

Li Hao · Masatoshi Nei

## Genomic organization and evolutionary analysis of *Ly49* genes encoding the rodent natural killer cell receptors: rapid evolution by repeated gene duplication

Received: 5 May 2004 / Accepted: 1 July 2004 / Published online: 14 August 2004  
© Springer-Verlag 2004

**Abstract** *Ly49* genes regulate the cytotoxic activity of natural killer (NK) cells in rodents and provide important protection against virus-infected or tumor cells. About 15 *Ly49* genes have been identified in mice, but only a few genes have been reported to date in rats. Here we studied all *Ly49* genes in the entire rat genome sequence and identified 17 putative functional and 16 putative non-functional genes together with their genomic locations in a 1.8-Mb region of chromosome 4. Phylogenetic analysis of these genes indicated that the *Ly49* gene family expanded rapidly in recent years, and this expansion was mediated by both tandem and genomic block duplication. The joint phylogenetic analysis of mouse and rat genes suggested that the most recent common ancestor of the two species had at least several *Ly49* genes, but that the majority of current duplicate genes were generated after divergence of the two species. In both species *Ly49* genes are apparently subject to birth-and-death evolution, but the birth and death rates of *Ly49* genes are higher in rats than in mice. The rate of gene expansion in the *Ly49* gene family in rats is one of the highest among all mammalian multigene families so far studied. The biochemical function of *Ly49* genes is essentially the same as that of *KIR* genes in primates, but the molecular structures of the two groups of NK cell receptors are very different. A hypothesis was presented to explain the origin of the differential use of *Ly49* and *KIR* genes in rodents and primates.

**Keywords** *Ly49* · Rat · Mouse · NK cell receptors · Fast-evolving genes

### Introduction

Natural killer (NK) cells are an important component of the innate immune system in mammals and are capable of discriminating virus-infected or tumor cells from healthy cells. This discrimination is accomplished by a large number of NK cell receptors that are expressed on the surface of NK cells (Trowsdale et al. 2001). Two of those NK cell receptors are killer-cell immunoglobulin-like receptors (KIR) and C-type lectin-like receptors (*Ly49*). The KIR and *Ly49* receptors have very different molecular structures and are encoded by different gene families. Yet, they have the same function and regulate the cytotoxic activity of NK cells through binding to major histocompatibility complex (MHC) class I molecules or MHC-related molecules (Colonna and Samaridis 1995; Brennan et al. 1996; Dorfman and Raulet 1996; Idris et al. 1999; Brown et al. 2001). Therefore, KIR and *Ly49* receptors represent a rare case of functional convergence at the molecular level. Interestingly, KIRs are used primarily in primates, whereas *Ly49*s are used in rodents. At present, how these different NK receptor gene families evolved in primates and rodents remains a mystery.

The diversity and evolution of the *KIR* gene family have been studied extensively in several primate species, and we now know that the intraspecific variation of KIRs is generated by duplicate genes, haplotype variation, allelic polymorphism, and alternative splicing (Kwon et al. 2000; Martin et al. 2000; Wilson et al. 2000; Gardiner et al. 2001). However, the primary factor of evolution of the *KIR* gene family is the rapid expansion of member genes by repeated gene duplication (Hershberger et al. 2001; Guethlein et al. 2002). A similar mechanism has been suggested about the evolution of the *Ly49* gene family in mice. The genetic diversity of mouse *Ly49* genes has been studied extensively with different strains such as C57BL/6 and 129/J (Brown et al. 1997; Depatie et al. 2000; Makrigiannis et al. 2002; Wilhelm et al. 2002). There is substantial variation in gene number and gene content among different strains. Rats are also likely to have many *Ly49* genes, but only six genes have been reported to date

L. Hao (✉) · M. Nei  
Department of Biology, Institute of Molecular Evolutionary  
Genetics, Pennsylvania State University,  
208 Mueller Laboratory,  
University Park, PA 16802, USA  
e-mail: lxh186@psu.edu  
Tel.: +1-814-8652796  
Fax: +1-814-8637336



## Identification and sequence analysis of rat *Ly49* genes

A draft assembly of the rat genome sequence (RGSC v3.1) covering more than 90% of the entire genome is available from NCBI. The sequences were obtained from two female rats (BN/SsNHsd) of a highly inbred line (Rat Genome Sequencing Project Consortium 2004). This means that the *Ly49* cluster identified here represents a single haplotype. Using the known rat *Ly49* gene (*Ly49.9*) as the query sequence, we first located the *Ly49* gene cluster on the supercontig (NW\_043770, updated by NW\_047696 recently) of rat Chr 4 with the computer program BLAST (Zhang et al. 1998a, <ftp://ftp.ncbi.nih.gov/blast/>). To identify the complete set of *Ly49* genes, a relaxed search was performed with a relatively high *E*-value (0.01). The number of BLAST hits suggested that about 30 different *Ly49* genes are located in a 1.8-Mb region of NKC. After extracting each potential gene segment, we used PipMaker (Schwartz et al. 2000, <http://bio.cse.psu.edu/pipmaker/>) to identify the exon boundaries

for each gene, comparing the extracted genomic sequences with the *Ly49* cDNA (*Ly49.9*).

The genes identified in this way were named according to the order of the genes on the chromosome starting from the telomeric end of the *Ly49* gene cluster (Fig. 1B). The list of the genes is available in Table 1, and their sequences can be obtained from the Nei lab database (<http://mep.bio.psu.edu/databases/Ly49>). The amino acid sequence for each rat *Ly49* gene was deduced from the nucleotide sequences of exons by the nucleotide-translation program, ExPASy (<http://us.expasy.org/tools/dna.html>). The repetitive elements in the *Ly49* gene cluster were identified by the computer program RepeatMasker (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker.html>).

## Phylogenetic analysis

All the sequence alignments were generated by using the computer program CLUSTAL X (Thompson et al. 1997).

**Table 1** Rat *Ly49* genes and their predicted characteristics

Gene	Previous gene	Group	Inhibitory/ activating	Note
<i>Ly49-1</i>	–	VI	I	Orthologous to mouse <i>Ly49-B</i>
<i>Ly49-2</i>	–	III	I	
<i>Ly49-3</i>	<i>Ly49.9</i>	III	I	
<i>Ly49-4</i>	<i>Ly49.29</i>	III	A	
<i>Ly49-5</i>	–	III	–	Insertion, premature stop codon in exon 3
<i>Ly49-6</i>	–	III	A	
<i>Ly49-7</i>	–	II	I	
<i>Ly49-8</i>	–	I	–	Deletion, premature stop codon in exon 1
<i>Ly49-9</i>	<i>Ly49i2</i>	V	–	Mutation, premature stop codon in exon 4 (CAG → TAG)
<i>Ly49-10</i>	<i>Ly49.19</i>	I	–	Mutation, premature stop codon in exon 1 (TCA → TGA)
<i>Ly49-11</i>	–	V	I	
<i>Ly49-12</i>	–	V	–	Mutation, premature stop codon in exon 5 (AAA → TAA)
<i>Ly49-13</i>	–	II	–	Insertion of ten nucleotides, premature stop codon in exon 1
<i>Ly49-14</i>	–	I	–	Deletion, premature stop codon in exon 2
<i>Ly49-15</i>	–	V	I	
<i>Ly49-16</i>	–	V	–	Mutation, premature stop codon in exon 5 (CGA → TGA)
<i>Ly49-17</i>	–	V	–	Mutation, premature stop codon in exon 5 (TTG → TAA)
<i>Ly49-18</i>	<i>Ly49.12</i>	I	A	
<i>Ly49-19</i>	–	V	–	Deletion of exon 6
<i>Ly49-20</i>	–	V	–	Mutation, premature stop codon in exon 4 (TAT → TAA)
<i>Ly49-21</i>	–	V	–	Deletion of exon 4
<i>Ly49-22</i>	–	IV	I	
<i>Ly49-23</i>	–	IV	I/A	
<i>Ly49-24</i>	–	IV	–	Deletion of exons 5 and 6
<i>Ly49-25</i>	–	IV	I	
<i>Ly49-26</i>	–	IV	I/A	
<i>Ly49-27</i>	–	IV	I/A	
<i>Ly49-28</i>	–	IV	I/A	
<i>Ly49-29</i>	–	IV	–	Deletion of exons 5 and 6
<i>Ly49-30</i>	–	IV	I	
<i>Ly49-31</i>	–	IV	–	Mutation, premature stop codon in exon 4 (TGG → TAG)
<i>Ly49-32</i>	–	IV	I	
<i>Ly49-33</i>	–	IV	–	Deletion of exon 1

Additional modifications were done by visual inspection. Phylogenetic analysis was conducted by using the neighbor-joining method (Saitou and Nei 1987) implemented in the computer program MEGA2 (Kumar et al. 2001). The bootstrap values were computed with 1,000 resamplings. We used Jukes–Cantor and Kimura distances for the nucleotide sequences of protein coding regions (Nei and Kumar 2000), but since these two distances generated virtually the same trees, we present only the trees obtained by Jukes–Cantor distances. (Note that the gene sequences used were closely related.) In addition, we used the parsimony method as implemented in PAUP\* with 1,000 bootstrap resamplings (TBR search) (Swofford 1998). In tree construction, all alignment gaps were eliminated (complete-deletion option). *NKG2A*, a C-type lectin-like gene, which is closely related to *Ly49* genes, was used as the out-group.

The nucleotide sequences for mouse *Ly49* genes (except *Ly49-B*) were obtained from the Mouse Genome Resources (<http://www.ncbi.nlm.nih.gov/genome/guide/mouse>). The GenBank accession number for mouse *Ly49-B* is AK017140. Two primate *Ly49* genes were also included in the analysis: Baboons (*Papio hamadryas*, AY028399) and humans (*Homo sapiens*, NM\_006611). The six known rat *Ly49* genes identified in different strains were obtained from GenBank with the following accession numbers. Strain F344: *Ly49.9*, U56863; *Ly49.12*, U56822; *Ly49.19*, U56823; *Ly49.29*, U56824. Strain PVG: *Ly49i2*, NM\_152848; *Ly49s3*, NM\_153726. The GenBank accession number for the rat *NKG2A* gene is AF021350.

## Results

Genomic organization of the gene families in the NKC region of rats

The gene families identified in the NKC regions of humans, mice, and rats are presented in Fig. 1A. Although the number of genes in rats is limited owing to the incomplete gene annotation, we observe a high level of conservation of NKC gene families with respect to the gene order and gene content among the three species, as was previously observed (Trowsdale et al. 2001). The conserved genes *NKG2A* and *CSDA* defined the boundary of the *Ly49* gene cluster.

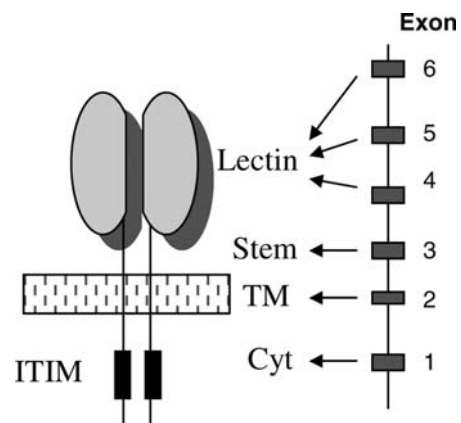
Our analysis of the genomic sequence of the rat *Ly49* gene cluster revealed a total of 33 full-length or nearly full-length genes located within an approximately 1.8-Mb region (Fig. 1B). In addition, there were several fragmentary genes containing one or a few exons, but these genes will not be considered in this paper. The number of rat *Ly49* genes is about twice as large as that of mice, which is about 15 (Fig. 1B). All rat genes are orientated with the same transcriptional direction as in the case of mouse genes. Of the 33 rat genes, five appear to be the same as those previously identified in different rat strains (Table 1). One of the previously identified genes, *Ly49s3*, did not match to any of our 33 genes probably

because different strains were used. We did not use this gene in the following study.

## Sequence and structure analysis of rat *Ly49* genes

The *Ly49* protein is a homodimer and consists of the extracellular lectin-like domain, stem, transmembrane, and cytoplasmic regions (Fig. 2). The coding region of rat *Ly49* genes is composed of six protein-coding exons and five introns as in the case of mouse genes (Takei et al. 1997). Exons 1, 2, and 3 encode the cytoplasmic, transmembrane, and stem regions, respectively. The last three exons encode the extracellular lectin-like domain, which is responsible for the interaction with ligands. The lengths of five introns vary with the gene, though the last intron (between exons 5 and 6) is generally longest. *Ly49-I* has the longest introns and the entire gene region extends over 30 kb. Excluding five incomplete genes (*Ly49-19*, *21*, *24*, *29*, and *33*), in which one or two exons are deleted, we found a total of 28 full-length genes. We assumed that the genes containing at least one stop codon in the first five exons are pseudogenes and the remaining genes are functional. Using this criterion, we found 17 full-length putative functional genes and 11 putative pseudogenes (Table 1). Therefore, the proportion of pseudogenes was 39% (11/28) among the full-length genes and 48% (16/33) among all genes including the incomplete ones. Note that there are some gaps in the supercontig, which we used to identify rat *Ly49* cluster. Hence, our estimates of the proportion of pseudogenes could be overestimated.

The mouse strain C57BL/6 is known to have a total of 15 full-length or nearly full-length *Ly49* genes, of which only four genes (*X*, *K*, *N*, and *M*) are probably non-functional (McQueen et al. 1999; Kane et al. 2001; Wilhelm et al. 2002). Therefore, the proportion of



**Fig. 2** The relationships between the exon–intron organization and protein domains in *Ly49* genes. *Cyt* Cytoplasmic domain, *TM* transmembrane domain, *Stem* stem domain, *Lectin* C-type lectin-like domain. Adapted from Takei et al. (1997). Note that the exon numbering here is different from the one commonly used for mouse *Ly49* genes because the 5'-non-coding exon (exon 1 in mouse *Ly49* genes) was not included in this organization. Here we only showed inhibitory receptors as an example. The activating receptors have the same exon–intron organization as inhibitory receptors

pseudogenes is about 27%. This mouse strain has another fragmentary gene (*L*) containing one exon (Wilhelm et al. 2002), but this gene will not be considered here. Using primers specific to two exons and one intron of known *Ly49* genes, Makrigiannis and co-workers (2002) identified 19 genes or gene segments in mouse strain 129/J, but their functionality is unknown. We therefore do not consider these genes in this paper.

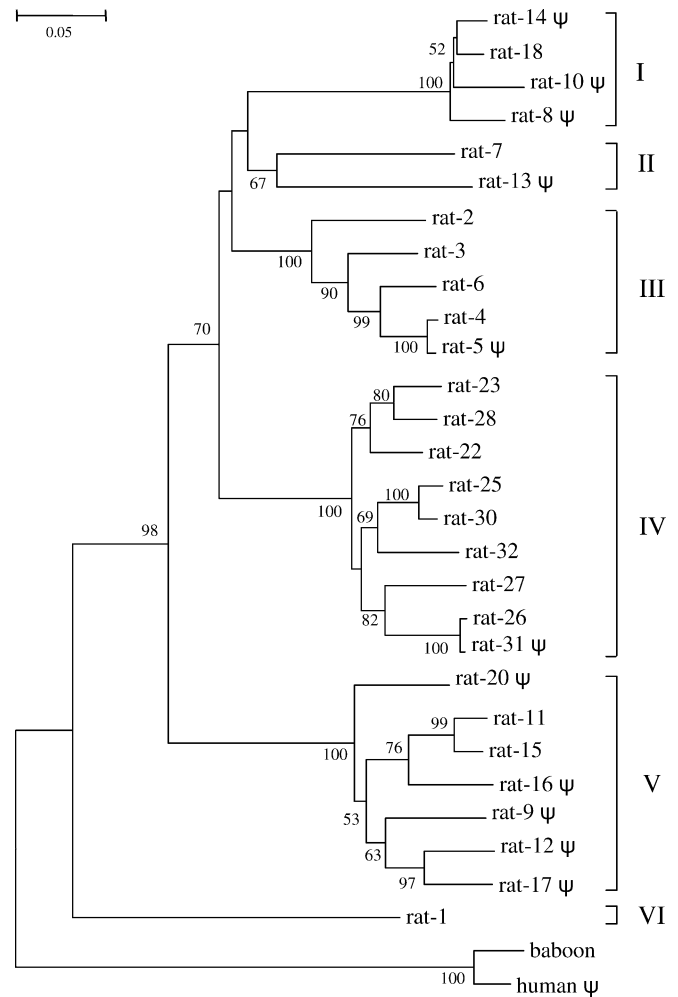
We then looked for the sequence signature of ITIM in all putative translated rat *Ly49* sequences. The results showed that ten genes are putative inhibitory receptors with ITIM (Table 1). Three genes (*Ly49-4*, 6, and 18) had no ITIM and are probably activating receptors because of the presence of arginine in the transmembrane region. Interestingly, the remaining four genes (*Ly49-23*, 26, 27, and 28) contained both the ITIM and transmembrane arginine. The function of these genes is unknown at present.

#### Evolutionary relationships of rat *Ly49* genes and genomic block duplication

To study the evolutionary relationships of rat *Ly49* genes, we constructed a phylogenetic tree of 28 full-length genes using nucleotide sequences (Fig. 3). For the out-group, we used primate *Ly49* genes. According to this tree, there are six different clades of *Ly49* genes with high bootstrap values (except for clade II). If we exclude the single gene clade VI (gene 1), the first splitting of phylogenetic groups occurred between clade V and other clades though the bootstrap value is not very high. The splitting order of the remaining clades, I–IV, is not clearly resolved. Essentially the same results were obtained by parsimony analysis. All of the five groups (I–V) include one or a few pseudogenes. According to the tree, some pseudogenes, such as genes *5ψ* and *31ψ*, were generated very recently, while some others (e.g., genes *9ψ* and *13ψ*) might have lost their function a long time ago.

As the genes were named according to their genomic location, we can easily see the relationships between the phylogenetic clades and genomic locations of the genes. If the number of *Ly49* genes increased mainly by tandem duplication, we would expect the genes which are physically clustered in the genome to form a monophyletic group. Figure 3 shows that all five genes (2–6) in clade III are closely related. This suggests that clade III genes were generated by tandem duplication.

However, other gene clusters do not necessarily show closely related phylogenetic relationships. A close examination of gene arrangement has led us to identify a genomic block duplication including four *Ly49* genes. Figure 4 shows the comparison of the two genomic blocks that were apparently generated by block duplication. The upper part of each row shows the genomic block starting with gene 23 and ending with gene 26, and the lower part represents the genomic block from gene 28 to gene 31. The genomic structures of the two blocks are virtually identical except for insertion or deletion of some repetitive

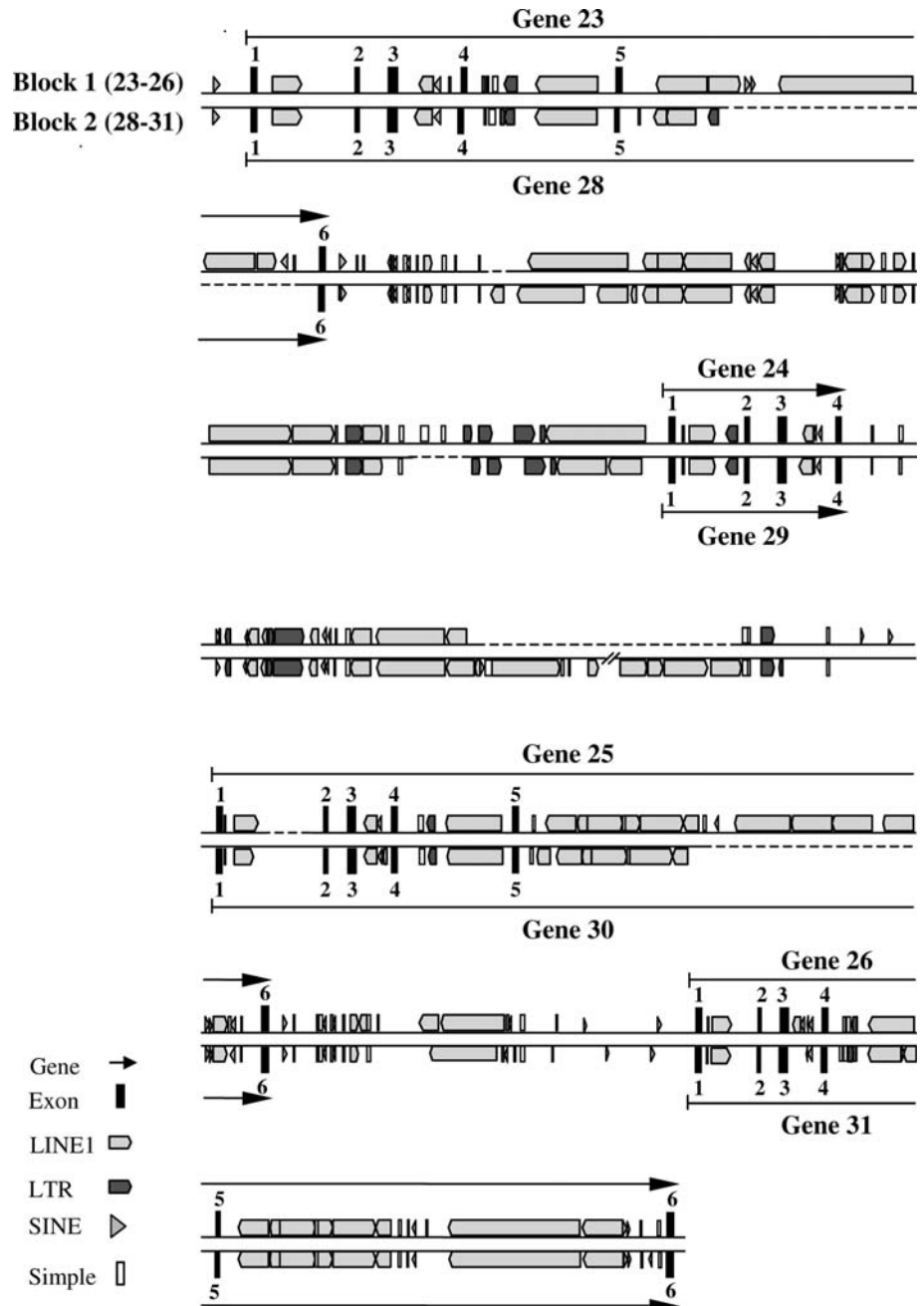


**Fig. 3** Phylogenetic tree of 28 full-length rat *Ly49* genes (790 nucleotides used). The tree was constructed by the neighbor-joining method using Jukes–Cantor distances. The bootstrap values based on 1,000 replications are shown above the branches (only values higher than 50% shown). Essentially the same tree topology was obtained by using parsimony methods. Each group is shown by a bracket, labeled with a Roman numeral. Letter  $\psi$  represents potential pseudogene

elements in the non-coding regions. The nucleotide identity between the coding regions was 95% between gene 23 and gene 28, 99% between gene 24 and gene 29, 98% between gene 25 and gene 30, and 99% between gene 26 and gene 31. These high identity values suggest that the block duplication occurred very recently.

A crude estimate of the time of occurrence of this duplication can be estimated by computing the Kimura distances between genomic blocks 1 and 2 and between the orthologous pair of mouse and rat genes (mouse-*B* and rat-*I*; see below). Since the divergence between the mouse and rat lineages ( $T_{MR}$ ) has been estimated to have occurred approximately 33 million years ago (MYA) (Nei et al. 2001), one can estimate the time of block duplication ( $T_D$ ) by  $(d_{12}/d_{MR}) \times 33$  MYA, where  $d_{12}$  is the distance estimate between genomic blocks 1 and 2 and  $d_{MR}$  is the distance between the orthologous mouse and rat genes. When we used the concatenate sequence of the coding regions of the

**Fig. 4** Comparison of the genomic organization of *Ly49* genes and repetitive elements between two recently duplicated genomic blocks. The *upper* sequence represents the first genomic block, which extends from the 5' end of the gene *Ly49-23* to the 3' end of the gene *Ly49-26*. The *lower* sequence represents the second genomic block, which extends from the 5' end of *Ly49-28* to the 3' end of *Ly49-31*. Each sequence is approximately 140 kb long. The exons of each gene and the repetitive elements were drawn roughly to scale on the genomic sequence. The *LINE* elements include two major subfamilies L1 and Lx, colored *light gray*. *LTRs* are long terminal repeat retrotransposons, including ERV and MaLR, etc. The *SINE* elements include primarily ID and B1 subfamilies. The simple repetitive elements represent various microsatellite loci



four *Ly49* genes from each of genomic blocks 1 and 2, we obtained  $d_{12}=0.025\pm 0.003$ . For  $d_{MR}$ , we obtained  $0.204\pm 0.017$ . Therefore, the genomic block duplication appears to have occurred about  $4.0\pm 0.1$  MYA.

The above examples of tandem duplication and genomic block duplication are probably not isolated cases, and it is possible that the other regions experienced these processes several times but we cannot see the trace of the occurrence anymore. Of course, it is also possible that some gene duplication has occurred by gene transposition. However, one thing is clear from the genomic maps of the *Ly49* clusters in mice and rats. That is, this gene family did not experience inverted gene duplication changing transcriptional direction, unlike many genes in other gene

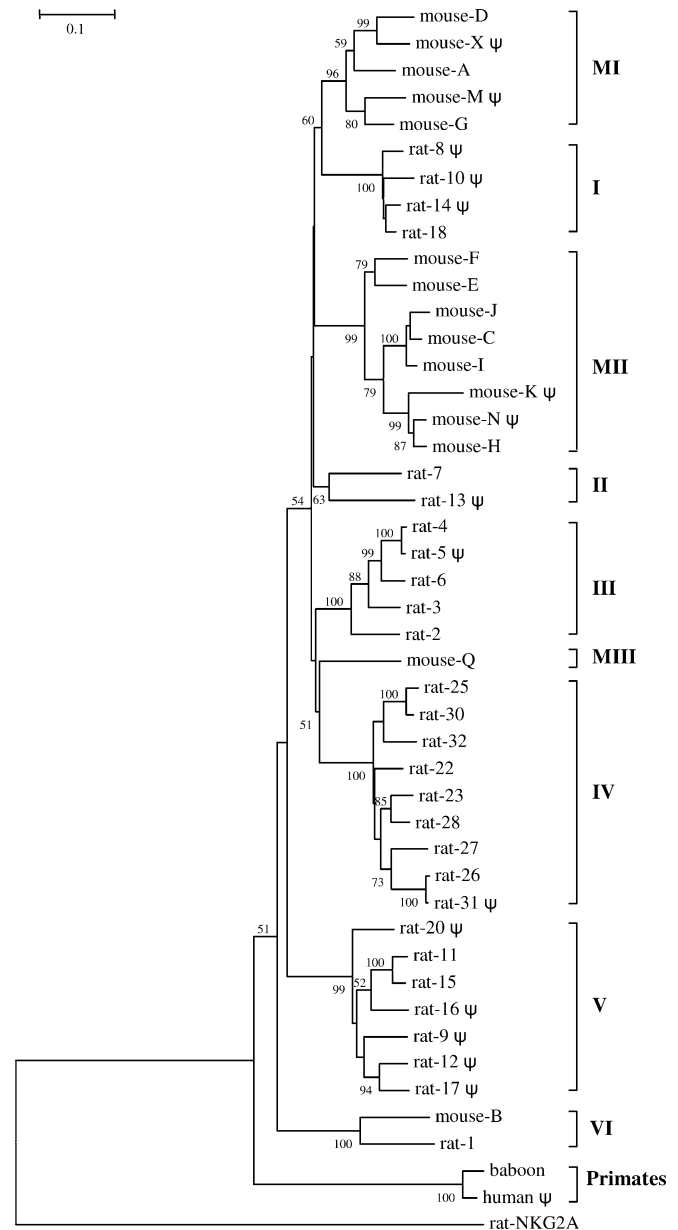
families such as the immunoglobulin and olfactory receptor gene families (Matsuda et al. 1993; Niimura and Nei 2003).

#### Evolutionary relationships of rat and mouse *Ly49* genes

To understand the evolutionary relationships of mouse and rat *Ly49* genes, we constructed a phylogenetic tree for the genes from mice, rats, baboons, and humans (Fig. 5). In this study, we used the mouse genes from strain C57BL/6 and rat genes from strain BN/SsNHsd/MCW. The root of the tree was determined by using the rat *NKG2A* gene,

which is closely related to rodent *Ly49* genes (see Fig. 1A). From this tree, we could identify nine groups of rodent genes, which are supported by relatively high bootstrap values, except for the single gene mouse clade MIII and rat clade II. Five of them (I–V) were rat-specific and are identical to those previously identified in Fig. 3. Group VI contains one mouse (gene *B*) and one rat (gene *I*) gene, and this group is supported by a bootstrap value of 100%. Both of these genes are located at the telomeric end of the *Ly49* gene cluster (Fig. 1B) and appear to be an orthologous pair. However, the branching order of the remaining eight phylogenetic clades is unclear, because the relevant bootstrap values are low. There are three mouse-specific clades (MI, MII, and MIII), and they are intermingled with rat-specific clades. This suggests that the most recent common ancestor (MRCA) of mice and rats already had several *Ly49* genes. However, the species-specific grouping suggests that the main rodent *Ly49* gene repertoires expanded rapidly after mice–rats divergence.

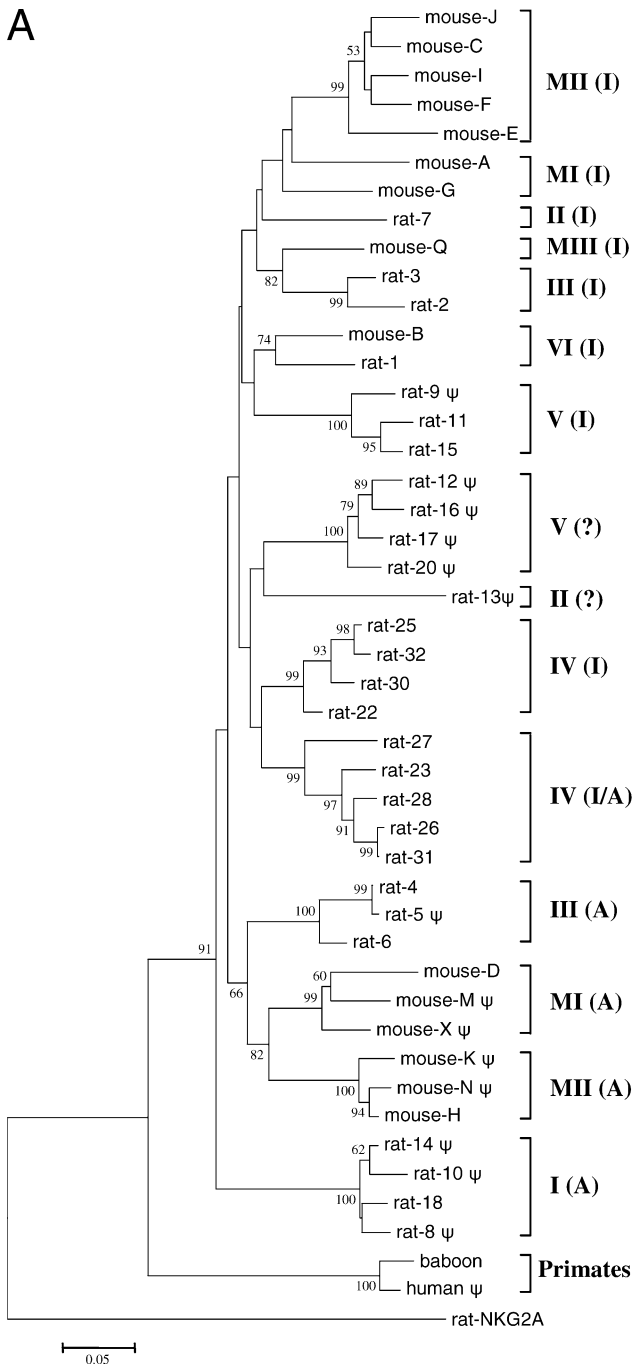
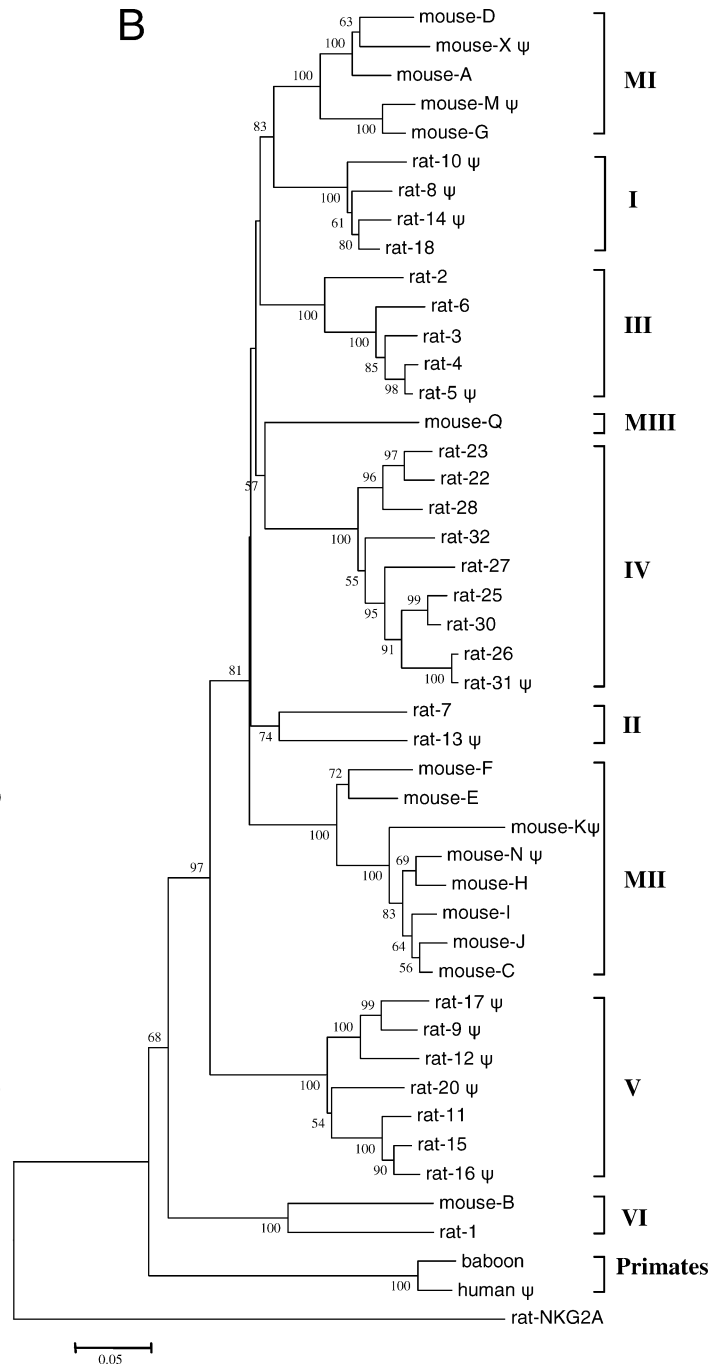
Gene conversion or recombination events might be involved in the evolution of *Ly49* genes. For example, conducting phylogenetic analysis of approximately 1.4-kb portions of intron 1 (between the first non-coding exon and exon 1) and intron 6 (between exons 5 and 6 in our definition) of the 14 mouse genes, Wilhelm and co-workers (2002) showed that the major groups of mouse genes were the same as those in Fig. 5 when the intron 6 region was used, but that a somewhat different grouping of genes was obtained when the intron 1 region was used. From this and other observations, they proposed that gene conversion or recombination occurred in the region of intron 3 (between exons 2 and 3 in our definition). However, since intron regions are often more susceptible to insertion and deletion than exon regions, we examined this hypothesis by constructing phylogenetic trees of rodent *Ly49* genes for exons 1–2 and exons 3–6, separately (Fig. 6). The major gene groups in the phylogenetic tree of extracellular region (Fig. 6B) are exactly same as those of Fig. 5 using full-length coding region. However, in the phylogenetic tree of cytoplasmic and transmembrane region (Fig. 6A), most major gene groups (except rat group I, VI, and mouse singleton gene group MIII) were divided into two subgroups. Interestingly, the two subgroups (i.e., rat group III and mouse group I, II) are often functionally different, and one subgroup includes mainly inhibitory receptors and the other represents activating receptors or pseudogenes. The *Ly49* genes with similar functionality (either inhibitory or activating) tend to cluster together. These results suggest that some degree of recombination might have occurred between extracellular and transmembrane regions in the evolution of rodent *Ly49* genes to promote the functional switch between the pair of inhibitory and activating receptors which share a highly similar lectin-binding domain, as suggested by Arase and Lanier (2004).



**Fig. 5** Phylogenetic tree of rodent and other mammalian *Ly49* genes. This tree was constructed by the neighbor-joining method using Jukes–Cantor distance for the coding regions of 15 mouse, 28 rat, two primate *Ly49* genes, and the rat *NKG2A* gene, which was used as an out-group (714 nucleotides used). Each group is shown by a bracket. The Roman numerals of rat *Ly49* gene groups are the same as that of Fig. 3. Each mouse *Ly49* group is labeled with a Roman numeral followed by the letter *M*. The letter  $\psi$  represents a potential pseudogene. The parsimony tree was virtually the same as the neighbor-joining tree if we disregard the branching patterns with low bootstrap values. Therefore, it is not presented here

#### Rates of gene expansion by duplication

Figure 5 shows that the number of *Ly49* genes in mice and rats expanded rapidly by gene duplication after divergence of the two species. To have a rough idea about how quickly the gene expansion occurred, we estimated the time of occurrence when the *Ly49* repertoire has expanded from a common ancestor. For this purpose, we used the

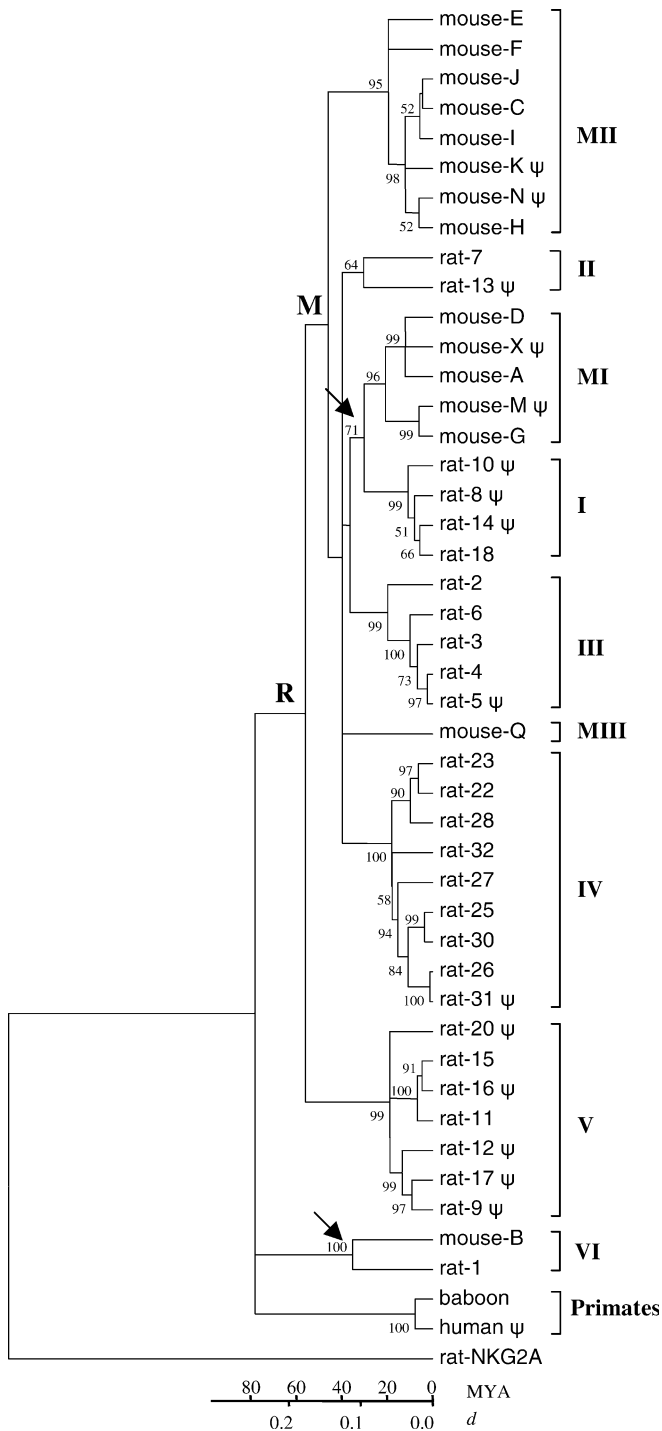
**A****B**

**Fig. 6A, B** Phylogenetic trees of rodent and primate *Ly49* genes using different region. The trees were constructed using *p*-distances. The bootstrap values based on 1,000 replications are shown *above* the branches (only values higher than 50% shown). **A** Tree based on

exons 1–2; 187 nucleotides used. (*I*) represents inhibitory receptors, (*A*) activating receptors, (?) either a pseudogene or functionally unknown. **B** Tree based on exons 3–6. Five hundred twenty-seven nucleotides used

linearized tree method (Takezaki et al. 1995). Using the branch length test of the LINTREE program (<http://mep.bio.psu.edu>), we first tested the molecular clock hypothesis and found that the mouse gene *Kψ* evolved faster than the average rate at the 1% significance level. However, elimination of this gene did not make much difference in the time estimates of branching points. We therefore constructed a linearized tree under the assumption of molecular clock using the entire set of genes except for the

out-group sequence. The linearized tree using the extra-cellular region of *Ly49* genes thus obtained is presented in Fig. 7. The timescale for this tree was obtained under the assumption that the putative orthologous mouse and rat *Ly49* gene lineages (indicated by arrows in Fig. 7) diverged about 33 MYA. The rate of nucleotide substitution was estimated to be approximately  $3.1 \times 10^{-9}$  per site per year.



**Fig. 7** Linearized tree of *Ly49* genes from rodent and primate species based on exons 3–6. The timescale was calibrated with 33 MYA under the assumption that the mouse and rat lineages diverged about 33 MYA (arrows)

Although the time estimates obtained from this tree are very crude, it appears that the mouse group MI, MII, and MIII genes were generated by repeated gene duplication from a sister gene (labeled *M*) of the mouse-*B* (also rat-*I*) gene. This sister gene apparently existed about 47 MYA, and during the last 47 million years (MY) the number of genes expanded from one to 14. Therefore, the rate of gene

expansion by duplication is  $(14-1)/47=0.28$  per gene per MY. Figure 7 suggests that the rat group I, II, III, IV, and V genes expanded from a sister gene (labeled *R*) of the rat-*I* (also mouse-*B*) to 27 duplicate genes. There are five additional incomplete genes (genes 19, 21, 24, 29, and 33) that were not included in Fig. 7 because one or two exons were missing. These genes belong to group IV or V genes (data not shown). Therefore, if we include these genes, there must have been 31 new genes generated during the last 57 MY. This gives an expansion rate of  $31/57=0.54$  per gene per MY. This rate is about two times higher than the mouse rate. We performed a similar analysis on the linearized tree of cytoplasmic and transmembrane region and the expansion rates were not significantly different (data not shown).

Table 2 shows the rates of gene expansion for some other gene families that are known to have experienced rapid gene duplication. It shows that human *KIR* genes expanded nearly at the same rate as that of the rat *Ly49* genes. MHC class I and class II genes are also known to have been subject to rapid gene duplication. Table 2, however, shows that the rates of gene expansion for these gene families in primates are less than half the rate of mouse *Ly49* genes. Of course, the estimates of MHC rates are very crude because they are based on only a few duplicate genes. The hominoid gene family *Morpheus*, whose function is unknown, has been acclaimed as the fastest evolving gene family so far identified (Johnson et al. 2001). Table 2 indicates that the rate of gene expansion for this gene family is nearly twice as high as that of rat *Ly49*.

Lynch and Conery (2003) estimated the rates of gene duplication for genome-wide genes considering recently duplicated genes (less than 2 MY old). Their estimates for animal and plant model organisms were 0.001–0.03 per gene per MY. These estimates are supposed to include duplicate genes that may soon become pseudogenes. In our computation, however, we considered a relatively longer evolutionary time and may not have included pseudogenes, which have already been deleted from the genome. Therefore, other things being equal, our estimates are expected to be smaller than Lynch and Conery's duplicate rates. Yet, the *Ly49* or *KIR* rates are much higher than those of Lynch and Conery. This suggests that the *Ly49* and *KIR* genes have duplicated much faster than the genome-wide genes. However, note that our rates are not the same as the mathematical rate of increase of genes by gene duplication (see Nei 1969). We did not use this rate here, because the actual evolutionary pattern is quite different from the mathematical model.

## Discussion

We have seen that the BN rat genome contains a total of 33 *Ly49* genes, of which 17 are putative functional genes and 16 are putative pseudogenes. The genome of mouse strain C57BL/6 is known to have 15 *Ly49* genes, of which 11 are functional and four are non-functional. These results

**Table 2** Rates of gene expansion by duplication for different gene families

Gene family	Evolutionary time (MYA)	No. of genes increased/ancestor	Expansion rate per MY	Data used
Class II MHC (human DRB)	50	4	0.08	Satta et al. (1996) and Takahashi et al. (2000)
Class I MHC (human)	50	4	0.08	Piontkivska and Nei (2003)
Human <i>KIR</i>	21	11 <sup>a</sup>	0.52	L. Hao and M. Nei, unpublished
Mouse <i>Ly49</i>	47	13	0.28	This study
Rat <i>Ly49</i>	57	31 <sup>b</sup>	0.54	This study
<i>Morpheus</i> (human)	13	13	1.00	Johnson et al. (2001)
Genome-wide genes			0.001–0.03 <sup>c</sup>	Lynch and Conery (2003)

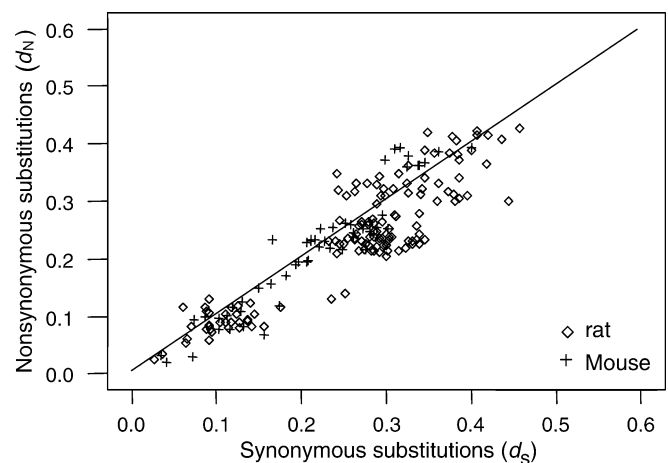
<sup>a</sup>This number was obtained by considering only KIR domain D1.

<sup>b</sup>Five incomplete genes (19, 21, 24, 29, and 33) are not included in the tree of Fig. 6. However, since genes 19 and 21 belong to group V, and the other three genes belong to group IV, these five genes were added.

<sup>c</sup>Gene duplication rates were obtained from *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Arabidopsis thaliana*. Strictly speaking, these rates are not the same as the gene expansion rate (see text).

suggest that the *Ly49* gene family has been subject to birth-and-death evolution and the birth and death rates of genes are higher in rats than in mice. These estimates are very crude, but if this is true, why should rats have a higher rate? One possible answer to this question is the difference in functional constraints that may exist between mice and rats. In the case of the MADS-box gene family, which controls the development of flowers and other characters in plants, the type-II gene subfamily is subject to stronger functional constraints than the type-I gene subfamily and shows a slower rate of birth-and-death evolution (Nam et al. 2004). To examine whether a similar mechanism is involved in the evolution of *Ly49* genes, we examined the rates of synonymous and non-synonymous nucleotide substitution in mice and rats separately using the method of Zhang and co-workers (1998b). Contrary to our expectation, however, the rates of synonymous and non-synonymous substitution were virtually the same for both mice and rats (Fig. 8). This finding also suggests that *Ly49* genes have evolved essentially in a neutral fashion at the nucleotide level. It is then possible that the higher rate of birth-and-death evolution in rats is merely due to the stochastic errors of the birth-and-death process of genes. Note that our analysis of rat *Ly49* gene cluster is based on a single haplotype (rat strain BN/SsNHsd/MCW). It is possible that rat *Ly49* genes have multiple haplotypes, as in the case of mouse *Ly49* genes. It will be interesting to explore the *Ly49* gene cluster in other rat strains and compare its haplotype diversity between mice and rats.

The *Morpheus* gene family, which has the highest gene duplication rate in Table 2, is known to have a rate of non-synonymous substitution ( $r_N$ ) significantly higher than the rate of synonymous substitution ( $r_S$ ), and this was taken as evidence that the rate of gene duplication has been accelerated by natural selection (Johnson et al. 2001). This is a reasonable explanation. However, this explanation may not be the whole story, because the gene duplication rate can be enhanced substantially without a significant increase of  $r_N$  relative to  $r_S$  as in the case of rat *Ly49* genes. Note that the gene duplication rate is sometimes



**Fig. 8** Relationships between the number of synonymous substitutions per synonymous site ( $d_S$ ) and the number of non-synonymous substitutions per non-synonymous site ( $d_N$ ) for mouse ( $n=11$ ) and rat ( $n=17$ ) putative functional *Ly49* genes

influenced substantially by genomic block duplication (e.g., Schable and Zachau 1993; Su and Nei 2001).

As mentioned earlier, the *KIR* genes in primates are structurally very different from the *Ly49* genes in rodents, but they have the same function. The number of member genes in the *KIR* family in primates is roughly similar to that of the *Ly49* family in mice, and the primate *KIR* genes have also evolved very rapidly by repeated gene duplication (L. Hao and M. Nei, unpublished). The similarity of the evolutionary patterns of *KIR* and *Ly49* genes suggests that their rapid evolution is caused by the function of the genes, that is, the interaction with MHC class I molecules, rather than by the protein structure encoded by the genes. It is interesting to note that the number of functional genes in the *KIR* gene family in humans and chimpanzees is about ten though it depends on haplotypes (Hsu et al. 2002), and this number is only slightly lower than that of the rodent *Ly49* gene family. This finding reminds us that the number of highly polymorphic MHC class I loci is only a few (often one to four), though this gene family contains a large number of genes, including pseudogenes

(Klein and Figueroa 1986; Anzai et al. 2003). It is possible that there is some kind of upper limit for the number of functional *KIR* or *Ly49* genes caused by interaction with MHC molecules.

The origins of *KIR* genes in primates and *Ly49* genes in rodents are quite mysterious. The molecular structures of *KIR* and *Ly49* receptor molecules are so different that it is difficult to know how they originated. However, it is possible to speculate how these different molecules have come to be used in primates and rodents separately. Actually, although *Ly49* genes are concentrated primarily in rodents, the human and baboon genomes contain at least one *Ly49* gene, whether they are functional or not (Fig. 1A). Similarly, the mouse genome is known to have at least two *KIR*-like genes (Hoelsbrekken et al. 2003; Welch et al. 2003). Furthermore, multiple *KIR* genes and a single *Ly49* gene have been identified in cattle (McQueen et al. 2002; Storset et al. 2003). It is therefore quite likely that the common ancestor of primates and rodents used both *KIR* and *Ly49* receptors. This view is supported by the fact that *KIR*-related genes, such as *ILT* genes in humans and *PIR* in mice, and *Ly49*-related genes, such as *NKG2* genes, are used in both primates and rodents (Trowsdale et al. 2001). After the primate and rodent lineages diverged, however, *KIR* genes apparently happened to be predominant in the primate lineage, and *Ly49* genes in the rodent. This differential use of NK cell receptors may have happened by chance or for some adaptive reasons. It is also likely that once one type of NK cell receptors became predominant in an evolutionary lineage, it was probably more efficient to produce them exclusively and the other type of receptors therefore gradually ceased to be used.

At the present time, of course, this is merely a hypothesis. However, this hypothesis can be tested by examining the presence or absence of *KIR* or *Ly49* genes in other orders of placental mammals, marsupials, or even birds and reptiles and studying whether they are expressed as NK cell receptors.

**Note added in proof** After submission of this paper, we learned that Wilhelm and Mager (2004) also identified rat *Ly49* genes. They reported 36 rat *Ly49* genes which are different from ours. This difference might have occurred because they did not use the updated rat genome sequence.

**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

- Anzai T, Shiina T, Kimura N et al (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc Natl Acad Sci USA* 100:7708–7713
- Arase H, Lanier LL (2004) Specific recognition of virus-infected cells by paired NK receptors. *Rev Med Virol* 14:83–93
- Brennan J, Mager DL, Jefferies WA, Takei F (1996) Recognition of class I major histocompatibility complex molecules by Ly-49: specificities and domain interactions. *J Exp Med* 183:1553–1559
- Brown MG, Fulmek S, Matsumoto K, Cho R, Lyons PA, Levy ER, Scalzo AA, Yokoyama WM (1997) A 2-Mb YAC contig and physical map of the natural killer gene complex on mouse chromosome 6. *Genomics* 42:16–25
- Brown MG, Dokun AO, Heusel JW, Smith HRC, Beckman DL, Blattenberg EA, Dubbelde CE, Stone LR, Scalzo AA, Yokoyama WM (2001) Vital involvement of a natural killer cell activation receptor in resistance to viral infection. *Science* 292:934–937
- Colonna M, Samaridis J (1995) Cloning of immunoglobulin-superfamily members associated with HLA-C and HLA-B recognition by human natural killer cells. *Science* 268:405–408
- Depatie C, Lee SH, Stafford A, Avner P, Belouchi A, Gros P, Vidal SM (2000) Sequence-ready BAC contig, physical, and transcriptional map of a 2-Mb region overlapping the mouse chromosome 6 host-resistance locus *Cmv1*. *Genomics* 66:161–174
- Dissen E, Ryan JC, Seaman WE, Fossum S (1996) An autosomal dominant locus, Nka, mapping to the Ly-49 region of a rat natural killer (NK) gene complex, controls NK cell lysis of allogeneic lymphocytes. *J Exp Med* 183:2197–2207
- Dorfman JR, Raulet DH (1996) Major histocompatibility complex genes determine natural killer cell tolerance. *Eur J Immunol* 26:151–155
- Gagnier L, Wilhelm BT, Mager DL (2003) *Ly49* genes in non-rodent mammals. *Immunogenetics* 55:109–115
- Gardiner CM, Guethlein LA, Shilling HG, Pando M, Carr WH, Rajalingam R, Vilches C, Parham P (2001) Different NK cell surface phenotypes defined by the DX9 antibody are due to *KIR3DL1* gene polymorphism. *J Immunol* 166:2992–3001
- Guethlein LA, Flodin LR, Adams EJ, 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
- Hershberger KL, Shyam R, Miura A, Letvin NL (2001) Diversity of the killer cell Ig-like receptors of Rhesus Monkeys. *J Immunol* 166:4380–4390
- Hoelsbrekken SE, Nylenna O, Saether PC, Slettedal IO, Ryan JC, Fossum S, Dissen E (2003) Cutting edge: molecular cloning of a killer cell Ig-like receptor in the mouse and rat. *J Immunol* 170:2259–2263
- Hsu K, Liu XR, Selvakumar A, Mickelson E, O'Reilly RJ, Dupont B (2002) Killer Ig-like receptor haplotypes analysis by gene content: evidence for genomic diversity with a minimum of six basic framework haplotypes, each with multiple subsets. *J Immunol* 169:5118–5129
- Idris AH, Smith HR, Mason LH, Ortaldo JR, Scalzo AA, Yokoyama WM (1999) The natural killer gene complex genetic locus Chok encodes Ly-49D, a target recognition receptor that activates natural killing. *Proc Natl Acad Sci USA* 96:6330–6335
- Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE (2001) Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413:514–515
- Kane KP, Silver ET, Hazes B (2001) Specificity and function of activating Ly-49 receptors. *Immunol Rev* 181:104–114
- Klein J, Figueroa F (1986) Evolution of the major histocompatibility complex. *Crit Rev Immunol* 6:295–386
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Kwon D, Chwae YJ, Choi IH, Park JH, Kim SJ, Kim J (2000) Diversity of the p70 killer cell inhibitory receptor (KIR3DL) family members in a single individual. *Mol Cells* 10:54–60
- Lynch M, Conery JS (2003) The evolutionary demography of duplicate genes. *J Struct Funct Genomics* 3:35–44

- Mager DL, McQueen KL, Wee V, Freeman JD (2001) Evolution of natural killer cell receptors: coexistence of functional *Ly49* and *KIR* genes in baboons. *Curr Biol* 11:626–630
- Makrigiannis AP, Pau AT, Schwartzberg PL, McVicar DW, Beck TW, Anderson SK (2002) A BAC contig map of the *Ly49* gene cluster in 129 mice reveals extensive differences in gene content relative to C57BL/6 mice. *Genomics* 79:437–444
- Martin A, Freitas EM, Witt CS, Christiansen FT (2000) The genomic organization and evolution of the natural killer immunoglobulin-like receptor (*KIR*) gene cluster. *Immunogenetics* 51:268–280
- Mason LH, Anderson SK, Smith HR, Yokoyama WM, Winkler-Pickett R, Ortaldo JR (1996) The *Ly49D* receptor activates murine natural killer cells. *J Exp Med* 184:2119–2128
- Mason LH, Gosselin P, Anderson SK, Fosler WE, Ortaldo JR, McVicar DW (1997) Differential tyrosine phosphorylation of inhibitory versus activating *Ly49* receptors and their recruitment of SHP-1 phosphatase. *J Immunol* 159:4187–4196
- Matsuda F, Shin EK, Nagaoka H et al (1993) Structure and physical map of 64 variable segments in the 3' 0.8-megabase region of the human immunoglobulin heavy-chain locus. *Nat Genet* 3:88–94
- McQueen KL, Lohwasser S, Takei F, Mager DL (1999) Expression analysis of new *Ly49* genes: most transcripts of *Ly49j* lack the transmembrane domain. *Immunogenetics* 49:685–691
- McQueen KL, Wilhelm BT, Harden KD, Mager DL (2002) Evolution of NK receptors: a single *Ly49* and multiple *KIR* genes in the cow. *Eur J Immunol* 32:810–817
- Nam J, Kim J, Lee S, An G, Ma H, Nei M (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc Natl Acad Sci USA* 101:1910–1915
- Naper C, Hayashi S, Joly E, Butcher GW, Rolstad B, Vaage JT, Ryan JC (2002a) *Ly49i2* is an inhibitory rat natural killer cell receptor for an MHC class Ia molecule (RT1-A1c). *Eur J Immunol* 32:2031–2036
- Naper C, Hayashi S, Kveberg L, Niemi EC, Lanier LL, Vaage JT, Ryan JC (2002b) *Ly49-s3* is a promiscuous activating rat NK cell receptor for nonclassical MHC class I-encoded target ligands. *J Immunol* 169:22–30
- Nei M (1969) Gene duplication and nucleotide substitution in evolution. *Nature* 221:40–42
- Nei M, Kumar S (2000) *Molecular evolution and phylogenetics*. Oxford University Press, Oxford
- Nei M, Xu P, Glazko G (2001) Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc Natl Acad Sci USA* 98:2497–2502
- Niimura Y, Nei M (2003) Evolution of olfactory receptor gene clusters in the human genome. *Proc Natl Acad Sci USA* 100:12235–12240
- Piontkivska H, Nei M (2003) Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Mol Biol Evol* 20:601–609
- Rat Genome Sequencing Project Consortium (RGSC) (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521
- Renedo M, Arce I, Montgomery K, Roda-Navarro P, Lee EE, Kucherlapati R, Fernandez-Ruiz E (2000) A sequence-ready physical map of the region containing the human natural killer gene complex on chromosome 12p12.3–p13.2. *Genomics* 65:129–136
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Satta Y, Mayer WE, Klein J (1996) Evolutionary relationship of *HLA-DRB* genes inferred from intron sequences. *J Mol Evol* 42:648–657
- Schable KF, Zachau HG (1993) The variable genes of the human immunoglobulin k locus. *Biol Chem Hoppe Seyler* 374:1001–1022
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W (2000) PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res* 10:577–586
- Storset AK, Slettedal IO, Williams JL, 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
- Su C, Nei M (2001) Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. *Mol Biol Evol* 18:503–513
- Swofford DL (1998) PAUP: phylogenetic analysis using parsimony. Sinauer, Sunderland, Mass.
- Takahashi K, Rooney AP, Nei M (2000) Origins and divergence times of mammalian class II MHC gene clusters. *J Hered* 19:198–204
- Takahashi T, Yawata M, Raudsepp T, Lear TL, Chowdhary BP, Antczak DF, Kasahara M (2004) Natural killer cell receptors in the horse: evidence for the existence of multiple transcribed *Ly49* genes. *Eur J Immunol* 34:773–784
- Takei F, Brennan J, Mager DL (1997) The *Ly-49* family: genes, proteins and recognition of class I MHC. *Immunol Rev* 155:67–77
- Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* 12:823–833
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface, flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882
- Trowsdale J, Barten R, Haude A, Stewart CA, Beck S, Wilson MJ (2001) The genomic context of natural killer receptor extended gene families. *Immunol Rev* 181:20–38
- Welch AY, Kasahara M, Spain LM (2003) Identification of the mouse killer immunoglobulin-like receptor-like (*Kir1*) gene family mapping to chromosome X. *Immunogenetics* 54:782–790
- Wilhelm BT, Gagnier L, Mager DL (2002) Sequence analysis of the *Ly49* cluster in C57BL/6 mice: a rapidly evolving multigene family in the immune system. *Genomics* 80:646–661
- Wilhelm BT, Mager DL (2004) Rapid expansion of the *Ly49* gene cluster in rat. *Genomics* 84:218–221
- Wilson MJ, Torkar M, Haude A, Milne S, Jones T, Sheer D, Beck S, Trowsdale J (2000) Plasticity in the organization and sequences of human *KIR/ILT* gene families. *Proc Natl Acad Sci USA* 97:4778–4783
- Zhang Z, Schaffer AA, Miller W, Madden TL, Lipman DJ, Koonin EV, Altschul SF (1998a) Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res* 26:3986–3990
- Zhang J, Rosenberg HF, Nei M (1998b) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA* 95:3708–3713