

Rapid expansion of killer cell immunoglobulin-like receptor genes in primates and their coevolution with MHC Class I genes

Li Hao, Masatoshi Nei*

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park, PA 16802, USA

Received 9 September 2004; received in revised form 24 November 2004; accepted 2 December 2004

Received by H.E. Roman

Abstract

The gene family of killer cell immunoglobulin-like receptors (KIRs) in primates provides the first line of defense against virus infection and tumor transformation. Interacting with MHC class I molecules, KIRs can regulate the cytotoxic activity of natural killer (NK) cells and distinguish the tumor and virus infected cells from normal body cells. Phylogenetic analysis and comparison of domain structures identified three major groups of KIR genes (group I, II, and III genes). These groups of KIR genes, generated by a series of gene duplications, have acquired different MHC-binding specificity. Inference of ancestral KIR sequences suggested that the functional divergence of group I genes from group II genes occurred by positive selection at the MHC-binding sites after duplication. Our evolutionary study has shown that group I genes diverged from group II genes about 17 million years ago (Mya) apparently after separation of hominoids from Old World (OW) monkeys. Around the same time, gene duplication generating the class I MHC-C locus appears to have occurred. These findings suggest that KIR and MHC class I genes have coevolved as an interacting system. The KIR gene family has experienced a rapid expansion in primate species. The rate of expansion of this gene family seems to be one of the highest among all hominoid gene families. The KIR gene family is also subject to birth-and-death evolution.

© 2004 Elsevier B.V. All rights reserved.

Keywords: KIR; MHC; Adaptive evolution; Positive selection; Ancestral sequence

1. Introduction

Natural killer (NK) cells are a critical component of the innate immune system. They are essential for the early immune response against tumor and virus-infected cells. The cytotoxic activity of NK cells is regulated through the interaction of the receptors on the NK-cell surface with major histocompatibility complex (MHC) class I molecules. It has been shown that a large number of different receptors are expressed on the surface of NK cells (see review by [Trowsdale et al., 2001](#)). On the basis of the protein structure,

the NK cell receptors can be classified into two groups: the immunoglobulin-like receptor superfamily (IgSF) and the calcium independent C-type lectin receptor superfamily. The family of killer cell immunoglobulin-like receptors (KIR) belongs to the IgSF family, and in humans the member genes of the family are located tandemly in a genomic region called the leukocyte receptor complex (LRC) on chromosome 19q13.4. Many other genes belonging to the IgSF family are also located in the LRC.

Interestingly, primates and rodents use different NK cell receptors (KIRs for primates and Ly49s for rodents) to regulate the activity of NK cells, although they have a similar signaling pathway ([Trowsdale et al., 2001](#)). The KIR gene family is composed of many member genes, and these member genes are highly diversified in hominoids and Old World (OW) monkeys (e.g., [Khakoo et al., 2000](#); [Hershberger et al., 2001](#); [Guethlein et al., 2002](#)). In cattle, multiple

Abbreviations: KIR, killer cell immunoglobulin-like receptor; MHC, major histocompatibility complex; ITIM, immunoreceptor tyrosine inhibitory motif; Mya, million years ago.

* Corresponding author. Tel.: +1 814 863 7334; fax: +1 814 863 7336.

E-mail address: nxm2@psu.edu (M. Nei).

KIR genes and a single Ly49 gene have been identified (McQueen et al., 2002; Storset et al., 2003). The KIR genes appear to be absent in rodents except one or two presumably nonfunctional genes (Hoelsbrekken et al., 2003).

There is a high degree of diversity in primate KIR genes, which is generated by multiple genes, multiple alleles, multiple haplotypes, multiple splicing sites, and multiple recombinations (e.g., Toneva et al., 2001; Rajalingam et al., 2004). The expansion of gene members is caused primarily by repeated gene duplication (Martin et al., 2004). Human and chimpanzee KIR genes are quite different from each other (Khakoo et al., 2000). Examination of KIR genes in other primates also showed a pattern of species-specific diversification (Hershberger et al., 2001; Guethlein et al., 2002). The rapid evolution of high diversity of primate KIR genes seems to be exceptional if we consider their recent origin of duplicate genes.

Parham (1997) proposed a ‘catch-up’ model of evolution, in which KIR genes evolve rapidly because of the pressure for catching up with the evolutionary changes of their ligands, highly polymorphic MHC class I molecules. Hughes (2002) studied the evolution of human KIR genes using Ig-like domains as the evolutionary units and identified positive Darwinian selection in one of the Ig-like domains. However, it remains unclear how KIR and MHC genes have coevolved in the presence of their interaction. It is also interesting to investigate whether the positive selection plays any role in the coevolution.

The purpose of this paper is three-fold: (1) investigation of the coevolution between KIR and MHC genes, (2) examination of possible involvement of positive selection on the evolution of human KIR genes, and (3) estimation of the divergence time of different lineages of primate KIR genes to determine the rate of expansion of KIR genes by

gene duplication. Because we are primarily interested in the long-term evolution of the KIR gene family, we used genes rather than individual Ig-like domains as the units of evolution.

2. Materials and methods

2.1. Background information

In the human population, there are a large number of haplotypes and hundreds of different (diploid) genotypes (e.g., Trowsdale et al., 2001; Hsu et al., 2002). The number of KIR genes varies considerably with haplotype. Eight major haplotypes are presented in Fig. 1. The first haplotype (H1) with a single activating receptor (2DS4) has the highest frequency (~50%) in the Caucasian population. The other haplotypes have at least two activating receptors (represented by ‘S’; see below) (Hsu et al., 2002). Two pairs of KIR sequences, 3DL1/3DS1 and 2DL2/2DL3, are likely to represent different alleles, because they are always located on the same genomic regions.

In this study the standard nomenclature of KIR genes was used (Marsh et al., 2003), and it is related to their organization of Ig-like domains (Fig. 2). Each KIR molecule contains the extracellular, transmembrane, and cytoplasmic regions. There are two or three Ig-like domains in the extracellular region. Each KIR gene is denoted by four alphanumeric characters. The first two letters (2D or 3D) represent the number of Ig-like domains of the molecule. In some KIR molecules, the cytoplasmic region contains one or two immunoreceptor tyrosine inhibitory motifs (ITIM). The KIRs with ITIM are inhibitory receptors, and they have the ability to inhibit the cellular activity when they bind to ligands. By



Fig. 1. Genomic organizations of human KIR genes in 8 representative haplotypes (H1–H8) (adapted from Hsu et al., 2002). Each haplotype has 8–14 KIR genes in a ~200 kb region. The black box represents the framework locus which is present on all the examined haplotypes. The open box represents the KIR gene which varies with haplotype. Only gene order is shown, and the distance between genes is not to scale. Each haplotype has occurred at least twice in the population studies (Hsu et al., 2002).

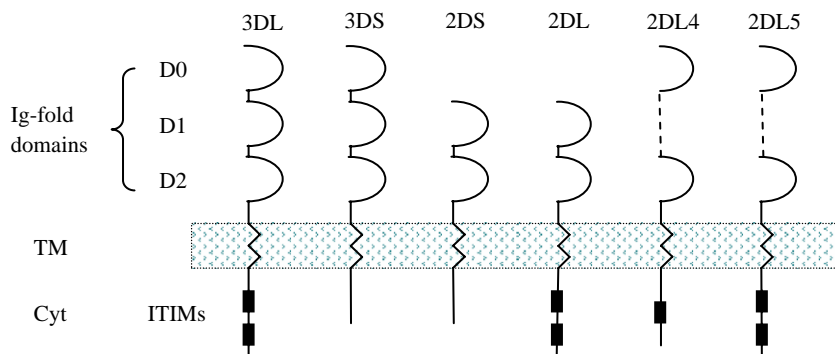


Fig. 2. Domain organizations of KIRs. Most KIRs can be classified into 4 subgroups according to the domain organization. One recently identified cattle KIR molecule (2DS1) has an unusual domain organization (D0 and D1) (Storset et al., 2003), which was not shown here. TM, transmembrane region. Cyt, cytoplasmic region. ITIM, the immunoreceptor tyrosine inhibitory motif.

contrast, the activating receptors do not contain ITIM. Instead, they recruit an adaptor protein through a charged amino acid in the transmembrane region and contribute to the activation of NK cells. ‘L’ denotes long cytoplasmic tail and is most likely to represent an inhibitory receptor with ITIM, whereas ‘S’ represents an activating receptor without ITIM. The letter ‘P’ represents a pseudogene. Primate KIRs are divided into four subgroups (3DL, 3DS, 2DL, and 2DS) according to the number of Ig-like domains and the presence or absence of ITIM (Fig. 2). The last digit of the nomenclature designates the order of identification of the gene in each subgroup. The three Ig-like domains in subgroups 3DL and 3DS are designated as D0, D1, and D2, starting from the N-terminal (Fig. 2). The KIRs in subgroups 2DL and 2DS contain domains D1 and D2, except 2DL4 and 2DL5 which have domains D0 and D2 instead.

2.2. Phylogenetic analysis

All the sequence alignments were obtained by using computer program CLUSTAL X (Thompson et al., 1997). Additional modifications were done by visual inspection. Phylogenetic analysis was conducted by using the computer program MEGA2 (Kumar et al., 2001). We constructed phylogenetic trees for human genes and for primate genes separately by the neighbor-joining (NJ) method (Saitou and Nei, 1987) after elimination of all alignment gaps (complete deletion option). We used p-distance (proportion of uncorrected nucleotide differences) to construct phylogenetic trees because p-distance has a smaller variance than other distances and often gives a better resolution of the topology (Nei and Kumar, 2000). In the estimation of divergence times, however, we used the Jukes–Cantor, Kimura’s 2-parameter, and Tamura–Nei distances to take care of multiple substitutions and the transition/transversion and GC content biases. However, the results obtained by these three distance measures were virtually identical because the extent of sequence divergences was small and there were not much transition/transversion and GC content biases. We therefore present only the results obtained by Kimura’s distance in this paper. To examine the reliability of topologies generated by

the NJ method, we also constructed MP trees using PAUP* 4.0 (Swofford, 1998). For the dataset of human KIR genes, the branch-and-bound search was used with 1000 bootstrap replications. For the primate KIR genes, we used the standard stepwise addition and TBR search with 500 bootstrap replications. However, since the major pattern of the topology of the parsimony tree was essentially the same as that of the NJ tree, we presented only the NJ tree in this paper.

The amino acid and nucleotide sequences of all ancestral nodes of the phylogenetic tree were inferred from the present-day sequences using the programs ANCESTOR and ANCGENE (Zhang and Nei, 1997; <http://mep.bio.psu.edu/>).

When we studied the coevolution of KIR and class I MHC genes in humans, we used 15 human KIR sequences excluding pseudogenes 2DP1 and 3DP1. Two pairs of potential KIR alleles (3DL1/3DS1 and 2DL2/2DL3) were included in this study, because they are functionally different and we wanted to include every sequence to identify the potential positive selection and functional coevolution between KIR and MHC genes.

In the study of the rate of gene family expansion, we used 47 KIR sequences from humans (10), common chimpanzees (6), pygmy chimpanzees (3), gorillas (7), orangutans (6), and macaques (15) after the following data filtering processes. (a) We excluded the KIR genes (2DL4 and 2DL5) that lacked the D1 domain because inclusion of these genes decreased the number of informative nucleotide sites in our phylogenetic analysis (complete-deletion option). (b) The distinction between loci and alleles was not always clear for non-human primate KIR sequences. Because we are interested in the estimation of gene expansion rate, the allelic sequence should not be included. We used only one of the potentially allelic sequences, which had a pairwise p-distance of less than 0.01. This criterion is conservative because the average nucleotide difference between different alleles in humans is about 0.003 (Li and Sadler, 1991). For human KIR sequences, 2DL3 and 3DS1 were excluded because they are the alleles of 2DL2 and 3DL1, respectively.

In our phylogenetic analysis, we included cattle KIR genes and recently identified rodent KIR-like sequences. The names of the sequences in the primate dataset and their

Table 1
Gene sequences used

Humans (hum)	2DS4 (AF258807)	3DL5 (AF334620)
3DL1 (NM_013289)	Pygmy chimpanzees (bono)	3DL6 (AF334621)
3DL2 (AJ276125)	3DLb (AF266733)	3DL7 (AF334622)
3DL3 (AF352324)	3DL4 (AF266731)	3DL8 (AF334623)
3DS1 (NM_014514)	3DSa (AF266735)	3DL9 (AF334624)
2DL1 (L41267)	Gorillas (gori)	3DL10 (AF334625)
2DL2 (NM_014219)	3DL7 (AY122869)	3DL11 (AF334626)
2DL3 (L41268)	2DLb (AY122871)	3DL16 (AF361084)
2DL4 (NM_002255)	2DLc (AY122872)	3DL18 (AF361086)
2DL5A (AL133414)	2DLd (AY122873)	3DH1 (AF334648)
2DL5B (AF217486) ^a	2DLe (AY122874)	3DH2 (AF334649)
2DS1 (X89892)	2DL6 (AY122870)	3DH3 (AF334650)
2DS2 (U24079)	2DSa (AY122875)	3DH4 (AF334651)
2DS3 (NM_012313)	Orangutans (oran)	Cattle (cattle)
2DS4 (U24077)	3DLA (AF470365)	3DL1 (AF490402) ^b
2DS5 (NM_014513)	3DLE (AF470370) ^c	3DL2 (AY075103) ^d
2DP1 (AL133414) ^a	2DLA (AF470358)	3DS1 (AF490401)
Common chimpanzees (chimp)	2DLB (AF470359)	Mice (mice)
3DL1 (AF258798) ^e	2DSA (AF470360)	3DL1 (NM_177748)
3DL4 (AF258800)	2DSC (AF470362) ^c	2DL1 (AY152727)
3DL5 (AF258801)	Macaques (maca)	Rats (rat)
3DL6 (AF258802)	3DL1 (AF334616)	3DL1 (AF527797)
3DS2 (AF258803)	3DL2 (AF334617)	

^a The gene sequences were extracted from the genomic contig.

^b Cattle 3DL1 and 3DS1 were obtained from Storset et al. (2003).

^c In Guethlein et al. (2002)'s paper, the gene 3DLE was 3DLE1 and 2DSC was 2DSC1.

^d Cattle 3DL2 is equal to the 3DL1 identified by McQueen et al. (2002).

^e The gene 3DL1 was denoted as 3DL1/2 in Khakoo et al. (2000)'s analysis.

GenBank accession numbers are given in Table 1. The letter 'H' in the name of some macaque KIR genes represents a molecule, in which the extracellular region is more similar to those of subgroup 3DL genes, whereas the cytoplasmic region has only one ITIM that is similar to that of 2DL4.

3. Results

3.1. Evolutionary relationships of KIR and MHC class I genes

The NJ tree for 15 human KIR sequences is presented in Fig. 3. Since this NJ tree is based on only the D2 domain, transmembrane, and cytoplasmic regions because of the complete deletion option used, we constructed another NJ tree using the nucleotide sequences of domain D1 only (Fig. 4). Genes 2DL4, 2DL5A, and 2DL5B were not used in this case because they lacked the D1 domain. This tree consisted of the same sequence groups as those of the tree in Fig. 3. According to these trees, there are three distinct groups of human KIR genes. This observation is consistent with previous results (Khakoo et al., 2000; Guethlein et al., 2002), although our trees were rooted unlike the previous trees and the correlation

between the phylogenetic grouping and the ligand binding specificity was clearly shown. Group I is the most recently diverged group with the largest number of genes, while group III is the most ancient. Eight KIR sequences are clustered into group I with a high bootstrap value and all of them share a pseudo-exon encoding a 'silent' D0 domain which is absent in the mature mRNA due to alternative splicing (Vilches et al., 2000). This supports that group I KIR genes are monophyletic.

The phylogenetic groups of human KIR genes are well correlated with the domain organization and ligand-binding specificity of their encoded proteins except for gene 3DL3. The KIRs encoded by group II genes share the organization of three Ig-like domains (D0, D1, and D2). Both group I and group III KIRs have the structure of two Ig-like domains, but with different combinations (Fig. 3). Furthermore, evolutionarily closely related KIRs tend to share similar MHC-binding properties. Six group I KIRs have been identified to bind to MHC-C alleles. In group II genes, 3DL1 and 3DS1 bind to MHC-B alleles, and 3DL2 to MHC-A alleles. As for group III genes, 2DL4 is known to bind to non-classical MHC-G molecules (see review by Boyington and Sun, 2001). Interestingly, the ligand-binding specificity of chimpanzee KIRs is similar to that of human KIRs in each group (Khakoo et al., 2000). This suggests that these different KIR gene groups have acquired different ligand-binding specificities before the divergence of the two species. In addition, the evolution of new MHC-binding specificity for KIR genes is correlated with the evolution of MHC class I genes in primates (Piontkivska and Nei, 2003). Therefore, we investigated the role of MHC ligands on the diversification of different groups of KIR genes and inferred the potential MHC-binding specificity of the ancestral KIR sequences.

3.2. Positive selection during the early expansion of group I human KIR genes

The MHC class I molecules are the only ligands of KIRs so far identified. The interaction of groups I and II KIRs with MHC molecules requires mainly D1 and D2 domains of KIR proteins (Boyington and Sun, 2001). The comparison of pairwise nonsynonymous (d_N) and synonymous (d_S) nucleotide substitutions for the individual Ig-like domain of human and other primate KIR genes (data not shown) has suggested that the d_N/d_S ratio is higher than 1 in domain D1. Furthermore, Hughes (2002) suggested that the natural selection apparently operates at the MHC-binding region of D1 domain. These results are consistent with the observation that the less conserved D1 domain is primarily responsible for the MHC locus or allotype binding specificity of KIRs (Gumperz et al., 1997; Winter and Long, 1997). We therefore focused on domain D1 in the following study and studied the potential role of positive selection on the functional diversification of different groups of KIR genes.

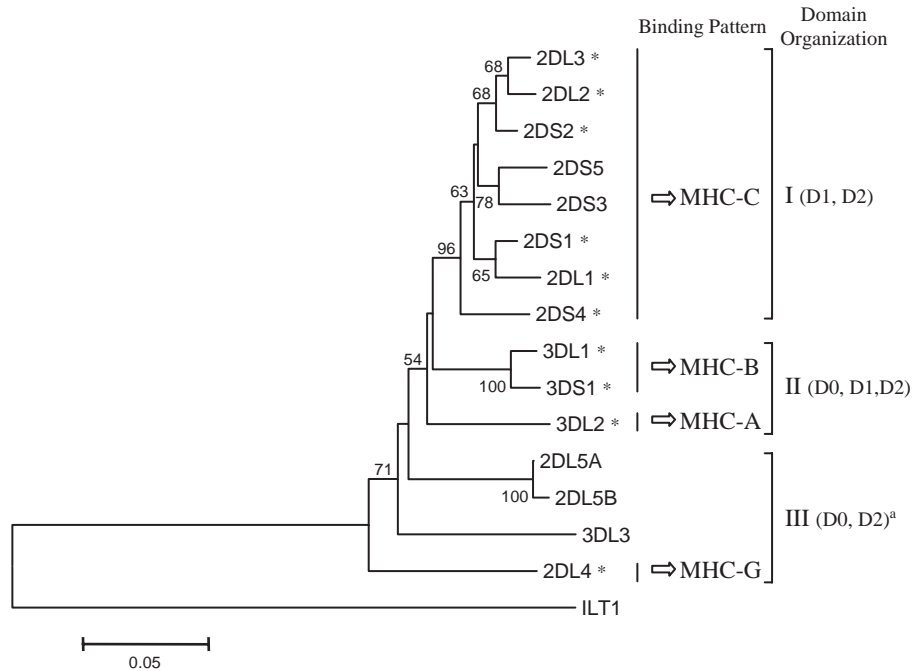


Fig. 3. Neighbor-joining tree of 15 human KIR nucleotide sequences. The p-distance was used. The bootstrap value based on 1000 replications is shown for each interior branch wherever it is greater than 50%. A human IgSF gene closely related to KIR genes, called ILT1 (immunoglobulin-like transcript), was used as the outgroup. Its Genbank accession number is U82275. The second Ig-like domain of this gene has been excluded in the alignment, since the gene has four Ig-like domains on the extracellular region and the first, third, and fourth domains are most similar to D0, D1, and D2 domains of KIRs, respectively (data not shown). The asterisk “*” represents the KIR protein with known MHC-binding specificities. The inferred MHC-binding property for each group is shown on the right hand side of the tree. ^a3DL3 has the structure of 3 Ig-like domains, which is different from 2DL4 and 2DL5.

The phylogenetic analysis of domain D1 suggested that group I KIR genes with MHC-C binding specificity diverged from group II KIR genes with MHC-A or B binding specificity after gene duplication occurred at node **a**

of the tree in Fig. 4. We therefore inferred the ancestral nucleotide sequences for each interior node of this tree and estimated the numbers of nonsynonymous (a_N) and synonymous (a_S) substitutions per sequence per branch.

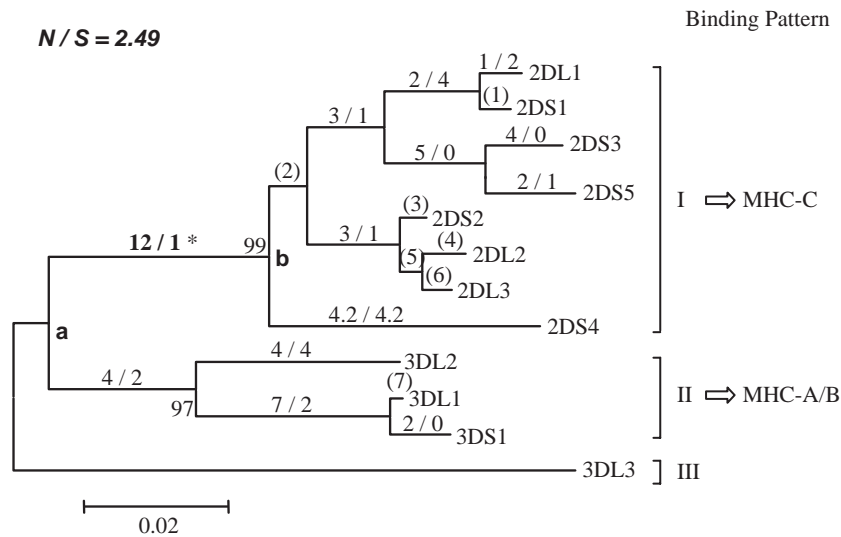


Fig. 4. Neighbor-joining tree of nucleotide sequences of the D1 domain from 12 human KIRs. The topology of this tree was obtained by using p-distance, and the same topology was obtained by using JC distance. The pseudogenes 3DP1 and 2DP1 were not included because of the nonsynonymous/synonymous substitution rate test. The number of nucleotides of D1 domain is 300. The gene duplication occurred at node **a**, from which 2 different groups (I and II) of KIR genes are diverged. The numbers of nonsynonymous (a_N) and synonymous (a_S) substitutions per sequence per branch are presented as a_N/a_S above the branches. The N/S ratio for all the KIR sequences including ancestral sequences is given above the tree. (1) (2/2). (2) (1.5/0.5). (3) (1/0). (4) (1/2). (5) (0/2). (6) (1/1). (7) (1/0). The statistical significance of the difference between observed ratio a_N/a_S and expected ratio N/S was determined by Fisher’s Exact Test which showed that the difference is significant at the 7% level (*).

There are 12 nonsynonymous substitutions but only 1 synonymous substitution on branch **a–b**. The Fisher’s Exact Test that compares the ratio of a_N/a_S with the expected ratio gives substantial statistical support ($P=0.07$) for the involvement of positive selection during the functional diversification of group I KIR genes (branch **a–b**).

3.3. Nonrandom amino acid substitutions for branch **a–b**

To examine what kind of amino acid changes occurred between nodes **a** and **b** in Fig. 4, we compared the ancestral amino acid sequences of domain D1 (Fig. 5). It appears that the amino acid changes from the ancestral sequences **a** to **b** are not random. Of a total of 9 amino acid substitutions between them, 4 are located at the MHC-binding sites (MBS) (labeled by asterisks in Fig. 5). The MBS of human groups I and II KIRs is known to include 18 amino acid sites, 7 of which are located in domain D1 and the remainder in D2 (Boyington and Sun, 2001). There are 100 amino acids in the D1 domain and 7 of them are at the MBS. If we assume that the amino acid substitutions occur randomly along the sequence, the probability of 4 out of 9 total substitutions occurring at the MBS can be computed by using the hypergeometric distribution, and it is only 0.00096. This result supports

the idea that an excessive number of amino acid substitution occurred at the MBS. A closer examination showed that all of the four amino acid substitutions between sequences **a** and **b** at the MBS were polarity changes. For example, the hydrophobic isoleucine (I) in **a** changed to the hydrophilic lysine (K) in **b** (labeled by the arrow sign in Fig. 5). These changes of polarity at the ligand-binding sites might be responsible for a change of the ligand-binding specificity.

Furthermore, to infer the potential ligand-binding specificity of ancestral sequences at nodes **a** and **b**, we compared their D1 domains with those of the present-day groups I and II KIR sequences (Fig. 5). The ancestral sequence **a** has a high degree of sequence similarity with group II KIR genes, while **b** is similar to group I KIR genes. Therefore, it is very likely that the ancestral sequence **a** might have recognized MHC-A or B-related molecules similar to that of group II KIRs. After the gene duplication at node **a**, one gene appears to have retained the old binding specificity, and the other (sequence **b**) evolved the new MHC-C binding specificity of the present-day group I KIRs. Further support for this argument comes from the extended comparison of the MBS in domain D1 among different primate KIR genes (Fig. 6). The amino acid compositions of groups I and II KIRs at the MBS are similar to those of the

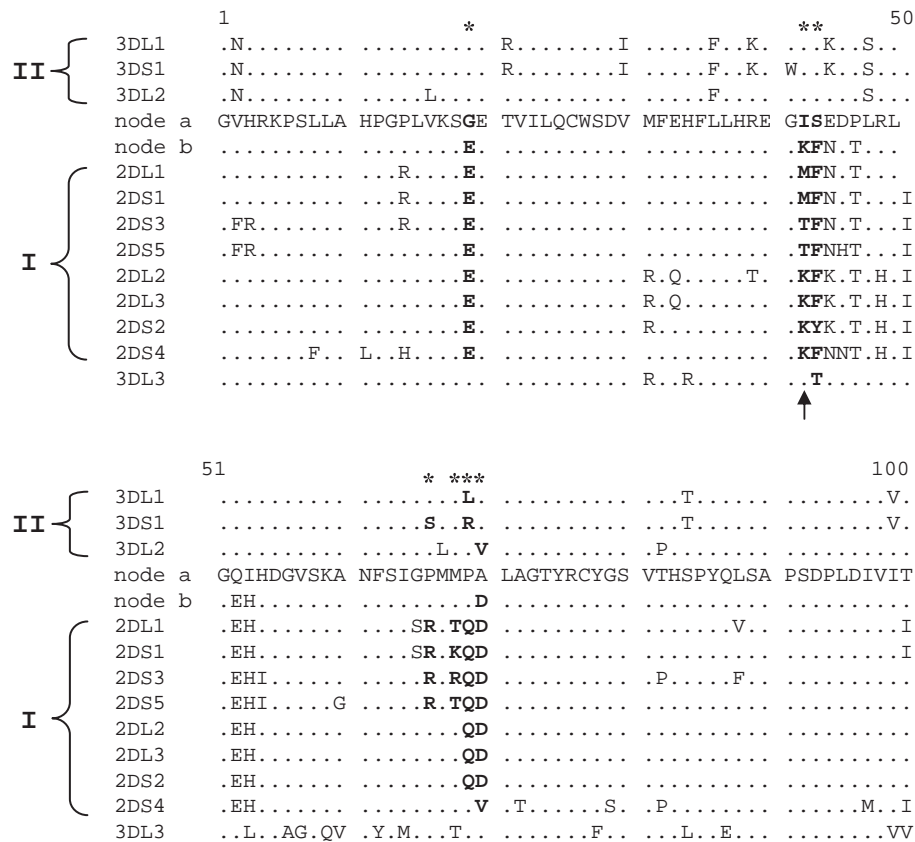


Fig. 5. Comparison of the D1 domain between ancestral KIR sequences (nodes **a** and **b** in Fig. 4) and present-day groups I and II human KIR sequences. Amino acids are presented by single-letter codes and dots represent the same amino acids as those of the sequence at node **a**. The MHC-binding sites (MBS) are shown in bold letters and also labeled with asterisks. The arrow sign represents the key amino acid which determines MHC-C allotypic binding specificities.

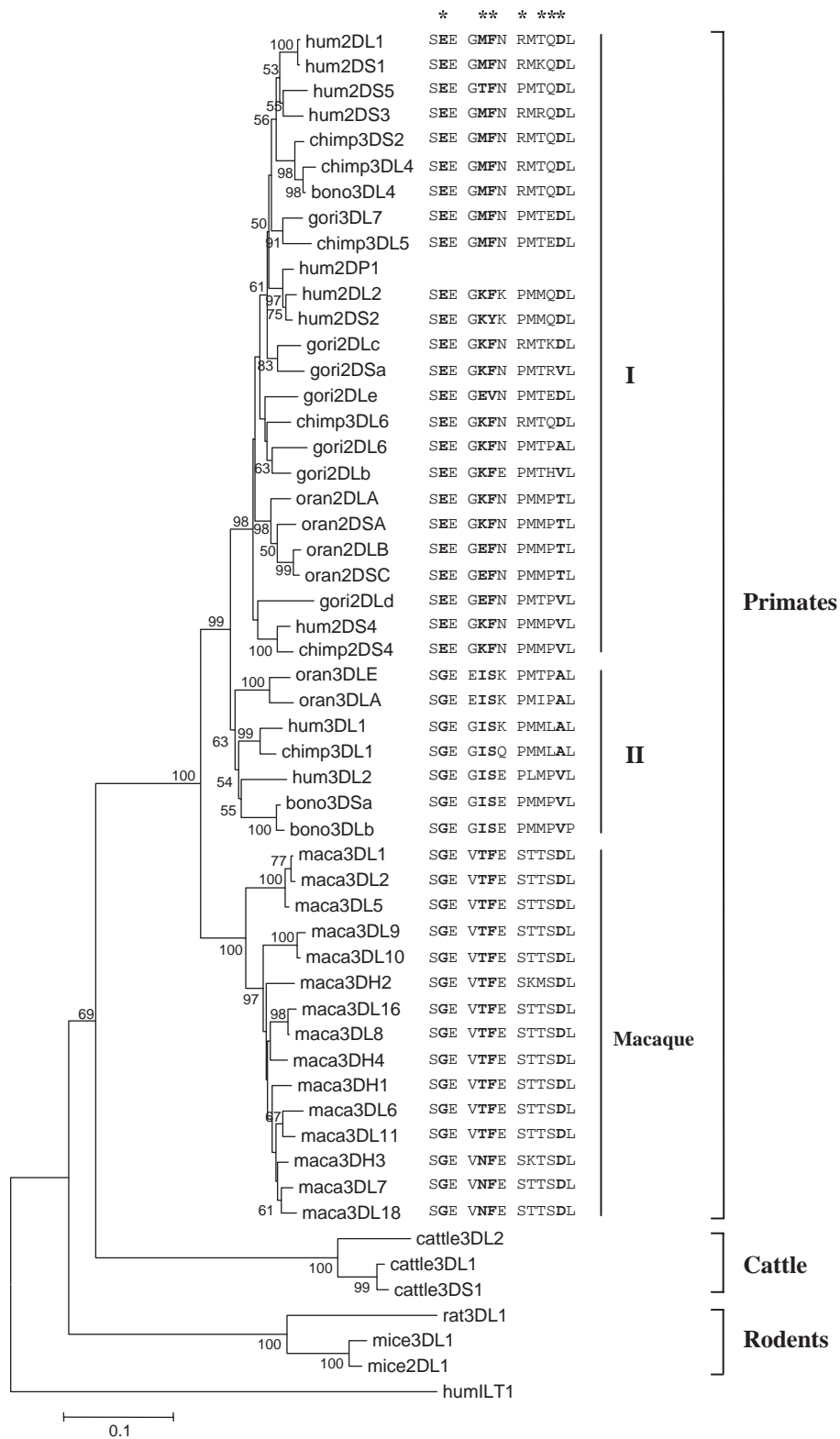


Fig. 6. Neighbor-joining tree of nucleotide sequences of 47 primate, 3 cattle, and 3 rodent KIR genes. We used both JC distance and p-distance and both trees had the same topology. The tree constructed by p-distance was shown here. Only extracellular regions were used and the number of nucleotide sites is 587. The three groups of KIR genes in primates are shown with vertical lines. The bootstrap value based on 1000 replications is shown for each interior branch wherever it is greater than 50%. The amino acids of MHC-binding sites (MBS) in the D1 domain (labeled with asterisks on the top) are shown for each primate KIR gene except for the pseudogene 2DP1. The 4 sites which are different between two inferred ancestral sequences (nodes **a** and **b** in Fig. 4) are shown in bold letters. The human ILT1 gene was used as the outgroup.

ancestral sequences **b** and **a**, respectively. The functional studies of chimpanzee KIRs showed that the KIR molecule encoded by a group II gene 3DL1 recognizes MHC-A and B, while two group I KIR genes, 2DL6 and 3DL4, have the MHC-C binding specificity (Khakoo et al., 2000). Overall, these results suggest that the amino acid substitutions in the MBS of D1 domain are related to the functional divergence from group II to group I KIR genes.

As mentioned earlier, the D1 domain is important in determining the MHC allotype specificity. For example, the change of a single amino acid at the MBS (labeled by the arrow sign in Fig. 5) switches the binding specificities from one allotype of MHC-C to another (Winter and Long, 1997). The human KIR molecules, 2DL2 and 2DL3, have lysine (K) at this site and they bind to the MHC-C allotypes (Cw1, 3, 7, and 8) which have asparagine (Asn) at amino acid position 80 (Asn⁸⁰) of the α 1 domain. However, 2DL1 has methionine (M) at this site and binds to the other MHC-C allotypes (Cw2, 4, 5, 6, and 15) with lysine at position 80 (Lys⁸⁰). Therefore, we speculate that the ancestral sequence **b** might have recognized the specific MHC-C allotypes (Asn⁸⁰) as in the case of 2DL2 and 2DL3.

3.4. Rapid differentiation of KIR genes by repeated gene duplication

One of the most distinctive features of evolution of the KIR gene family is the rapid increase of member genes by gene duplication. To study the rate of gene family expansion, we constructed an NJ tree of primate KIR genes using p-distance (Fig. 6). The cattle and rodent KIR genes which form distinct clusters from primate KIR genes were also included. According to the tree, both group I and group II KIR genes are shared by all hominoid species used (humans, chimpanzees, gorillas, and orangutans). In each of these two groups, the orangutan KIR genes form a cluster separate from human and chimpanzee genes. Even between humans and chimpanzees, there are only two pairs of KIR genes (2DS4 and 3DL1) that might be considered as the real orthologs between these two species. All macaque KIR genes form a clade separate from hominoid groups I and II KIR genes. Therefore, the KIR gene family in primates was diversified very rapidly and many gene duplication events have occurred in different primate species.

To examine the expansion rate of this gene family, we constructed a linearized tree (Fig. 7, Takezaki et al., 1995) and estimated the divergence times of duplicate genes using the human and orangutan divergence (13 Mya) as the calibration point (Glazko and Nei, 2003). To test the rate constancy among sequences, we used the branch length test as implemented in the LINTREE program (<http://mep.bio.psu.edu/>). This test indicated that there is no KIR sequence evolving significantly faster or lower than the average rate ($P < 0.01$). The molecular time scale in Fig. 7 shows that group I and group II KIR genes diverged about 17 Mya (16.6 ± 1.4 , node **B**), which is after the divergence between

OW monkeys and hominoids (23 Mya) and before the divergence of the human–chimpanzee lineage from the orangutan lineage (13 Mya). This explains why no macaque KIR genes are clustered with these two groups. Although the group III KIR genes (2DL4 and 2DL5) were not included in the analysis because of the lack of the D1 domain, they have been identified in both hominoids and OW monkeys (e.g., Khakoo et al., 2000; Hershberger et al., 2001) and diverged from group I and group II hominoid KIR genes as a basal lineage when the macaque-specific KIR group diverged (data not shown). Therefore, we speculate that the expansion of primate KIR genes started about at least 25 Mya (25.4 ± 3.0 at node **A** in Fig. 7).

There are a total of 12 groups I and II KIR genes excluding potential alleles in human genome but including pseudogenes 2DP1 and 3DP1 (Fig. 1). Because a gene family with n members should have been generated by $n-1$ gene duplication events, the rate of gene expansion for the KIR gene family in the human lineage is roughly estimated to be 0.65 (11/17) per million years (My). For the most diversified group (I), the gene expansion rate is even higher and becomes 0.69 (=9/13) per My under the assumption that humans and orangutans diverged 13 Mya. We have done similar computations for other primate species. The results have shown that in all primate species considered here the rate of expansion of KIR genes is very high (0.3–0.6, Table 2). Particularly the macaque KIR gene family showed a rate as high as 1.1 duplications per My. It is known that MHC class I and class II genes also have experienced rapid gene duplications. The results obtained by a similar approach indicate that the rates of MHC gene expansion are more than 5 times lower than those of KIR genes (Table 2, Takahashi et al., 2000; Piontkivska and Nei, 2003). There is only one reported gene family that shows a rate of gene expansion as high as that for KIR genes. It is the morpheus gene family whose function is unknown (Johnson et al., 2001). In this gene family, gene duplication occurred 13 times in the human lineage after separation from orangutans. So, the rate becomes one duplication per one My. The Ly49 gene family in rodents, which is the functional equivalent of primate KIR genes, has an expansion rate as high as that of KIR genes in primates (Hao and Nei, 2004).

3.5. Coevolution between KIR and MHC class I genes

The evolutionary pattern of primate KIR genes appears to coincide with that of classical MHC class I genes within primates. The group I KIR genes diverged from group II KIR genes soon after the gene duplication that generated the MHC-C locus 21–28 Mya (Piontkivska and Nei, 2003). The orthologs of human MHC-C genes could be identified in other hominoids but not in OW monkeys and New World (NW) monkeys. Furthermore, although macaques have true orthologs of hominoid MHC-A and B genes, they are distantly related to each other (Adams and Parham, 2001). These observations explain why no group I KIR genes are

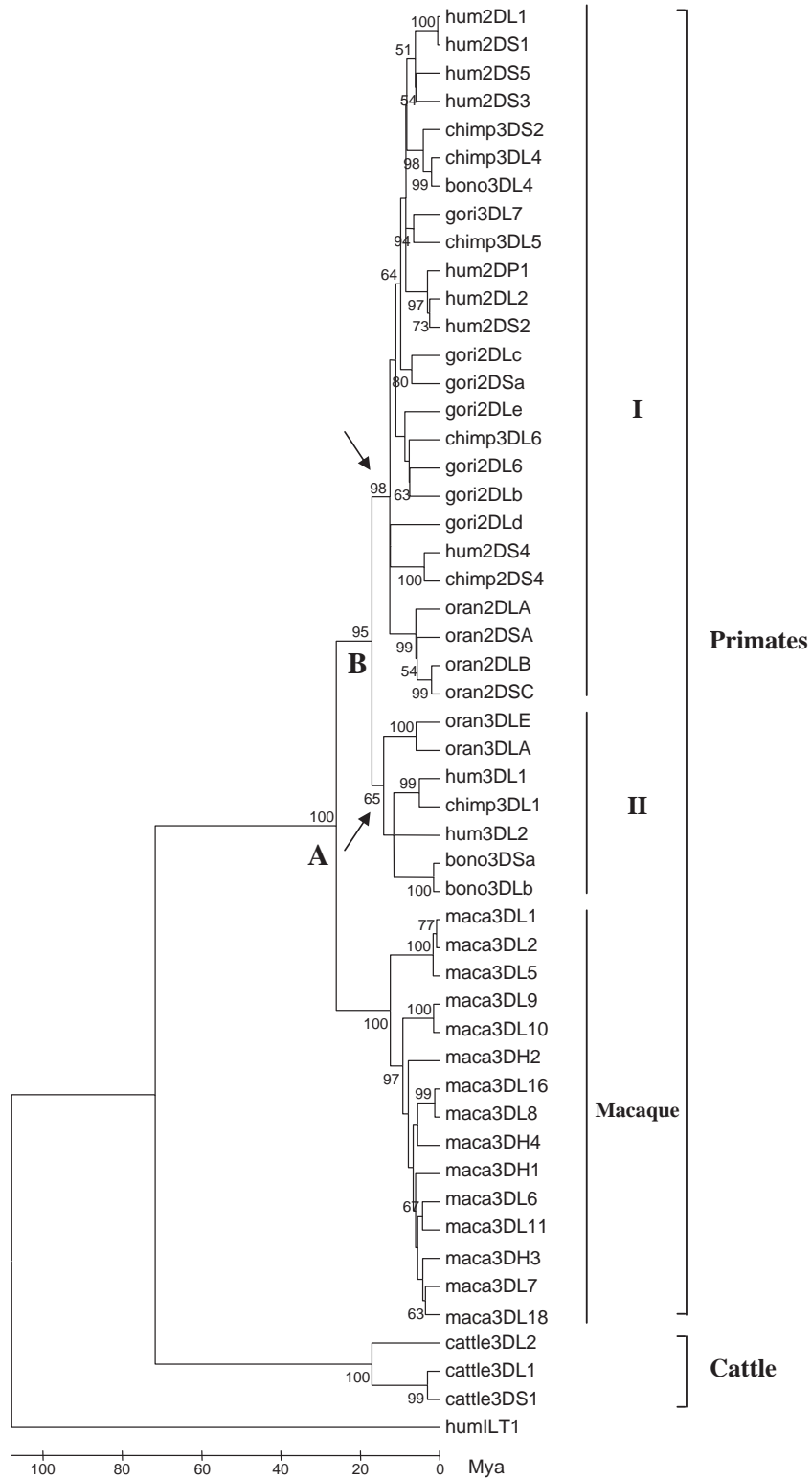


Fig. 7. Linearized tree of the primate KIR genes obtained by using Kimura's distance. Rodent KIR-like genes were excluded because the function of these genes is unknown. The time scale was calibrated with the divergence time between humans and orangutans (13 Mya, labeled with arrow signs). The nodes of which the divergence times we are interested in are labeled as capital letters in bold.

found in macaques and all macaque KIR genes form a species-specific cluster different from groups I and II KIR genes (Fig. 6). Orangutan KIR genes form a separate cluster

different from human and chimpanzee KIR genes in both groups I and II. This is consistent with the observation that the classical MHC loci in orangutans are distantly related to

Table 2

Rates of gene expansion for the groups I and II hominoid KIR lineage, the macaque-specific KIR lineage, and other gene families

Gene family	Evolutionary time (My)	Duplication events	Expansion rate per my	Reference
Class II MHC	50	4	0.08	Takahashi et al. (2000)
Class I MHC	50	4	0.08	Piontkivska and Nei (2003)
Groups I and II human KIR	17	11	0.65	Our studies
Group I common chimpanzee KIR	13	4 ^a	0.31	Our studies
Group I gorilla KIR	13	6 ^a	0.46	Our studies
Groups I and II orangutan KIR	17	5 ^a	0.29	Our studies
Macaque KIR	13	14 ^a	1.08	Our studies
Morpheus	13	13	1.00	Johnson et al. (2001)

^a The number of KIR genes might not be fully identified.

the orthologs in chimpanzees and humans. It has been shown that the presence of the MHC-C is polymorphic in orangutan populations. In light of the positive selection at MBS of KIR genes shown earlier, the correlation of evolutionary pattern between KIR genes and MHC class I genes seems to suggest a coevolutionary relationship between them during the evolution of primates.

4. Discussion

As mentioned earlier, the genetic variation of the KIR gene family is manifested primarily as haplotype diversity generated by tandem duplication, block duplication, nucleotide substitution, gene translocation, recombination, and gene loss, etc. Without having more information about the KIR haplotype structure in other primates, it is difficult to evaluate the contribution of each of these factors. However, our rough estimates of the divergence times of different KIR lineages provide us with an idea about how quickly the KIR gene family expanded and how their evolution correlated with the evolution of their ligands during the primate evolution.

We have emphasized the increase of functional genes in current species of hominoids and OW monkeys. However, it is likely that their ancestral species had many KIR genes and many previously functional genes have been lost or become pseudogenes. Indeed, the KIR gene family has several pseudogenes in humans, chimpanzees, and macaques. Therefore, this gene family appears to be subject to birth-and-death evolution, like the other multigene families such as the MHC and immunoglobulin gene families (Nei et al., 1997). In our study, however, it was difficult to study the death rate of genes, because we could not estimate the number of genes in the genome of ancestral organisms.

The biological function of a haplotype must depend on the number of inhibitory and activating genes, allelic sequences, and splicing properties, etc. We can measure the extent of haplotype diversity (H) in a population by a quantity similar to gene diversity (or heterozygosity) of a gene (Nei, 1987, pp.178). In the present case, it is given by $H = n \left(1 - \sum_{i=1}^N X_i^2 \right) / (n-1)$, where n , N , and X are the total number of genomes examined, the number of different

haplotypes observed, and the frequency of the i -th haplotype, respectively. In a Caucasian population studied by Hsu et al. (2002), we have estimated H to be 0.81. This value is much higher than the gene diversity (0.135) observed by protein electrophoresis in human populations (Nei, 1987, pp.192). However, this haplotype diversity must be minimum because only a small number of individuals have been studied so far.

It is interesting to note that the genetic variability of MHC class I molecules is generated primarily by allelic variation rather than by haplotype diversity. If KIR genes and MHC alleles coevolved to cope with antigen evolution as suggested by Parham (1997) and by the results presented here, why is haplotype diversity important in KIR genes? If the amino acid sequences of the MHC-binding regions of KIR coevolve with the corresponding regions of MHC molecules, there is no need for a high level of haplotype diversity. Probably, one reason for the occurrence of many haplotypes is that both inhibitory and activating genes are required in this system, and the two types of genes apparently must be relatively closely related to each other. If this is the case, unequal interlocus recombination may occur frequently, and this will increase the number of different haplotypes. At the present time, however, the real mechanism of the interaction between KIRs and MHC molecules is not well understood, and it would be difficult to know the reason for generating many different haplotypes.

In rodents a structurally different gene family called the Ly49 gene family plays essentially the same function as that of KIRs in primates as mentioned. As in the case of KIR genes, the Ly49 gene family in rodents has experienced a rapid expansion of member genes by repeated gene duplication (Wilhelm et al., 2002; Hao and Nei, 2004). It would be interesting to know how these two entirely different gene families have come to play the same biological function.

Acknowledgements

We thank Jan Klein, Jongmin Nam, Yoshihito Niimura, and Nikolas Nikolaidis for their discussion. This work was supported by NIH grant GM20293 (to M.N.).

References

- Adams, E.J., Parham, P., 2001. Species-specific evolution of MHC class I genes in the higher primates. *Immunol. Rev.* 183, 41–64.
- Boyington, J.C., Sun, P.D., 2001. A structural perspective on MHC class I recognition by killer cell immunoglobulin-like receptors. *Mol. Immunol.* 38, 1007–1021.
- Glazko, G., Nei, M., 2003. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* 20, 424–434.
- Guethlein, L.A., Flodin, L.R., Adams, E.J., Parham, P., 2002. NK cell receptors of the orangutan (*Pongo pygmaeus*): a pivotal species for tracking the coevolution of killer cell Ig-like receptors with MHC-C. *J. Immunol.* 169, 220–229.
- Gumperz, J.E., et al., 1997. Conserved and variable residues within the Bw4 motif of HLA-B make separable contributions to recognition by the NKB1 killer cell-inhibitory receptor. *J. Immunol.* 158, 5237–5241.
- Hao, L., Nei, M., 2004. Genomic organization and evolutionary analysis of Ly49 genes encoding the rodent natural killer cell receptors: rapid evolution by repeated gene duplication. *Immunogenetics* 56, 343–354.
- Hershberger, K.L., Shyam, R., Miura, A., Letvin, N.L., 2001. Diversity of the killer cell Ig-like receptors of rhesus monkeys. *J. Immunol.* 166, 4380–4390.
- Hoelsbrekken, S.E., et al., 2003. Cutting edge: molecular cloning of a killer cell Ig-like receptor in the mouse and rat. *J. Immunol.* 170, 2259–2263.
- Hsu, K.C., Chida, S., Geraghty, D.E., Dupont, B., 2002. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunol. Rev.* 190, 40–52.
- Hughes, A.L., 2002. Evolution of the human killer cell inhibitory receptor family. *Mol. Phylogenet. Evol.* 25, 330–340.
- Johnson, M.E., et al., 2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413, 514–519.
- Khakoo, S.I., et al., 2000. Rapid evolution of NK cell receptor systems demonstrated by comparison of chimpanzees and humans. *Immunity* 12, 687–698.
- Kumar, S., Tamura, K., Jakobsen, I., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245.
- Li, W.-H., Sadler, L.A., 1991. Low nucleotide diversity in man. *Genetics* 129, 513–613.
- Marsh, S.G., et al., 2003. Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Tissue Antigens* 62, 79–86.
- Martin, A.M., et al., 2004. Comparative genomic analysis, diversity and evolution of two KIR haplotypes A and B. *Gene* 335, 121–131.
- McQueen, K.L., Wilhelm, B.T., Harden, K.D., Mager, D.L., 2002. Evolution of NK receptors: a single Ly49 and multiple KIR genes in the cow. *Eur. J. Immunol.* 32, 810–817.
- Nei, M., 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei, M., Kumar, S., 2000. *Molecular evolution and phylogenetics*. Oxford University Press, Oxford, New York.
- Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate system. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7799–7806.
- Parham, P., 1997. Events in the adaptation of natural killer cell receptors to MHC class I polymorphisms. *Res. Immunol.* 3, 190–194.
- Piontkivska, H., Nei, M., 2003. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Mol. Biol. Evol.* 20, 601–609.
- Rajalingam, R., Parham, P., Abi-Rached, L., 2004. Domain shuffling has been the main mechanism forming new hominoid killer cell Ig-like receptors. *J. Immunol.* 172, 356–369.
- Saitou, N., Nei, M., 1987. The neighbour-joining method: a new method for constructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Storset, A.K., Slettedal, I.O., Williams, J.L., Law, A., Dissen, E., 2003. Natural killer cell receptors in cattle: a bovine killer cell immunoglobulin-like receptor multigene family contains members with divergent signaling motifs. *Eur. J. Immunol.* 33, 980–990.
- Swofford, D.L., 1998. *PAUP. Phylogenetic analysis using parsimony*. Sinauer Associates, Sunderland, Massachusetts.
- Takahashi, K., Rooney, A.P., Nei, M., 2000. Origins and divergence times of mammalian class II MHC gene clusters. *J. Hered.* 91, 198–204.
- Takezaki, N., Rzhetsky, A., Nei, M., 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12, 823–833.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24, 4876–4882.
- Toneva, M., et al., 2001. Genomic diversity of natural killer cell receptor genes in three populations. *Tissue Antigens* 57, 358–362.
- Trowsdale, J., Barten, R., Haude, A., Stewart, C.A., Beck, S., Wilson, M.J., 2001. The genomic context of natural killer receptor extended gene families. *Immunol. Rev.* 181, 20–38.
- Vilches, C., Pando, M.J., Parham, P., 2000. Genes encoding human killer-cell Ig-like receptors with D1 and D2 extracellular domains all contain untranslated pseudoexons encoding a third Ig-like domain. *Immunogenetics* 51, 639–646.
- Wilhelm, B.T., Gagnier, L., Mager, D.L., 2002. Sequence analysis of the Ly49 cluster in C57BL/6 mice: a rapidly evolving multigene family in the immune system. *Genomics* 80, 646–661.
- Winter, C.C., Long, E.O., 1997. A single amino acid in the p58 killer cell inhibitory receptor controls the ability of natural killer cells to discriminate between the two groups of HLA-C allotypes. *J. Immunol.* 158, 4026–4028.
- Zhang, J., Nei, M., 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.* 44 (Suppl. 1), S139–S146.