88 locus in Golden Retriever and Boxer breeds is limited. *Tissue Antigens.* 2012; 80(2):175–183. [PubMed: 22571293]
- Saito Y, Naruse TK, Akari H, Matano T, Kimura A. Diversity of MHC class I haplotypes in cynomolgus macaques. *Immunogenetics.* 2012; 64:131–141. [PubMed: 21881951]
- Saper MA, Bjorkman PJ, Wiley DC. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J Mol Biol.* 1991; 219:277–319. [PubMed: 2038058]
- Slade RW. Limited MHC polymorphism in the southern elephant seal: implications for MHC evolution and marine mammal population biology. *Proc Biol Sci.* 1992; 249:163–171. [PubMed: 1360677]
- Staden R, Beal KF, Bonfield JK. The Staden package, 1998. *Methods Mol Biol.* 2000; 132:115–130. [PubMed: 10547834]
- Tallmadge RL, Campbell JA, Miller DC, Antczak DF. Analysis of MHC class I genes across horse MHC haplotypes. *Immunogenetics.* 2010; 62:159–172. [PubMed: 20099063]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
- 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]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997; 25:4876–4882. [PubMed: 9396791]
- Trowsdale J. The MHC, disease and selection. *Immunol Lett.* 2011; 137:1–8. [PubMed: 21262263]
- Vivian JP, Duncan RC, Berry R, O'Connor GM, Reid HH, Beddoe T, Gras S, Saunders PM, Olshina MA, Widjaja JM, Harpur CM, Lin J, Malveste SM, Price DA, Lafont BA, McVicar DW, Clements CS, Brooks AG, Rossjohn J. Killer cell immunoglobulin-like receptor 3DL1-mediated recognition of human leukocyte antigen B. *Nature.* 2011; 479:401–405. [PubMed: 22020283]
- Wagner JL. Molecular organization of the canine major histocompatibility complex. *J Hered.* 2003; 94:23–26. [PubMed: 12692158]
- Wagner JL, Palti Y, DiDario D, Faraco J. Sequence of the canine major histocompatibility complex region containing non-classical class I genes. *Tissue Antigens.* 2005; 65:549–555. [PubMed: 15896203]
- Wang J, Whitman MC, Natarajan K, Tormo J, Mariuzza RA, Margulies DH. Binding of the natural killer cell inhibitory receptor Ly49A to its major histocompatibility complex class I ligand. *J Biol Chem.* 2002; 277:1433–1442. [PubMed: 11696552]
- Weber DS, Stewart BS, Schienman J, Lehman N. Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Mol Ecol.* 2004; 13:711–718. [PubMed: 14871373]
- Welch AY, Kasahara M, Spain LM. Identification of the mouse killer immunoglobulin-like receptor (Kirl) gene family mapping to chromosome X. *Immunogenetics.* 2003; 54:782–790. [PubMed: 12618911]
- Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 1997; 13:555–556. [PubMed: 9367129]
- Yang Z, Wong WS, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 2005; 22:1107–1118. [PubMed: 15689528]
- Yuhki N, O'Brien SJ. Molecular characterization and genetic mapping of class I and class II MHC genes of the domestic cat. *Immunogenetics.* 1988; 27:414–425. [PubMed: 2897330]
- Yuhki N, O'Brien SJ. Exchanges of short polymorphic DNA segments predating speciation in feline major histocompatibility complex class I genes. *J Mol Evol.* 1994; 39:22–33. [PubMed: 8064870]

- Yuhki N, Heidecker GF, O'Brien SJ. Characterization of MHC cDNA clones in the domestic cat. Diversity and evolution of class I genes. *J Immunol.* 1989; 142:3676–3682. [PubMed: 2715636]
- Yuhki N, Beck T, Stephens R, Neelam B, O'Brien SJ. Comparative genomic structure of human, dog, cat MHC: HLA, DLA, and FLA. *J Hered.* 2007; 98:390–399. [PubMed: 17675392]
- Yuhki N, Mullikin JC, Beck T, Stephens R, O'Brien SJ. Sequences, annotation and single nucleotide polymorphism of the major histocompatibility complex in the domestic cat. *PLoS One.* 2008; 3:e2674. [PubMed: 18629345]
- Zhong JF, Harvey JT, Boothby JT. Characterization of a harbor seal class I major histocompatibility complex cDNA clone. *Immunogenetics.* 1998; 48:422–424. [PubMed: 9799341]

\$watermark-text

\$watermark-text

\$watermark-text

a

Twelve groups of harbor seal MHC class I variants												
Animal ID	N*01:	N*02P:	N*03:	N*04:	N*05:	N*06:	N*07:	N*08:	N*09:	N*10:	N*11:	N*12:
Pv149*	01	01 & 02	01	01	01	01	01	01	01	01	01	01
Pv150*	01	01	01	01	01	01	01	01:02	01	01	01	
Pv151	01 & 01 [†]	01	01	01	02	01	01			01		
Pv152	01	01:02	01	01	01	01:02	01		01	01	02	02
Pv1912	01	01	02 & 04	03			01			01	01	
Pv1913	01	01	04	03			01		01	01	01:02	
Pv3541	01	01	03 & 04	01			01		01	01	01	
Pv55855	01	01	02 & 04:02	02			01		01	01	01	
Alleles	2	3	5	3	2	2	1	2	1	1	3	2
Allotypes	1	na	4	3	2	1	1	1	1	1	2	2

b

Animal ID	Sex	gDNA	mRNA	Groups	Site
Pv149*	M	13	11	12	AB
Pv150*	F	11	10	11	AB
Pv151	M	9	8	8	AB
Pv152	M	11	10	11	AB
Pv1912	M	8	7	7	WSH
Pv1913	F	8	7	8	WSH
Pv3541	?	9	8	8	WSH
Pv55855	M	9	8	8	WSH

c

Domain		α_1		α_2							
Residue		62	83	108	114	116	147	152	155	156	
N*03:	01	R	G	K	S	D	R	V	Q	D	
	02	.	.	.	.	.	W	.	.	.	
	03	.	S	.	.	.	W	W	R	H	
	04	L	.	N	E	F	W	W	R	H	
N*04:	01	L	G	N	S	D	R	V	Q	D	
	02	R	.	.	.	.	.	.	.	.	
	03	.	.	.	.	.	W	.	.	.	

Fig. 1.

The harbor seal has a diverse system of MHC class I genes. **a** Each column corresponds to 1 of 12 groups (Phvi-N*01–12) of *MHC class I* variant, predicted from sequence comparisons to represent different genes. Each of the first eight rows corresponds to an individual seal, with the presence of an MHC class I group indicated by *gray shading*. Variants within a group are distinguished by four digit no-menclature: for example, N*03:01 and N*03:02. To save space, the third and fourth digits are only included when needed to discriminate variants. Rows 9 and 10 give the numbers of alleles and allotypes for each group. *Dagger*, the two alleles of *Phvi-N*01* differ by a single synonymous nucleotide substitution in exon 5. The *dashed horizontal line* separates the Abertay Sands (AB) seals from the Wash (WSH) seals. **b** Under “gDNA” is given the number of *MHC class I* variants detected in the genome of each seal; under “mRNA” is the estimated number of transcribed genes for each seal. Under “Groups” is given the number of MHC class I groups represented in each animal. Under “Site” is given the geographical location (AB or WSH) where the blood or spleen sample from each animal was taken. **c** Amino acid sequence differences between Phvi-N*03 and Phvi-N*04 allotypes. *Presence of a dot* denotes identity with the N*03:01 or N*04:01 allotype. *Shaded gray* are positions predicted to contact bound peptide. Extrapolating from the human system, contacts could be made between residue 62 and TCR, and between residue 83 and the single seal killer cell immunoglobulin-like receptor (Hammond et al.

2009). Allotypic differences not shown in **c**: Phvi-N*05:02 differs from Phvi-N*05:01 by substitution of proline for alanine at position 185 in the α_3 domain; Phvi-N*11:02 differs from Phvi-N*11:01 by substitution of histidine for tyrosine at position 320 in the cytoplasmic tail; Phvi-N*12:02 differs from Phvi-N*12:01 by substitution of glutamine for glutamate at position 154 in the α_2 domain

\$watermark-text

\$watermark-text

\$watermark-text

Fig. 2. *Phvi-N*01* is a recombinant: the 5'-segment resembles dog *DLA-79* and giant panda *Aime-1906*, the 3'-segment resembles other transcribed seal *MHC class I*. **a-d** Phylogenetic trees constructed from carnivore *MHC class I* sequences. The recombinant *Phvi-N*01* gene taxa are boxed, whereas the likely parental genes are shown with gray shading (*Phvi-N*08*) or black shading with white lettering (*DLA-79*/ *Aime-1906*). **a** 5'-segment: comprising exon 2, intron 2, and exon 3. **b** 3'-segment: comprising exons 4–8, introns 4–7, and 3'-untranslated region. Intron 3 was not included in the analysis because it contains the recombination break point. **c** 5'-segment coding region sequence: comprising exons 2 and 3.

d 3'-segment coding region sequence: comprising exons 4–8 and 3'-untranslated region. *DLA*, *FLA*, and *Aime* denote dog (Burnett et al. 1997), cat (Beck et al. 2005), and giant panda (Pan et al. 2008) *MHC class I* genes, respectively. All other *MHC class I* taxa are from harbor seal for which the species designation, *Phvi*, is only given for *Phvi-N*01* and *Phvi-N*08*. "P" denotes pseudogene. NJ and ML analysis gave trees with similar topology. Shown are NJ trees, rooted at the midpoint; for nodes with bootstrap support >50 % the value of the support is shown. **e** Cartoon illustrating the structure of the *Phvi-N*01* recombinant, which has a 5'-segment resembling dog *DLA-79* (and *Aime-1906*) and a 3'-segment resembling *Phvi-N*08*. *Black-filled boxes with white lettering* denote exons with greater sequence similarity to their *DLA-79* and *Aime-1906* counterparts, whereas *gray-shaded boxes* denote exons with greater sequence similarity to their *Phvi-N*08* counterparts. The predicted site of recombination is within intron 3 between exons 3 and 4

Fig. 3.

The gray seal has a diverse system of *MHC class I* genes. Each column corresponds to 1 of 15 groups (Hagr-N*01–15) of MHC class I variant, predicted from sequence comparisons to represent different genes. Each of the first four rows corresponds to an individual seal, with the presence of an MHC class I group indicated by *gray shading*. Variants within a group are distinguished by four digit nomenclature, for which the third and fourth digits are only included when needed to discriminate variants. Rows 5 and 6 give the numbers of alleles and allotypes for each group. **b** Under “gDNA” is given the number of genes detected in the genome of each seal; under “mRNA” is the estimated number of transcribed genes for each

seal. Under “Groups” is the number of *MHC class I* groups represented in each seal. Under “Site” is given the geographical location (Abertay Sands) where the seals were captured. **c** Differences in the amino acid sequences of the N*05, N*06, N*07, and N*15 allotypes. *Presence of a dot* denotes sequence identity with the :01 allotype. *Shaded gray* are positions predicted to contact bound peptide. Allotypic differences not shown in **c**: Hagr-N*03:02 differs from Hagr-N*03:01 by substitution of threonine for isoleucine at position 124 in the α_2 domain; Hagr-N*08:01 differs from Hagr-N*08:02 by eight substitutions: six in α_1 (positions 43, 45, 60, 61, 69, and 70) and two in α_2 (positions 145 and 158); Hagr-N*09:02 differs from Hagr-N*09:01 by eight substitutions: three in α_1 (positions 21, 69, and 70) and five in α_2 (positions 108, 127, 134, 152, and 174); Hagr-N*09:03 differs from Hagr-N*09:01 by substitution of phenylalanine for valine at position 21 of α_1 ; Hagr-N*14:02 differs from Hagr-N*14:01 by substitution of serine for alanine at position 40 in α_1 , and two substitutions in the cytoplasmic domain (glycine for serine at position 328 and threonine for methionine at position 334) that are the consequence of recombination with Hagr-N*04

Fig. 4.

Lineages and sublineages of seal *MHC class I* genes. Phylogenetic trees were constructed from nucleotide sequences of the coding regions of *MHC class I* genes: NJ and ML analysis giving trees with similar topology. Shown is a NJ tree, which is rooted at the midpoint. For nodes with bootstrap support >50 %, the values of the support are shown. Species designations are Phvi (*harbor seal*); Hagr (*gray seal*), Mosc (Hawaiian monk seal) (Aldridge et al. 2006), PLA-A1 (Eastern Pacific harbor seal) (Zhong et al. 1998), domestic dog (DLA), giant panda (Aime), and domestic cat (FLA). Excluded from the analysis was exon 1 that encodes the leader peptide, which is not a part of the mature MHC class I molecule. The incomplete sequences included in the analysis had at least 50 % of the coding region

sequence. Only the sequence for the 5'-segment of *Phvi-N*01* and *Hagr-N*01* was used because the 3'-segment of *Phvi-N*01* is recombinant (see Fig. 2), and the 3'-segment of *Hagr-N*01* eluded our methodology. Pairwise sequence comparison showed *Mosc1.2* is in sublineage 1 of the diversity lineage, whereas *Mosc1.1*, *1.4*, *1.5*, and *1.6* are in sublineage 2. *Mosc1.3* is a recombinant, with exon 2 coming from sublineage 2, and exons 3–8 being like *Mosc1.2* of sublineage 1. The recombinant nature of *Mosc1.3* contributes ambiguity to the node separating the two groups of Hawaiian monk seal *MHC class I* genes

\$watermark-text

\$watermark-text

\$watermark-text

a		α_1				α_2							
Diversity lineage		24	69	79	80	97	150	161	162	166	167	169	
Sublineage 1		I	N	N	V	K	V	K	D	Q	S	R	
Sublineage 2		A	D	Q	T	L	A	E	G	E	W	G	
						W							
		G	S		I	R							
		E			N	E							
		I	G		V	V			D	Q			
Function		P	T K	T K	P T K	P	T K	T	T	T	P T		

b		α_1				α_2	α_3			Ct		
Sublineage 1		40	41	53	90	116	185	240	251	319	358	333
Phvi	N*05:01	A	A	E	A	Y	A	T	S	H	Q	T
	N*05:02	A	A	E	A	Y	P	T	S	H	Q	T
	N*10:01	S	V	E	A	W	P	T	S	H	Q	T
	N*11:01	S	V	E	D	W	P	T	S	Y	Q	T
	N*11:02	S	V	E	D	W	P	T	S	H	Q	T
Hagr	N*05:01	A	A	E	A	W	P	T	S	Y	Q	T
	N*05:02	A	A	E	A	W	P	T	P	Y	Q	T
	N*05:03	A	A	E	A	W	P	S	S	Y	Q	T
	N*05:04	A	A	G	A	W	P	S	S	Y	Q	T
	N*14:01	A	A	E	A	W	P	T	S	H	S	M
	N*14:02	S	A	E	A	W	P	T	S	H	Q	T
Function						P						

Fig. 5.

Sequence motifs at functional sites in the α_1 and α_2 domains distinguish two sublineages within the diversity lineage of seal MHC class I molecules. **a** Shown are the amino acid residues present at 11 positions distinguishing the two sublineages of diversity lineage MHC class I proteins. For sublineage 2, the relative frequencies of alternative amino acids at a position are given by the size of the *colored rectangles*. **b** Shown is the diversity of amino acid residues at positions that distinguish the members of sublineage 1. *Colors* denote the types of amino acid: aliphatic (I, L, V): *olive green*; basic (K, R): *dark blue*; acidic (D, E): *red*; polar (N, Q, S, T): *gray*; small hydrophobic (A): *light blue*; small polar (G): *gold*; bulky hydrophobic (W): *green*; hydrophobic (M): *light green*; aromatic hydrophobic (Y): *mid blue*;

aromatic charged (H): *pink*; proline (P): *purple*. Under “Function,” P denotes contact with peptides, T denotes contact with T cell receptors, and K denotes contact with killer cell immunoglobulin-like receptors (KIR), as deduced from structural and functional studies on other species, mainly human and mouse (Bjorkman et al. 1987; Boyington and Sun 2002; Marrack et al. 2008; Saper et al. 1991; Vivian et al. 2011; Wang et al. 2002)

\$watermark-text

\$watermark-text

\$watermark-text

Fig. 6.

Harbor and gray seal *MHC class I* genes of the diversity lineage have been subject to positive diversifying selection. **a** Comparison of mean dN and dS for exons 2–8 encoding the mature *MHC class I* protein, for exons 2 and 3 encoding the peptide binding site (α_1 and α_2 domains) and for exon 4 encoding the α_3 domain. The null hypothesis (H0) was tested against the alternative hypothesis (H1), using a one-tailed Z test. The significance level (α) is indicated when H0 was significantly rejected. **b** Shown are the results of likelihood ratio tests of positive selection, comparing site models without (M1a and M7) and with (M2a and M8) positive selection. Comparisons were performed for the harbor seal diversity lineage, *Phvi-N*03* alone, and the gray seal diversity lineage. Shown are the likelihood ratio test

statistic ($2\Delta I$) and the significance level (α) indicating the statistical support for positive selection. **c** Shown are positively selected codons identified by the Bayes empirical Bayes approach. *Black boxes* denote positively selected residues detected by models M8 and M2a and *light gray boxes* denote positively selected residues identified with model M8 only. Confidence levels are indicated by *asterisks*: * $P < 0.05$ and ** $P < 0.01$. Below each residue are the contacts it makes with peptides (P), $\alpha\beta$ T cell receptors (T), killer cell immunoglobulin-like NK cell receptors (K), and Ly49 NK cell receptors (Bjorkman et al. 1987; Boyington and Sun 2002; Marrack et al. 2008; Saper et al. 1991; Vivian et al. 2011; Wang et al. 2002). **d** The *ribbon diagrams* show positions of positive selection in the α_1 and α_2 domains of harbor seal (*left*) and gray seal (*right*) MHC class I molecules of the diversity lineage. *Yellow color* denotes a significance level of $P < 0.05$, and *red color* denotes a significance level of $P < 0.01$. The bound peptide is *colored blue*. The ribbon diagrams were derived from the crystal structure of HLA-C*03:04 (Boyington et al. 2000)

Fig. 7. Comparison of MHC class I diversity in closely related species of phocid seal and hominid. For various groups of MHC class I coding region sequences, the complete set of pairwise differences was calculated and the distribution of values plotted as a histogram. **a** Comparison of all gray seal *MHC class I* variants. **b** Comparison of all harbor seal *MHC class I* variants. **c** Comparison of the combined sets of gray and harbor seal *MHC class I* variants. The *horizontal lines* above the histogram denote the range of values for particular types of comparison: *I* between alleles, orthologs, and between diversity lineage genes of sublineage 1; *II* other comparisons between diversity lineage genes; *III* between *N*01* and

other genes, 5'-segment only; and *IV* between *N*02P* and other genes. **d** Comparison of the combined set of functional *MHC class I* variants from two hominid species: humans, *Homo sapiens* (*Hosa*) and chimpanzee, *Pan troglodytes* (*Patr*). The range values are for the following comparisons: *V* between alleles of the same gene, *VI* between alleles from different *classical class I* genes, *VII* between *classical* and *nonclassical class I* genes, and *VIII* between alleles from different *nonclassical class I* genes. **e** Comparison of differences between alleles of the three classical *MHC class I* genes of humans and chimpanzees. All comparisons are between alleles of the same gene, including interspecies comparisons. The results from the three genes are combined in the histogram. **f** Comparison of the differences between pairs of alleles from different hominid classical *MHC class I* genes. **g** Comparison of the differences between pairs of hominid *MHC class I* alleles, in which one allele comes from a classical *MHC class I* gene and the other comes from a nonclassical *MHC class I* gene. **h** Comparison of the differences between pairs of alleles from hominid nonclassical *MHC class I* genes. This includes pairwise differences between alleles from the same gene as well as those between alleles of different genes

Table 1

GenBank accession numbers

Allele	Accession number
Harbor seal MHC class I	
<i>Phvi-N*01:01:02</i>	JX218867
<i>Phvi-N*01:01:02</i>	JX218871
<i>Phvi-N*02:01:01P</i>	JX218872
<i>Phvi-N*02:01:02P</i>	JX218873
<i>Phvi-N*02:02P</i>	JX218874
<i>Phvi-N*03:01</i>	JX218878
<i>Phvi-N*03:02</i>	JX218879
<i>Phvi-N*03:03</i>	JX218880
<i>Phvi-N*03:04:01</i>	JX218881
<i>Phvi-N*03:04:02</i>	JX218882
<i>Phvi-N*04:01</i>	JX218883
<i>Phvi-N*04:02</i>	JX218884
<i>Phvi-N*04:03</i>	JX218885
<i>Phvi-N*05:01</i>	JX218886
<i>Phvi-N*05:02</i>	JX218887
<i>Phvi-N*06:01:01</i>	JX218894
<i>Phvi-N*06:01:02</i>	JX218895
<i>Phvi-N*07:01</i>	JX218896
<i>Phvi-N*08:01:01</i>	JX218875
<i>Phvi-N*08:01:02</i>	JX218876
<i>Phvi-N*09:01</i>	JX218877
<i>Phvi-N*10:01</i>	JX218888
<i>Phvi-N*11:01:01</i>	JX218889
<i>Phvi-N*11:01:02</i>	JX218890
<i>Phvi-N*11:02</i>	JX218891
<i>Phvi-N*12:01</i>	JX218893
<i>Phvi-N*12:02</i>	JX218892
Gray seal MHC class I	
<i>Hagr-N*01:01</i>	JX218870
<i>Hagr-N*01:02</i>	JX218868
<i>Hagr-N*01:03</i>	JX218869
<i>Hagr-N*02:01P</i>	JX218937
<i>Hagr-N*03:01:01</i>	JX218918
<i>Hagr-N*03:01:02</i>	JX218919
<i>Hagr-N*03:01:03</i>	JX218923
<i>Hagr-N*03:02</i>	JX218920
<i>Hagr-N*04:01</i>	JX218917
<i>Hagr-N*05:01</i>	JX218901

Allele	Accession number
<i>Hagr-N*05:02</i>	JX218902
<i>Hagr-N*05:03</i>	JX218903
<i>Hagr-N*05:04</i>	JX218904
<i>Hagr-N*05:05</i>	JX218916
<i>Hagr-N*06:01:01</i>	JX218907
<i>Hagr-N*06:01:02</i>	JX218909
<i>Hagr-N*06:02</i>	JX218910
<i>Hagr-N*06:03</i>	JX218908
<i>Hagr-N*07:01</i>	JX218898
<i>Hagr-N*07:02</i>	JX218899
<i>Hagr-N*07:03</i>	JX218900
<i>Hagr-N*07:04</i>	JX218897
<i>Hagr-N*08:01</i>	JX218922
<i>Hagr-N*08:02</i>	JX218906
<i>Hagr-N*09:01:01</i>	JX218924
<i>Hagr-N*09:01:02</i>	JX218925
<i>Hagr-N*09:01:03</i>	JX218929
<i>Hagr-N*09:02:01</i>	JX218926
<i>Hagr-N*09:02:02</i>	JX218928
<i>Hagr-N*09:03</i>	JX218927
<i>Hagr-N*10:01</i>	JX218933
<i>Hagr-N*11:01:01</i>	JX218930
<i>Hagr-N*11:01:02</i>	JX218931
<i>Hagr-N*11:01:03</i>	JX218932
<i>Hagr-N*12:01</i>	JX218911
<i>Hagr-N*12:02:01</i>	JX218912
<i>Hagr-N*12:02:02</i>	JX218913
<i>Hagr-N*13:01</i>	JX218935
<i>Hagr-N*14:01:01</i>	JX218914
<i>Hagr-N*14:01:02</i>	JX218905
<i>Hagr-N*14:02</i>	JX218915
<i>Hagr-N*15:01</i>	JX218934
<i>Hagr-N*15:02</i>	JX218921
<i>Hagr-N*15:03</i>	JX218936