Genes and Immunity (2005) 6, 481–492

www.nature.com/gene

FULL PAPER

Complete elucidation of a minimal class I MHC natural killer cell receptor haplotype

SK Anderson[1], K Dewar[2,3], M-L Goulet[4], G Leveque[3] and AP Makrigiannis[4,5,6,7]

[1]Basic Research Program, SAIC-Frederick, National Cancer Institute-Frederick, Frederick, MD, USA; [2]Department of Human Genetics, McGill University, Montréal, QC, Canada; [3]McGill University and Genome Québec Innovation Centre, Montréal, QC, Canada; [4]Laboratory of Molecular Immunology, IRCM, Montréal, QC, Canada; [5]Department of Medicine, Université de Montréal, Montréal, QC, Canada; [6]Department of Medicine, Division of Experimental Medicine, McGill University, Montréal, QC, Canada; [7]Department of Microbiology and Immunology, McGill University, Montréal, QC, Canada

The BALB/c inbred mouse is widely used in models of infectious disease, transplantation, and cancer. The differences in the immune responses of BALB/c compared to C57BL/6 mice are especially valuable for the identification of immune regulation genes. One striking immune variance between these mice is in the function of natural killer (NK) cells, and there is strong evidence implicating differential expression of Ly49 genes. In this study, the complete BALB/c Ly49 gene cluster has been sequenced and found to contain six functional genes and two pseudogenes. Compared to C57BL/6 mice, there is a 200 kb region absent in the BALB/c cluster including a complete lack of Ly49h-related genes, which explains the increased susceptibility of BALB/c to cytomegalovirus infection. In addition, there is no BALB/c Ly49d allele, explaining the inability of BALB/c NK cells to kill certain tumor cells. The Ly49 region has now been sequenced in three different inbred mouse strains, and comparisons indicate that the evolution of each haplotype is not straightforward and has involved large-scale deletions/insertions, gene recombination, and unequal crossing over between divergent haplotypes. This study confirms that relatively small murine class I MHC receptor haplotypes exist, analogous to observations made of human killer cell Ig-like receptor gene haplotypes.

Genes and Immunity (2005) **6,** 481–492. doi:10.1038/sj.gene.6364232; published online 2 June 2005

Keywords: *Ly49; natural killer cell; haplotype; comparative genomics; BALB/c*

Introduction

The BALB/*c* mouse is an important animal model for immunologists and has been widely used for the dissection of natural killer (NK) cell function in tumor surveillance, transplantation, and infectious disease studies. NK cells from BALB/*c* mice are especially informative due to their differing functional capabilities as compared to NK cells derived from other strains such as C57BL/6 (B6) mice. Early reports identifying the receptors that affect the ability of murine NK cells to spare or destroy potential target cells were performed mainly in B6 mice, although a smaller number of studies were conducted with other mice, including BALB/*c*. In mice the *Ly49* gene family codes for inhibitory and activating receptors that largely regulate NK cell function by the recognition of self (class I MHC) and non-self (pathogen-derived ligands), respectively. Restriction fragment length polymorphism (RFLP) analysis of the *Ly49a* gene produced the first evidence that the genomic structure, and possibly content, of the *Ly49* gene families is different in B6 and BALB/*c* mice.[1,2]

With respect to surface receptor phenotype, BALB/*c* and B6 NK cell subsets share some NK cell-specific mAb staining since they both react with the YE1/48 (Ly49A), 5E6 (Ly49C/I), and 4D11 (Ly49G) antibodies.[3] In contrast, B6 and not BALB/*c* NK cells are $4E5^+$ (Ly49D), $NK1.1^+$ (Nkrp1c), and also possess a $5E6^-1F8^+$ (Ly49H) population.[3–5] Cloning and sequencing of *Ly49* cDNAs showed that BALB/*c* NK cells share at least three *Ly49*-related sequences or alleles with B6 mice: *Ly49a*, *c*, and *g*, although the alleles have diverged and contain a significant number of differences, resulting in several amino-acid substitutions.[6–8] The novel activating Ly49L receptor was also found to be expressed in BALB/*c* mice, but was first identified in the *Ly49* haplo-identical CBA and C3H mouse strains.[8,9] The finding of Ly49L supported the contention that all activating receptors have an inhibitory counterpart with a highly similar

Correspondence: AP Makrigiannis, Laboratory of Molecular Immunology, Institut de Recherches Cliniques de Montréal, Rm. 1340, 110 avenue des Pins Ouest, Montréal, QC, H2W 1R7, Canada 514-987-5630.
E-mail: makriga@ircm.qc.ca

Animal care was provided in accordance with the procedures outlined in 'A Guide for the Care and Use of Laboratory Animals' (National Institutes of Health Publication No. 86-23, 1985). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Received 1 March 2005; revised and accepted 8 April 2005; published online 2 June 2005

extracellular domain that has probably arisen by gene hybridization between two different *Ly49* genes. For example, Ly49L has an extracellular domain highly similar to the inhibitory Ly49G.

Ligand identification studies using soluble class I MHC tetramers have shown that the specificities of the B6 and BALB/*c* alleles of Ly49A are similar, but differential killing of H-2D^{d+} targets is seen when comparing B6 *vs* BALB/*c* Ly49A^{+} NK cell cytotoxicity.[10] Similar to Ly49A, B6 and BALB/*c* alleles of Ly49C were indistinguishable in their broad MHC specificity in a cell-binding assay.[11] Ligand specificities of the Ly49G alleles were assessed in RNK cells transfected with chimeric activating receptors expressing the extracellular domains of Ly49G^{B6} or G^{BALB} attached to an activating cytoplasmic domain. These experiments demonstrated that both allele products recognized H-2D^{d}, but Ly49G-^{BALB} also displayed specificity for H-2D^{k}.[12] In a recent study, the affinity of Ly49G^{BALB} was found to be greater than Ly49G^{B6} for H-2D^{d} when allele-sorted NK cells from (B6 $\times$ BALB)F$_1$ mice were used in cytotoxicity assays.[13] Therefore, despite allelic divergence, key residues for specific MHC molecule interaction are maintained and ligand specificity is similar, if not identical, among some Ly49 alleles of B6 and BALB/*c* mice.

While RFLP studies indicated that the genetic complexity of the *Ly49* cluster was very different between B6 and BALB/*c* strains, there was no direct evidence to support this possibility, since three of the known BALB/*c* *Ly49* genes (*Ly49a*, *c*, and *g*) were very similar to their B6 counterparts and the fourth, *Ly49l*, showed close homology to the *Ly49m*^{B6} pseudogene. In order to assess the genetic content of the BALB/*c* mouse, a *Ly49* gene-mapping study was recently conducted and showed that the BALB/*c* *Ly49* gene repertoire is much smaller than the B6 cluster. The BALB/*c* mapping study provided evidence of only nine predicted genes: α, *Ly49a*, *c*, *l*, *g*, *i*, *x*, *e*, and *q*.[14] Further, the cDNA sequences for the BALB/*c* *Ly49e* and *q* genes were found to be highly similar to their B6 counterparts. The gene-mapping study also detected *Ly49i* and *x*-like exon fragments. Although the previous mapping effort was rigorous, it was not possible to know if all *Ly49* genes in the BALB/*c* mouse had been detected. In this study, the total sequence of the BALB/*c* *Ly49* cluster is revealed along with the full complement of BALB/*c* *Ly49* genes and their coding sequences. The elucidation of the BALB/*c* *Ly49* repertoire provides essential information on the entire possible contribution of *Ly49* genes during NK cell functional studies in this mouse strain. Furthermore, the complete BALB/*c* *Ly49* sequence contributes greatly to the understanding of the evolution of this interesting gene family.

Results

Sequencing of the BALB/*c* Ly49 cluster

The nearly complete genomic sequence from the BALB/*c* inbred mouse of the region on mouse chromosome 6 from $\sim$20 kb 5′ of the α pseudogene fragment until just beyond exon 5 of *Ly49q* was determined. An earlier gene-mapping effort that characterized the *Ly49* content of the BALB/*c* mouse provided multiple BAC clones containing different combinations of *Ly49* genes. Two clones from that prior study, CHORI28-340f19 and CHORI28-347i8, were chosen for complete sequencing in the present study. Purified DNA from each BAC was mechanically sheared and subcloned into pUC19 plasmids. Plasmid subclone libraries were then sequenced to approximately seven-fold coverage for each BAC. Residual captured sequence gaps in the *Ly49* region were elucidated following a single round of primer walking. The assembled sequence lengths were 147 652 and 211 097 bp for BACs CH28-340f19 and CH28-347i8, respectively. By pulsed-field gel electrophoresis, BAC clones CH28-340f19 and CH28-347i8 were previously estimated to be 143 and 203 kb, respectively.[14] Merging the two BAC sequences produced a regional BALB/*c* *Ly49* cluster sequence of 354876 bp. Based on assembly estimations of gap sizes and comparisons with the homologous regions in the B6 and 129S6/SvEvTac (129S6) *Ly49* cluster sequences, the estimated total missing sequence in the uncaptured gaps was approximately 15 kb. As the missing sequence is distributed across multiple gaps, and the size of the smallest known functional *Ly49* gene is 18 kb, it is highly unlikely that any *Ly49* genes remain undetected in the BALB/*c* regional sequence.

In the region sequenced, nine *Ly49*-related genes were identified with a relative order of: α, *Ly49a*, *c*, *l*, *g*, *i*, *y*, *e*, and *q*. This order and the number of genes confirm the gene content and organization predicted during the previous gene-mapping study for this cluster.[14] Like the B6 haplotype, it was predicted that the γ pseudogene fragment also exists in BALB/*c* mice, but this gene was not present on either of the BACs sequenced in this study. The *Ly49x*-like gene previously detected in the BALB/*c* genome was renamed *Ly49y*, as it has diverged considerably compared to *Ly49x*^{B6} (discussed below). A complete *Ly49i*-related gene was also identified as previously predicted. A BALB/*c* *Ly49b* cDNA has been deposited in GenBank (AF253057), but that gene was not present in the sequenced region; therefore it is likely that, similar to the B6 genome, *Ly49b* in BALB/*c* mice is located far from the main gene cluster.

Direct sequence comparison to the C57BL/6 Ly49 cluster

Direct comparison of the BALB/*c* and B6 *Ly49* cluster by percent identity plot (PIP[15]) analysis using single coverage revealed that the two haplotypes are highly related (Figure 1). Almost all regions of the BALB/*c* cluster have matches of 95% or greater sequence identity to the B6 cluster. The only regions that have no match are the BALB/*c*-specific LINE1 motifs in *Ly49a*-intron 6 and *Ly49i*-intron 2. The counterpart of the first 11 kb of the BALB/*c* sequence was not included in the B6 sequence and so has no match. From exon 3 of *Ly49i* to just beyond exon 6 of *Ly49y*, there is much less homology to the B6 sequence relative to the remaining BALB/*c* sequence. Surprisingly, most of the *Ly49l* gene had a very strong match (95–100%) to the B6 cluster, despite the fact that *Ly49l* is only represented in the B6 genome by a single exon (exon 7). The B6 *Ly49m* pseudogene appears to be responsible for this close homology match.

The most closely related regions in the two clusters that account for the high PIP scores are indicated in a dotplot comparison of the two regions (Figure 2). There is a very strong indication of homology from the α fragment to almost the end of *Ly49a*. The break in identity at the end

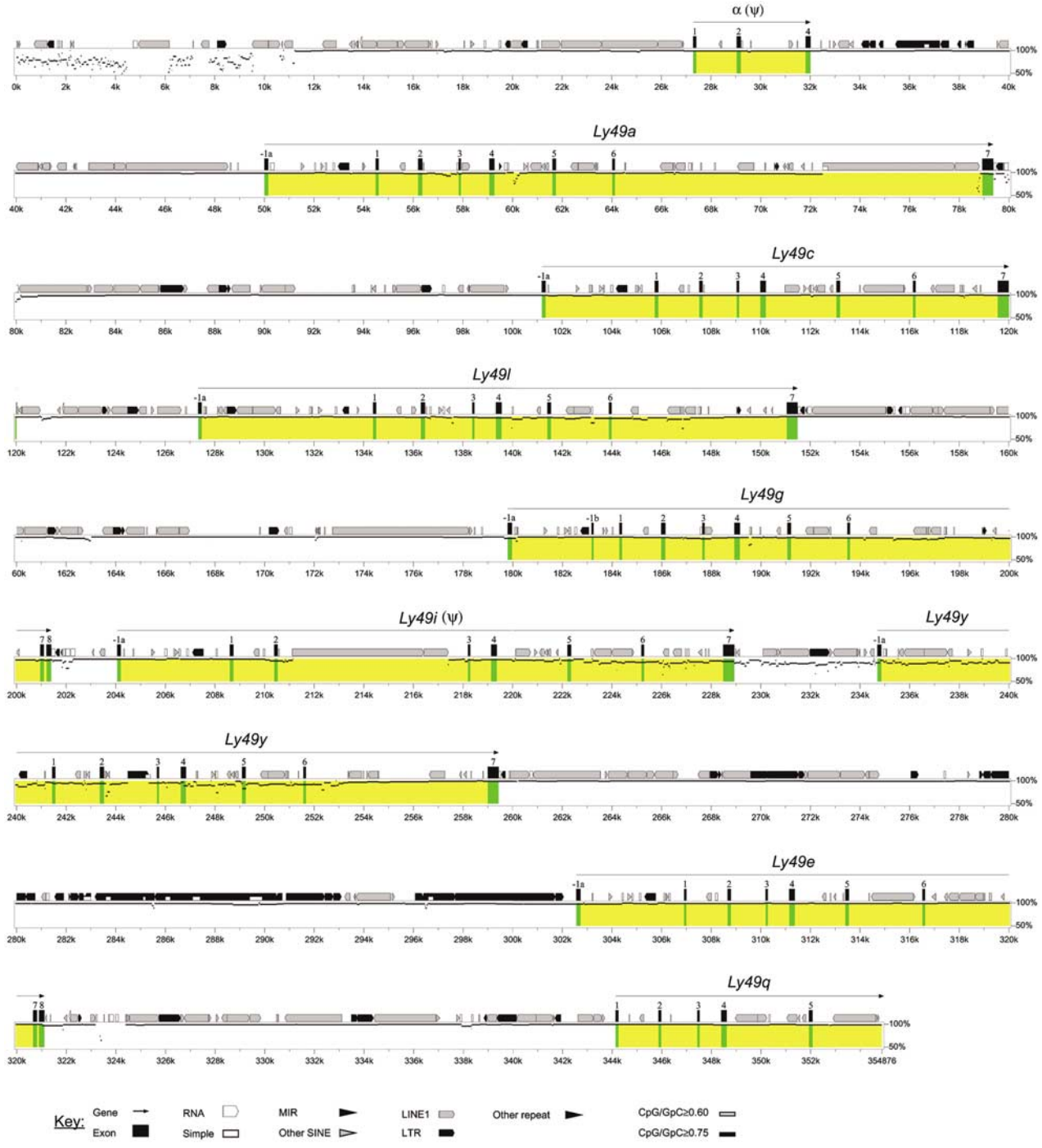

Figure 1 Organization of the BALB/*c Ly49* gene cluster. A 354 kb sequence of the BALB/*c Ly49* cluster is represented in graphical form using Pipmaker. The best match to the B6 *Ly49* cluster (single coverage) is shown in the rectangular area below the exon/repeat line and is represented as a line graph with % identity on the *Y*-axis. The length of the sequence is shown below the plot in kilobases. Exons 6 and 7 of *Ly49q* exist in the BALB/*c* genome, but were not present on the sequenced BAC clones. Regions that contain genes are marked in yellow and exons in green. The name of the gene is given above the yellow area and putative pseudogenes are denoted with a (ψ). See key for symbols of different types of repeats and other sequence elements that were identified by Repeatmasker.

of the *Ly49a* gene is caused by a LINE1 motif not present in the B6 *Ly49a* gene. After this small disruption, there is a very good match that includes all of *Ly49c* and *l*. The B6 counterpart genes in this region giving the match are *Ly49c* and *m*. Based on the PIP and dotplot results, the conserved locus organization and intergenic spacing, as well as the previously noted high cDNA and amino-acid homology, these data collectively suggest that *Ly49m*^{B6} and *Ly49l*^{BALB} can be considered as alleles.

Figure 2 also shows that the B6 and BALB/*c Ly49g* gene and upstream intergenic region are very well conserved. Surprisingly, *Ly49i*^{BALB} does not show strong

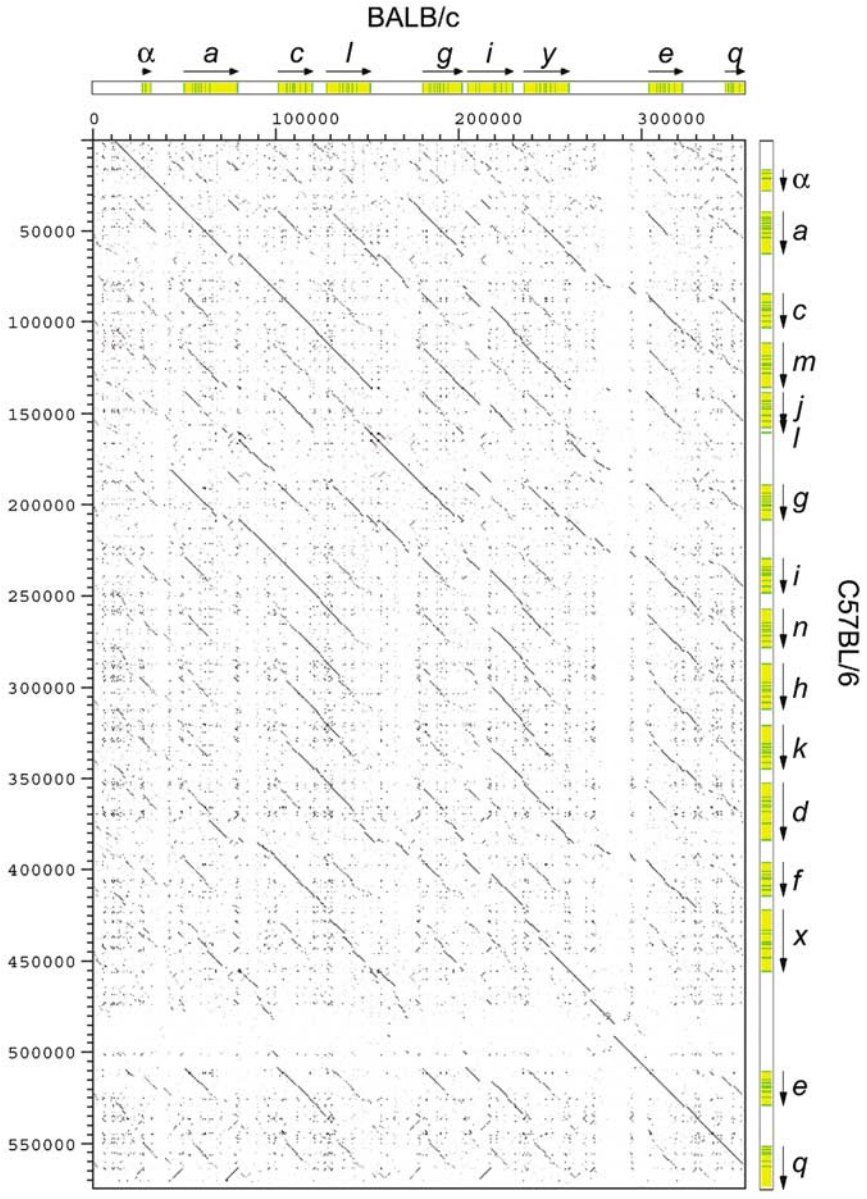

Figure 2 Direct sequence comparison between two divergent *Ly49* gene clusters. The BALB/*c* and B6 *Ly49* gene clusters were directly compared at the nucleotide level using the Dotter program. Diagonal lines represent regions of sequence homology. The location of genes (including exons) is shown for each of the haplotypes. The scale on both axes is in basepairs. Breaks in continuity of the homology lines indicate the locations of haplotype-specific deletions/insertions such as repetitive elements.

homology with *Ly49i*^{B6}. In fact *Ly49i*^{B6} has a much stronger and longer match to *Ly49c*^{BALB}. Beyond this, the homology matches break down and only short matches are evident due to the high overall relatedness of all *Ly49* genes. Strong conservation is only again observed for the *Ly49e* and *q* region. *Ly49x*^{B6} shows a weak match for *Ly49y*^{BALB}, again arguing for their separation as distinct (albeit related) genes. The only other available fully sequenced murine *Ly49* region is from 129S6 mice, and dotplot comparison of the BALB/*c* and 129S6 *Ly49* clusters showed even less conservation than to the B6 region. In the BALB/*c vs* 129S6 comparison, the best matches are again in the boundary regions containing α-*Ly49a-c* and *Ly49e-q* (data not shown). These direct sequence comparisons suggest that the BALB/*c* *Ly49* cluster is more closely related to the B6 haplotype, although the B6 region is more closely related to the 129S6 haplotype according to this type of analysis.[16]

Both B6 and BALB/*c* clusters possess a very high content and similar distribution of repetitive elements (Table 1). Approximately 50% of the *Ly49* gene cluster is composed of repetitive elements as compared to the murine genome total of 39%.[17] The most abundant elements were the LINE1 motifs at approximately 36% in both haplotypes. Despite the large contractions or expansions of the three sequenced murine *Ly49* haplotypes (B6, BALB/*c*, 129S6), the percentage of the different repetitive elements in each haplotype is similar. The number of repeats are significantly elevated relative to

Table 1 Summary of repeats in the BALB/*c* and B6 *Ly49* clusters[a]

Type of element	*Number of elements*		*Length occupied (bp)*		*Percentage of sequence (%)*	
	BALB[b]	*B6*[c]	*BALB*	*B6*	*BALB*	*B6*
SINEs						
Alu/B1	52	84	5848	9922	1.65	1.73
B2-B4	18	29	2248	3780	0.63	0.66
Ids	2	2	144	144	0.04	0.03
MIRs	0	0	0	0	0	0.00
Total	72	115	8240	13 846	2.32	2.42
LINEs						
LINE1	164	270	127 696	203 714	35.98	35.57
L3/CR1	1	1	92	92	0.03	0.02
Total	165	271	127 788	203 806	36.01	35.59
LTR elements						
MaLRs	20	26	10 885	12 771	3.07	2.23
ERVL	2	1	463	493	0.13	0.09
ERV_classI	12	18	11 602	14 117	3.27	2.46
ERV_classII	17	32	16 624	37 208	4.68	6.50
Total	51	77	39 574	64 589	11.15	11.28
Small RNA	3	8	204	497	0.06	0.09
Simple repeats	123	205	7818	13 367	2.20	2.33
Low complexity	77	106	3957	5086	1.01	0.89

[a]As detected by Repeatmasker.
[b]The region scanned for repeats in the BALB/*c* cluster was 354 kb and included all genes between the α gene fragment to intron 5 of *Ly49q*.
[c]The region analyzed for the B6 cluster was from the beginning of the α gene fragment to exon 6 of *Ly49q* (561 kb).

the total genome average, and this is especially true for the LINE1 motifs which are present at almost twice the level (~36%) of the mouse genome total of 19% (Table 1).[17] Thus, it appears that repetitive elements in general, and LINE1 motifs in particular, are intimately coupled with the evolution of the *Ly49* gene repertoire.

Novel BALB/*c* Ly49 exonic sequences

As a result of the BALB/*c* *Ly49* region sequencing, the complete predicted exon sequences for two new genes, *Ly49i*^{BALB} and *Ly49y*^{BALB}, were elucidated. Predicted amino-acid translations indicate that *Ly49y* would code for an activating receptor as it has an arginine codon in the correct location for DAP12 association in the putative transmembrane domain coding region (Table 2). DAP12 is the signaling adaptor that recruits Syk kinase upon ligand binding by the receptor.[18] The exonic coding sequence for *Ly49y* appears to be intact, but RT-PCR of cDNA derived from BALB/*c* adherent lymphokine-activated killer (ALAK) cells or total splenocytes with multiple *Ly49y*-specific primer sets did not result in the detection of *Ly49y* transcription (data not shown). These results suggest that *Ly49y* is an untranscribed pseudogene. In the previous BALB/*c* *Ly49* mapping study, exons 3 and 4 of *Ly49y* had been identified in BAC clones and named *Ly49x*^{BALB}.[14] When the complete coding sequence is analyzed, *Ly49y* is closest to *Ly49p*^{129} in nucleotide (~95%) identity, and the next closest cDNA match is *Ly49x*^{B6} (~93%). *Ly49y*^{BALB} was chosen as the name for this new gene due to its location relative to framework genes such as *Ly49g*. *Ly49p*^{129} is on the opposite side of *Ly49g*^{129} relative to *Ly49y*^{BALB}.

Table 2 Characteristics of BALB/*c* *Ly49* genes

Ly49	*Arginine codon in TM*[a]	*ITIM in IC*[b]	*Observed sequence anomalies*	*Transcript detected?*
α	(No exon 3)	Yes	No start codon, no exon 3, 5–7	NT[d]
a	No	Yes	None	Yes
b	No	Yes	None	Yes
c	No	Yes	None	Yes
e	No	Yes	None	Yes
g	No	Yes	None	Yes
i	No	Yes[c]	In-frame stop codon in exon 3	NT
l	Yes	No	None	Yes
q	No	Yes	None	Yes
y	Yes	No	None	No

[a]TM, transmembrane.
[b]IC, intracytoplasmic.
[c]DTYSTV.
[d]NT, not tested.

In contrast to *Ly49y*, the putative amino-acid translation of *Ly49i*^{BALB} does not have the DAP12-associating arginine residue, but instead has an ITIM-like motif in the cytoplasmic domain (DTYSTV), although the presence of a charged residue in the first position, normally occupied by a valine, may argue against the functionality of the motif. The closest match for *Ly49i*^{BALB} is *Ly49i*^{B6} with 96.5% nucleotide identity, and is in the correct location relative to *Ly49g* to be considered as the BALB/*c* allele of *Ly49i*. *Ly49i*^{BALB} appears to be a pseudogene as it

has an early in-frame stop codon in exon 3 and so suggests that the commonly used 5E6 mAb will only recognize Ly49C in BALB/*c* mice. In B6 mice, 5E6 binds to both Ly49C and Ly49I, thus making it difficult to separate the functions of these two receptors.[19] This can now be overcome simply by using BALB/*c* NK cells. As no other genes of the same subfamily (or even activating counterparts) exist in the BALB/*c* genome, it is highly unlikely that 5E6 will cross-react with any other BALB/*c* *Ly49* gene. Apart from the α fragment, which is common to all sequenced murine haplotypes, only *Ly49i* and *y* are predicted to represent pseudogenes in BALB/*c* mice. This is in stark contrast to the B6 and 129S6 *Ly49* haplotypes that have a significantly larger proportion of defective genes.[20,16]

Comparison of the total cDNA coding sequence of all BALB/*c*, 129S6, and B6 *Ly49* was performed by sequence alignment according to homology, and is displayed as a phylogram in Figure 3. Exonic coding sequence for all pseudogenes or genes for which no cDNA has been cloned was artificially spliced together. The BALB/*c* *Ly49a*, *b*, *c*, *e*, *g*, and *q* alleles grouped very closely with their B6 and 129S6 counterparts. $Ly49l^{BALB}$ paired with $Ly49m^{B6}$ 98% of the time in 1000 bootstrap sets,

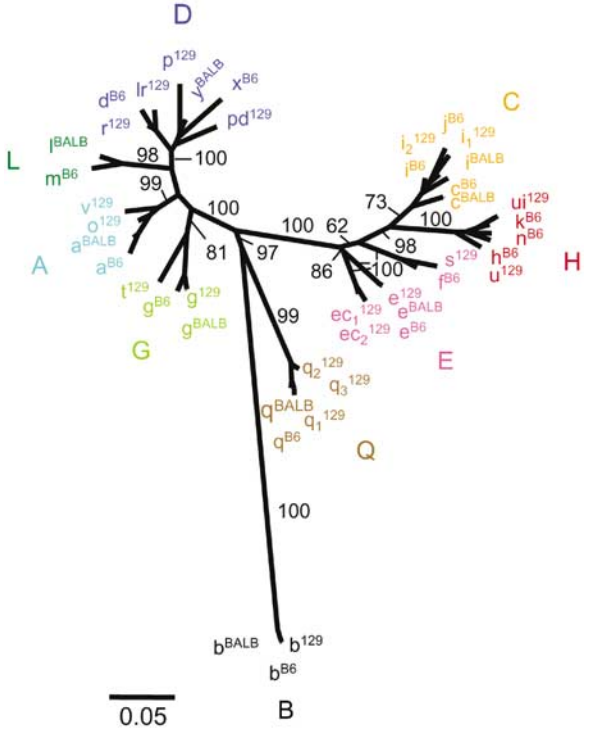

Figure 3 Phylogeny of all *Ly49* cDNAs from three murine inbred mouse strains. All known *Ly49* cDNA coding regions from B6, BALB/*c* and 129S6 mice, including artificially spliced cDNAs from nontranscribed pseudogenes, were aligned using ClustalX. Bootstrap analysis was performed on 1000 data sets with PHYLIP and the final consensus phylogram was visualized with TreeExplorer. The percentage bootstrap values are given for major branchpoints. The *Ly49* family of genes can be subdivided into the D, L, A, G, B, Q, E, H, and C-related groups. Each subfamily is represented by a distinct color and similar shades (light and dark) denote inhibitory *vs* activating receptors with similar carbohydrate recognition domains. The scale bar indicates the percentage of divergence between cDNAs.

supporting their proposed allelic relationship. $Ly49y^{BALB}$ always associated with the *Ly49d* subfamily, but the association with $Ly49x^{B6}$ was not stronger than to the next nearest relative $Ly49p^{129}$, again supporting a separate identity of *Ly49y* from *Ly49x*. $Ly49i^{BALB}$ was most similar in sequence to the other *Ly49i* alleles, but is still the most divergent member. The different subfamilies were color-coded according to the following groups: D, L, A, G, B, Q, E, H, and C-like. *Ly49s* and *f* were included in the E-like group even though they appear to be part of a separate branch in the phylogram. However, the bootstrap value for the average branching pattern for this section is low (62%) compared to the values for all other major branchpoints. In the remaining minority bootstrap alignments, *Ly49s/f* were on the same branch as the other E-like genes and so are represented as one subfamily.

Ly49 promoter elements in BALB/*c*

Three distinct sites of transcription initiation have been described in the *Ly49* genes. The promoter regions currently identified are named Pro1, Pro2, and Pro3.[21,22] The Pro1 element is active in immature NK cells, and a recent study has revealed that Pro1 is a bi-directional promoter complex that can act as a probabilistic switch.[23] The probabilistic nature of this element provides an explanation for the stochastic expression of Ly49 proteins. Pro2 is the most frequently studied *Ly49* promoter, and it is active in mature NK cells. Pro3 is an additional promoter located adjacent to the first coding exon (exon 2) of several *Ly49* genes, and it has been shown to represent the major site of transcriptional initiation of the *Ly49g* gene in splenic NK cells.[22] Figure 4 shows a comparison of the Pro1 and Pro2 regions of the BALB/*c* *Ly49* genes. The *Ly49q* gene has no discernable Pro1 element, as previously observed for the B6 and 129S6 *Ly49q* genes.[16] The comparison of the Pro1 elements shows that this region of the inhibitory *Ly49* genes is generally conserved, and the activating genes *Ly49l* and *Ly49y* have distinct Pro1 elements. Sequence analysis of the 129S6 *Ly49lr* gene revealed a 1660 bp deletion that removed the Pro1 element of this gene, and no transcripts of *Ly49lr* could be detected in 129S6 NK cells, supporting an important role for Pro1 activity in the initiation of *Ly49* transcription.[16] The *Ly49y* gene possesses a high degree of sequence homology to *Ly49lr* and it shares a similar position in the *Ly49* cluster. Remarkably, *Ly49y* contains a similar but nonidentical deletion of 1640 bp that removes the 5′ region of the Pro1 element including the TATA box associated with reverse transcription. *In vitro* studies have shown that reporter constructs lacking the reverse TATA box possess some forward promoter activity; however, like *Ly49lr*, *Ly49y* produces no detectable transcript in NK cells, suggesting that the deletion renders the Pro1 element nonfunctional. The Pro1 element of the BALB/*c* *Ly49i* gene is very similar to the 129S6 $Ly49i_1$ gene and the *Ly49j* gene of C57BL/6. All of these genes have a 19 bp deletion adjacent to the forward TATA element, and this alteration has been associated with decreased forward transcriptional activity of the *Ly49j* Pro1 element. Since *Ly49j* is transcribed in less than 5% of NK cells,[24] the BALB/*c* *Ly49i* gene may only be transcribed by a small percentage of cells. The possible reduced expression of the *Ly49i* gene in BALB/*c*

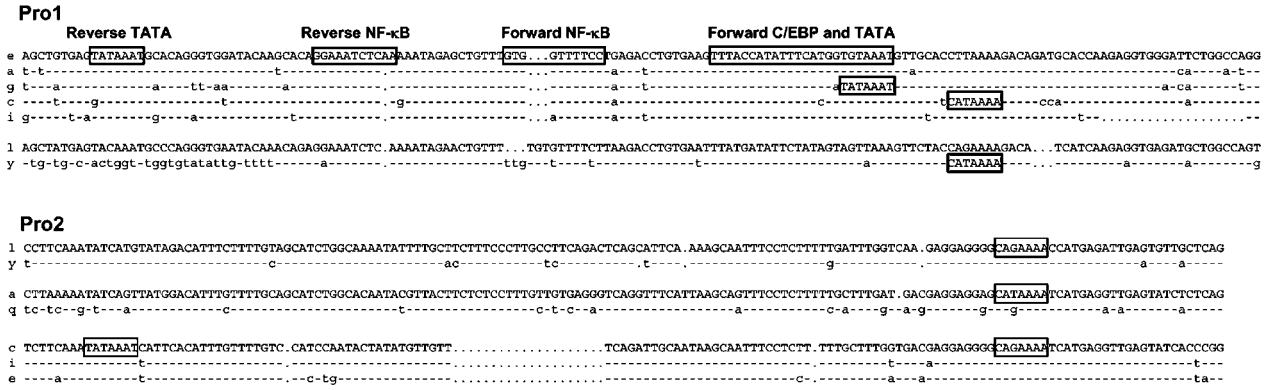

Figure 4 Promoter region analysis of BALB/*c Ly49* genes. The Pro1 and Pro2 regions of the BALB/*c Ly49* genes were aligned and separated into subgroups. The sequence of one member of each group is shown, and important putative transcription factor binding sites are boxed. A dash (−) indicates identity and a period (.) indicates absent nucleotides. Note that *Ly49q* has no identifiable Pro1 region, and *Ly49b* and the α fragments were not included in this analysis. The Pro2 and exon 1 regions of $Ly49g^{BALB}$ fall within an uncaptured sequence gap and so are not available.

probably has no functional consequence, since it represents a pseudogene due to the presence of a premature stop codon in exon 3.

The Pro2 elements of the BALB/*c Ly49* genes can be divided into three distinct groups: activating, *Ly49a*-related, and *Ly49c*-related genes. Comparison of the BALB/*c* Pro2 promoter sequences to the 129S6 and B6 alleles demonstrates a high degree of conservation of this promoter region between strains (data not shown).

Comparison of gene content between three mouse Ly49 haplotypes

The disparity with respect to gene content between the three sequenced murine *Ly49* haplotypes is remarkable. Figure 5 is a scale representation of the main *Ly49* genes (excluding the distant *Ly49b* gene) on chromosome 6 of these three haplotypes. The subfamily to which each gene belongs is indicated by the same color used in the phylogram of Figure 3. The most striking difference, aside from haplotype size, is the lack of any *Ly49h*-like genes in the BALB/*c* haplotype. All other *Ly49* subfamilies are represented in BALB/*c* mice, by at least one member. While the genomic content of each of the three haplotypes is dissimilar, it is obvious that the arrangement of several genes is conserved in each of the haplotypes. These conserved *Ly49* were previously described as 'framework' genes, and are composed of three gene pairs: *Ly49a* and *c*, *Ly49g* and *i*, and *Ly49e* and *q*.[14] The present sequence information from the BALB/*c* cluster supports the framework hypothesis. The regions containing framework gene pairs align very strongly between BALB/*c* and B6 haplotypes (see Figure 2), except for a BALB/*c*-specific deletion between the *Ly49g* and *i* genes.

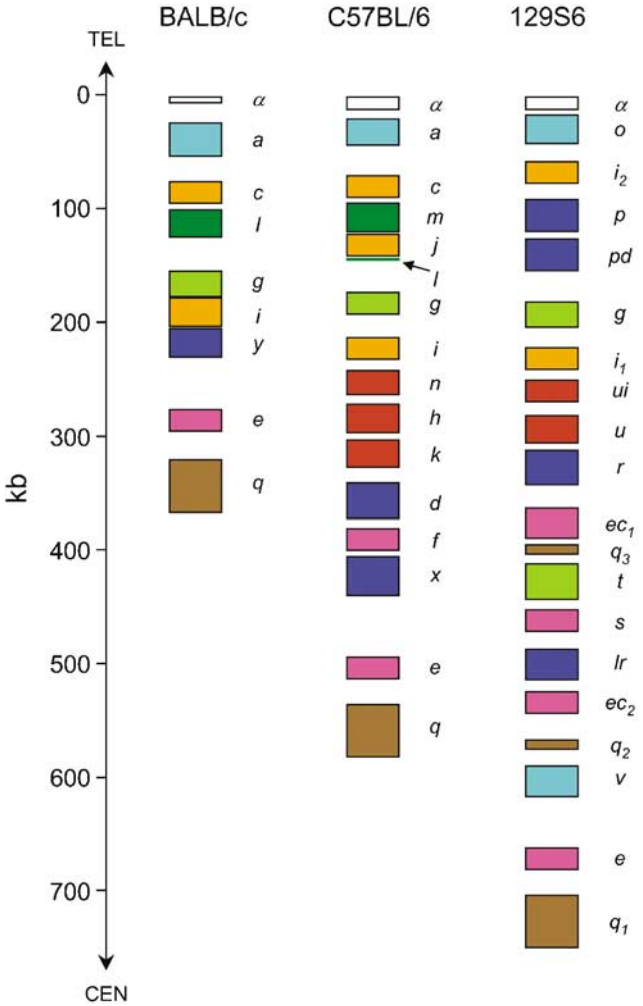

Figure 5 Comparison of gene content between three divergent murine *Ly49* haplotypes. The *Ly49* genes of BALB/*c*, B6, and 129S6 mice are compared graphically. The figure is drawn to scale and the scale in kilobases is shown on the left along with the relative locations of the chromosome 6 telomere (TEL) and centromere (CEN). The colors of the various *Ly49* genes follow the scheme of Figure 3. For each gene, the rectangle covers the first known exon to the last. The locations of the last exons of $Ly49q^{BALB}$ and $Ly49q1^{29}$ are inferred from the B6 haplotype.

Discussion

The sequence characterization conducted in this study provides a complete evaluation of the BALB/*c Ly49* repertoire. The known absence or presence of specific *Ly49* genes in this and other mouse strains will facilitate the matching of NK cell functions with specific receptors. For example, this study formally demonstrates that the lack of mAb 4E5 staining,[3] as well as the decreased level

of CHO tumor cell killing by BALB/*c* NK cells relative to B6 NK cells,[25] is due to the complete absence of a *Ly49d*-like gene in BALB/*c* mice. Similarly, the sequence data confirm the absence of a *Ly49h*-related gene in BALB/*c* mice as previously suggested by the lack of a NK cell subset staining with Ly49H-specific antibodies and the increased splenic viral titer of BALB/*c* mice relative to B6 during MCMV infection.[5,26,27] Furthermore, the complete sequencing of the BALB/*c* *Ly49* cluster indicates that the 5E6 mAb should only recognize Ly49C on NK cells in this inbred strain since the BALB/*c* allele for *Ly49i* is a pseudogene. Thus, the BALB/*c* mouse strain is superior to B6 in this respect, as all functional NK studies performed with B6 mice using the 5E6 antibody cannot differentiate between the contributions of Ly49C and I.

Where are the BALB/*c* pathogen receptors?

Ly49D recognizes MHC ligands from mice and other rodent species and confers to NK cells the ability to reject bone marrow cell allografts expressing H-2D^{d}.[25,28–30] However, the exact physiological role of the *Ly49d*-like genes is at present still unknown. In contrast, the identification of Ly49H as an activating receptor for the MCMV m157 gene product proves that Ly49H has a direct role in antiviral immunity mediated by B6 NK cells.[31,32] The lack of a gene coding for this type of receptor in the BALB/*c* haplotype is surprising. Assuming that MCMV is an endemic pathogen in wild mice as human cytomegalovirus is to humans, and that the absence of a *Ly49h*-like gene is not a result of inbreeding, then it appears that the NK cell-mediated arm of the innate immune system of the BALB/*c* mouse does not respond to MCMV infection.

It seems unlikely that such an NK cell receptor haplotype would have survived selection in the wild if MCMV were prevalent. If one assumes that each of the different *Ly49* haplotypes have properties advantageous to NK cell function, and therefore host defense, then it is possible that the BALB/*c* cluster has some, as yet, unknown protective component. The finding of a pathogen-derived ligand for the activating Ly49L receptor would support this hypothesis. BALB/*c* mice do possess at least one activating receptor for MCMV, namely TLR9. This receptor is expressed on dendritic cells and is able to recognize members of the herpesvirus family resulting in the rapid secretion of cytokines, especially interferon α/β by plasmacytoid dendritic cells.[33] Similar to *Ly49h*-deficient BALB/*c* mice, the absence of TLR9 results in mice that are much more susceptible to MCMV infection. Thus, it appears that activating receptors for MCMV on both dendritic cells and NK cells are required for an effective immune response to MCMV.

Ly49 allele inactivation

In this study, it is proposed that *Ly49l*^{BALB} and *Ly49m*^{B6} are actually alleles based on sequence homology and their relative location within the *Ly49* gene cluster. Therefore, it would appear that the single exon 7 fragment in the B6 haplotype currently termed *Ly49l* is misnamed and is in fact a remnant of a different gene. Exon 7-*Ly49l*^{B6} has several possible origins and include: (1) a small duplication of the region surrounding this exon from *Ly49m*, (2) duplication of the complete *Ly49m* gene followed by deletion of the promoter region and exons 1–6, or (3) *Ly49pd*[129] was originally a part of a B6 haplotype precursor and the exon 7 fragment is the residual evidence. Exon 7 of *Ly49pd* is almost identical in sequence to exon 7-*Ly49l*^{B6}.[16] The existence of complete *Ly49m* and *Ly49w* genes in the NOD inbred mouse suggests that the second scenario may be true. *Ly49w* is very similar to both *Ly49l*^{BALB} and *Ly49m*^{B6} and is obviously derived from a common progenitor.[8] If *Ly49w* were adjacent to *Ly49m* in the NOD haplotype, as *Ly49m* and exon 7-*Ly49l* are in the B6 genome, it would support such a hypothesis.

In addition to *Ly49l*^{B6} and *m*^{B6}, many genes in the three sequenced murine *Ly49* gene clusters have been inactivated. In particular, the selective inactivation of different members of the *Ly49c*-related subfamily during evolution is intriguing. In the B6 haplotype all three members, *Ly49c*, *j*, and *i*, are functional and expressed. In 129-strain mice only *Ly49i*$_1$ is present and functional; the putative *Ly49c*^{B6} allele, *Ly49i*$_2$, is a pseudogene, and there is no *Ly49j*^{B6} counterpart in 129S6 mice.[16] As shown in the present study, like the 129S6 haplotype, the BALB/*c* cluster only has two *Ly49c*-like genes and only one is functional. But unlike the 129S6 haplotype, the *Ly49i* allele has been inactivated in the BALB/*c* strain. The haplotype-specific inactivation of the various *c*-like inhibitory *Ly49* may be a consequence of inbreeding and 'trapping' of a *Ly49* gene with a MHC haplotype that does not supply a ligand. Thus, the inactivation of this *Ly49* by random mutation would be a neutral event with regard to natural selection. Another possible explanation for inhibitory *Ly49* gene inactivation is overlapping ligand specificities. Ligand redundancy could conceivably allow for the loss of a gene for which a related *Ly49* already provides the same specificity. This makes the expression of three C-like genes in the B6 strain all the more intriguing and argues for their having separate nonredundant functions despite evidence showing that Ly49C^{B6} can bind all the same types of MHC as Ly49I^{B6} and more.[11]

Evolution of the BALB/*c* Ly49 haplotype

The evolutionary pathways accounting for the B6, BALB/*c*, and 129S6 *Ly49* clusters are not straightforward due to the presence of 'shared' genes between the B6 and BALB/*c* haplotypes (*Ly49l* and *Ly49m* are considered to be allelic), and the B6 and 129S6 haplotypes (the *Ly49h*-like genes as well as *Ly49f/s* and *Ly49d/r*). At least three possible scenarios can account for the current haplotypes. In scenario 1, the 'Linear Progression Model' postulates that the BALB/*c* haplotype gave rise to an ancestral B6 haplotype by acquiring the *Ly49h*, *d*, and *f* precursors through duplications in the case of *Ly49d* and *f*, or through chimerization in the case of the *Ly49h* gene followed by tandem duplication (Figure 6a). The ancestral B6 cluster then underwent a small deletion of the *Ly49l/m* precursor resulting in an ancestral 129S6 haplotype. The individual ancestral B6 and 129S6 haplotypes then went on to acquire their present-day genetic structure.

In the second scenario, the 'Unequal Crossing Over Model' suggests that a minimal ancestral haplotype containing all the common genes of the three murine haplotypes gave rise to the BALB/*c* and ancestral

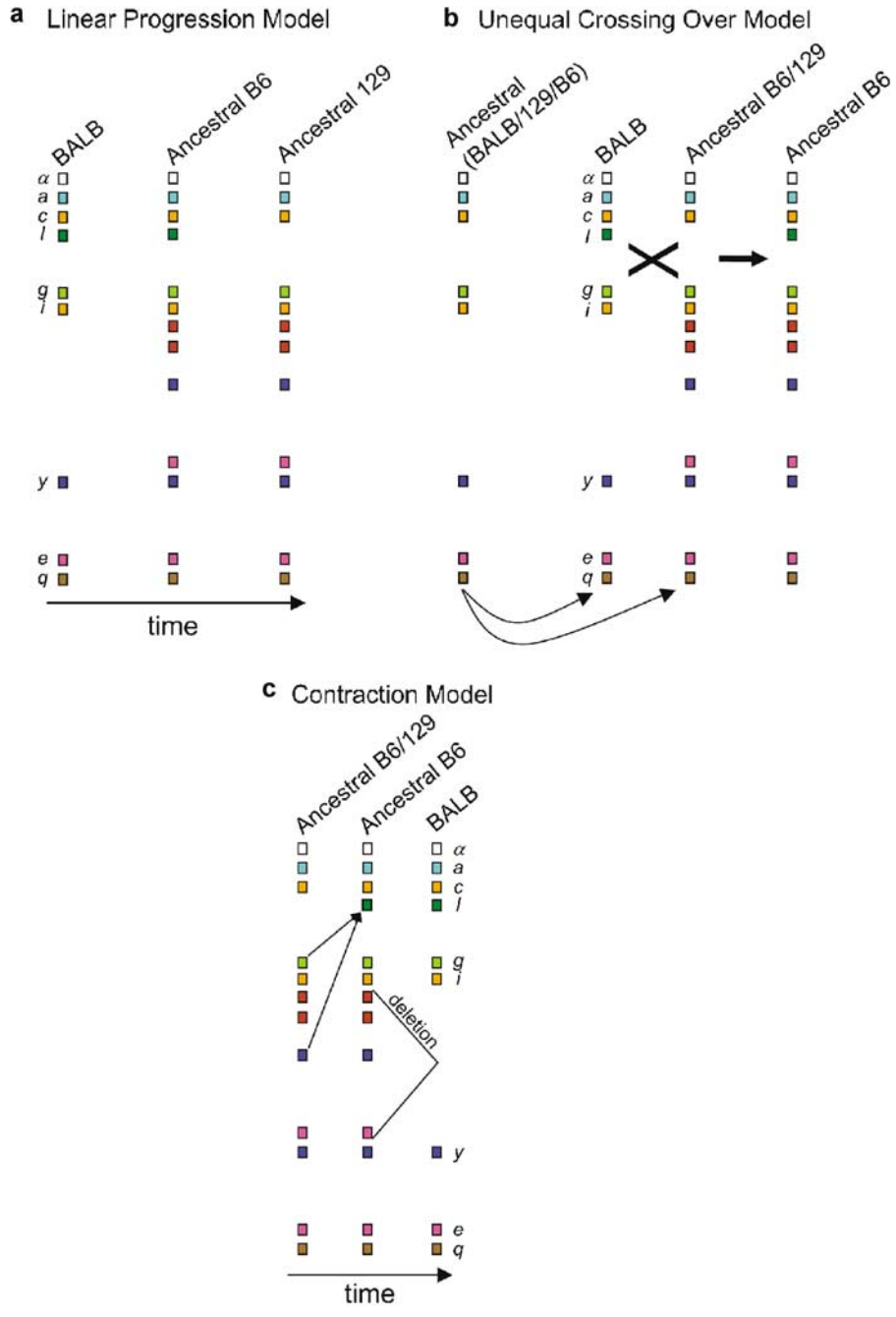

Figure 6 Three possible scenarios for the evolution of the murine *Ly49* gene cluster. Three hypothetical derivations of the present-day BALB/*c*, B6, and 129S6 *Ly49* haplotypes are presented. The type of each gene is indicated by the color, which was previously assigned to each subfamily in Figure 3. Genes and haplotypes are not drawn to scale. (**a**) The BALB/*c* haplotype preceded an ancestral B6 haplotype, which in turn gave rise to an ancestral 129S6 haplotype. All necessary duplications, deletions, and recombinations are assumed. (**b**) A minimal common ancestral haplotype gave rise to BALB/*c* and ancestral B6/129S6 haplotypes independently, which then recombined, giving rise to an ancestral B6 haplotype now containing the *Ly49l*/*m* gene. (**c**) Alternatively, the present-day BALB/*c* haplotype is the result of a large deletion in an ancestral B6 haplotype that removed the *Ly49h*, *d*, and *f* genes.

B6/129S6 haplotypes independently (Figure 6b). A mouse heterozygous for both haplotypes was generated and a meiotic crossing-over event within the *Ly49* cluster resulted in the unequal exchange of genetic information. This resulted in *Ly49l* being 'grafted' onto what became the ancestral B6 haplotype. Finally, in the third scenario, the 'Contraction Model' theorizes that in a haplotype ancestral to both B6 and 129S6 clusters a hybridization event gave rise to the *Ly49l*/*m* gene (Figure 6c). This ancestral B6 haplotype then experienced a large deletion between the *Ly49i* and *y* genes and resulted in the present-day BALB/*c* haplotype. Note that each model assumes that the present-day B6 and 129S6 haplotypes are derived from the ancestral haplotype

shown, and that missing gene duplication/hybridization events are assumed to occur later in the specific evolution of each haplotype. More complex scenarios can be envisioned for the formation of the current haplotypes, but these three models encompass the likely major events.

Rodent Ly49 evolution

There are now three fully sequenced murine *Ly49* haplotypes with a very wide disparity in gene content. In addition, a single rat *Ly49* haplotype has also been sequenced and described.[34–36] The BN/SsNHsd/MCW inbred rat possesses 34 uninterrupted *Ly49* genes with the same transcriptional orientation. In addition to a group of rat *Ly49* that have no obvious orthologues in the mouse, one gene, *rLy49i8*, is a clear orthologue of *mLy49b*, and *mLy49q* is clearly more related to a large group of rat *Ly49* than to the other mouse *Ly49*. Likewise, *rLy49i6* is more closely related to the mouse *Ly49* than to other rat *Ly49* genes. This suggests that the precursor species to mice and rats possessed a *Ly49* haplotype composed of *Ly49b*-, *Ly49q*-, *mLy49*-, and *rLy49*-like genes. After speciation, *mLy49* and *rLy49* expanded by duplication in mice and rats, respectively, while a single *mLy49* was maintained in rats and *rLy49* was lost in the mouse lineage. *Ly49b* was also maintained as a single gene in both species, but *Ly49q* preferentially duplicated many times in the rat. B6 and BALB/*c* mice have a single *Ly49q* gene, but 129S6 has two *Ly49q* duplications, both of which are missing two exons coding for most of the carbohydrate recognition domain and so casts doubt on their functionality.[16] Based on the gene content disparity of the three known murine haplotypes and the single rat haplotype, it would appear that *Ly49* cluster evolution is continuous, perhaps punctuated by major events, and that this evolution is quite rapid.

Concluding remarks

Mouse haplotypes can contain between 8–20 *Ly49* genes. This is highly reminiscent of the killer cell Ig-related receptor (*KIR*) genes in humans, whose protein products are the functional analogues of Ly49 in mice. There are two fully sequenced human *KIR* haplotypes, each containing nine or 10 *KIR* genes, but these haplotypes only have four framework *KIR* in common.[37] Segregation analysis of families suggests that *KIR* haplotypes can vary widely in gene content with a range of 4–14 genes per haplotype.[38] Why is there so much variation in NK cell receptor repertoires within species? At least two separate factors are driving the evolution of these gene clusters: (1) recognition of self by the inhibitory receptors and (2) recognition of non-self by the activating receptors. The evolution of the inhibitory receptors must keep pace with the evolution of the class I *MHC* genes. However, this is complicated by the fact that class I MHC has a second and equally important role as a ligand for the T-cell receptor. Therefore, all three gene families (*MHC*, *KIR*/*Ly49*, *TCR*) are co-evolving. Assuming that all the activating Ly49 are receptors for pathogens, then they must evolve as the pathogens evolve. The extremely high incidence of pseudogenes among the activating *Ly49* subfamily may reflect and/or assist the rapid evolution of this gene family.

Materials and methods

Identification of BAC clones for sequencing

The BALB/*c Ly49* gene cluster was sequenced using a sequence-ready minimal BAC tiling path that had been identified in a prior study from a BALB/*c*ByJ inbred mouse genomic BAC library (CHORI-28; BACPAC Resources, Oakland, CA, USA).[14] The following BAC clones were selected for complete sequencing: CH28-340f19 and CH28-347i8.

BAC clone sequencing

BAC plasmid DNA from each BAC clone was purified, mechanically sheared into 2–4 kb fragments, end-repaired, and cloned into pUC19 as previously described.[39] Subclones were sequenced on both strands with M13-forward and M13-reverse universal primers using Big Dye Terminators (Applied Biosystems, Foster City, CA, USA). Approximately seven-fold redundant sequencing was performed using ABI3730XL instruments (Applied Biosystems). Each BAC was individually assembled with CONSED version 13.0 (http://www.phrap.org/consed/consed.html),[40] and sequence contigs were ordered and oriented using subclone read-pair information. Residual sequencing gaps in the *Ly49* region were reduced following a single round of primer walking. The individual BAC sequences are available from GenBank using the following accession numbers: CH28-340f19 (AC159401) and CH28-347i8 (AC159402). Merging the BAC clone sequences into a single contiguous sequence of approximately 354 kb resulted in a regional sequence. The BALB/*c* regional sequence is available as Supplementary material at http://www.ircm.qc.ca/microsites/amakrigiannis/en/.

Sequence analysis

The B6 *Ly49* cluster sequence was assembled as described previously.[16] Dotplot comparison of the BALB/*c vs* B6 clusters was performed using Dotter (available from http://www.cgb.ki.se/cgb/groups/sonnhammer/Dotter.html).[41] Repetitive elements were identified with Repeatmasker version 3.0.0 (AFA Smit and P Green, unpublished; available at http://www.repeatmasker.org/). A PIP of the repeat-masked BALB/*c vs* B6 *Ly49* sequence was constructed using Advanced Pipmaker with a setting of single coverage (available at http://pipmaker.bx.psu.edu/pipmaker/).[15] To facilitate Dotter and Pipmaker analysis, some small gaps in noncoding, but highly conserved, areas were filled using B6 or 129S6 sequence. This sequence constituted less than 5% of the total BALB/*c* regional sequence and does not appear in the GenBank submission. The B6 *Ly49* cluster sequence and the BALB/*c vs* B6 PIP (Figure 1) are also available as Supplementary material at http://www.ircm.qc.ca/microsites/amakrigiannis/en/.

For *Ly49* cDNA phylogeny analyses, putative cDNA sequences were constructed from new exon data and aligned using ClustalX version 1.81.[42] Bootstrap analysis of 1000 replicates was then performed with Phylogeny Inference Package (PHYLIP) version 3.5c as described previously (available at http://evolution.genetics.washington.edu/phylip.html).[16,43] Phylograms were constructed with TreeExplorer version 2.12 (K Tamura, unpublished; available from http://evolgen.biol.metro-u.ac.jp/TE/TE_man.html).

Abbreviations
129S6, 129S6/SvEvTac; B6, C57BL/6; NK, natural killer; KIR, killer cell Ig-like receptor; PIP, percent identity plot; ALAK, adherent lymphokine-activated killer; RFLP, restriction fragment length polymorphism.

Acknowledgements

We gratefully acknowledge Vince Forgetta, Jessica Wasserscheid, Kevin McKee, and Xiaolan Zhang for sequencing and bioinformatics support. This project has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1CO12400. This work was also supported in part by an operating grant from the Canadian Institutes for Health Research. APM is a scholar of the Canadian Institutes for Health Research (New Investigator Award).

References

1 Chan PY, Takei F. Molecular cloning and characterization of a novel murine T-cell surface antigen, YE1/48. *J Immunol* 1989; **142**: 1727–1736.
2 Yokoyama WM, Jacobs LB, Kanagawa O, Shevach EM, Cohen DI. A murine T lymphocyte antigen belongs to a supergene family of type II integral membrane proteins. *J Immunol* 1989; **143**: 1379–1386.
3 Ortaldo JR, Mason AT, Winkler-Pickett R, Raziuddin A, Murphy WJ, Mason LH. Ly-49 receptor expression and functional analysis in multiple mouse strains. *J Leukoc Biol* 1999; **66**: 512–520.
4 Koo GC, Peppard JR. Establishment of monoclonal anti-Nk-1.1 antibody. *Hybridoma* 1984; **3**: 301–303.
5 Daniels KA, Devora G, Lai WC, O'Donnell CL, Bennett M, Welsh RM. Murine cytomegalovirus is regulated by a discrete subset of natural killer cells reactive with monoclonal antibody to Ly49H. *J Exp Med* 2001; **194**: 29–44.
6 Held W, Roland J, Raulet DH. Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. *Nature* 1995; **376**: 355–358.
7 Stoneman ER, Bennett M, An J *et al*. Cloning and characterization of 5E6(Ly-49C), a receptor molecule expressed on a subset of murine natural killer cells. *J Exp Med* 1995; **182**: 305–313.
8 Silver ET, Gong DE, Hazes B, Kane KP. Ly-49W, an activating receptor of nonobese diabetic mice with close homology to the inhibitory receptor Ly-49G, recognizes H-2Dk and H-2Dd. *J Immunol* 2001; **166**: 2333–2341.
9 Makrigiannis AP, Etzler J, Winkler-Pickett R, Mason A, Ortaldo JR, Anderson SK. Identification of the Ly49L protein: evidence for activating counterparts to inhibitory Ly49 proteins. *J Leukoc Biol* 2000; **68**: 765–771.
10 Mehta IK, Wang J, Roland J, Margulies DH, Yokoyama WM. Ly49A allelic variation and MHC class I specificity. *Immunogenetics* 2001; **53**: 572–583.
11 Hanke T, Takizawa H, McMahon CW *et al*. Direct assessment of MHC class I binding by seven Ly49 inhibitory NK cell receptors. *Immunity* 1999; **11**: 67–77.
12 Silver ET, Lavender KJ, Gong DE, Hazes B, Kane KP. Allelic variation in the ectodomain of the inhibitory Ly-49G2 receptor alters its specificity for allogeneic and xenogeneic ligands. *J Immunol* 2002; **169**: 4752–4760.
13 Makrigiannis AP, Rousselle E, Anderson SK. Independent control of Ly49*g* alleles: implications for NK cell repertoire selection and tumor cell killing. *J Immunol* 2004; **172**: 1414–1425.
14 Proteau M-F, Rousselle E, Makrigiannis AP. Mapping of the BALB/*c* Ly49 cluster defines a minimal natural killer cell receptor gene repertoire. *Genomics* 2004; **84**: 669–677.
15 Schwartz S, Zhang Z, Frazer KA *et al*. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res* 2000; **10**: 577–586.
16 Makrigiannis AP, Patel D, Goulet ML, Dewar K, Anderson SK. Direct sequence comparison of two divergent class I MHC natural killer cell receptor haplotypes. *Genes Immun* 2005; **6**: 71–83.
17 Waterston RH, Lindblad-Toh K, Birney E *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; **420**: 520–562.
18 McVicar DW, Taylor LS, Gosselin P *et al*. DAP12-mediated signal transduction in natural killer cells. A dominant role for the Syk protein-tyrosine kinase. *J Biol Chem* 1998; **273**: 32934–32942.
19 Brennan J, Lemieux S, Freeman JD, Mager DL, Takei F. Heterogeneity among Ly-49C natural killer (NK) cells: characterization of highly related receptors with differing functions and expression patterns. *J Exp Med* 1996; **184**: 2085–2090.
20 Wilhelm BT, Gagnier L, Mager DL. Sequence analysis of the Ly49 cluster in C57BL/6 mice: a rapidly evolving multigene family in the immune system. *Genomics* 2002; **80**: 646–661.
21 Saleh A, Makrigiannis AP, Hodge DL, Anderson SK. Identification of a novel Ly49 promoter that is active in bone marrow and fetal thymus. *J Immunol* 2002; **168**: 5163–5169.
22 Wilhelm BT, McQueen KL, Freeman JD, Takei F, Mager DL. Comparative analysis of the promoter regions and transcriptional start sites of mouse Ly49 genes. *Immunogenetics* 2001; **53**: 215–224.
23 Saleh A, Davies GE, Pascal V *et al*. Identification of probabilistic transcriptional switches in the Ly49 gene cluster: a eukaryotic mechanism for selective gene activation. *Immunity* 2004; **21**: 55–66.
24 Kubota A, Kubota S, Lohwasser S, Mager DL, Takei F. Diversity of NK cell receptor repertoire in adult and neonatal mice. *J Immunol* 1999; **163**: 212–216.
25 Nakamura MC, Naper C, Niemi EC *et al*. Natural killing of xenogeneic cells mediated by the mouse Ly-49D receptor. *J Immunol* 1999; **163**: 4694–4700.
26 Lee SH, Zafer A, de Repentigny Y *et al*. Transgenic expression of the activating natural killer receptor Ly49H confers resistance to cytomegalovirus in genetically susceptible mice. *J Exp Med* 2003; **197**: 515–526.
27 Scalzo AA, Fitzgerald NA, Simmons A, La Vista AB, Shellam GR. Cmv-1, a genetic locus that controls murine cytomegalovirus replication in the spleen. *J Exp Med* 1990; **171**: 1469–1483.
28 Furukawa H, Iizuka K, Poursine-Laurent J, Shastri N, Yokoyama WM. A ligand for the murine NK activation receptor Ly-49D: activation of tolerized NK cells from beta 2-microglobulin-deficient mice. *J Immunol* 2002; **169**: 126–136.
29 Nakamura MC, Linnemeyer PA, Niemi EC *et al*. Mouse Ly-49D recognizes H-2Dd and activates natural killer cell cytotoxicity. *J Exp Med* 1999; **189**: 493–500.
30 Raziuddin A, Longo DL, Mason L, Ortaldo JR, Bennett M, Murphy WJ. Differential effects of the rejection of bone marrow allografts by the depletion of activating *vs* inhibiting Ly-49 natural killer cell subsets. *J Immunol* 1998; **160**: 87–94.
31 Arase H, Mocarski ES, Campbell AE, Hill AB, Lanier LL. Direct recognition of cytomegalovirus by activating and inhibitory NK cell receptors. *Science* 2002; **296**: 1323–1326.
32 Smith HR, Heusel JW, Mehta IK *et al*. Recognition of a virus-encoded ligand by a natural killer cell activation receptor. *Proc Natl Acad Sci USA* 2002; **99**: 8826–8831.
33 Krug A, French AR, Barchet W *et al*. TLR9-dependent recognition of MCMV by IPC and DC generates coordinated cytokine responses that activate antiviral NK cell function. *Immunity* 2004; **21**: 107–119.

34 Wilhelm BT, Mager DL. Rapid expansion of the Ly49 gene cluster in rat. *Genomics* 2004; **84**: 218–221.
35 Hao L, Nei M. Genomic organization and evolutionary analysis of Ly49 genes encoding the rodent natural killer cell receptors: rapid evolution by repeated gene duplication. *Immunogenetics* 2004; **56**: 343–354.
36 Nylenna O, Naper C, Vaage JT *et al*. The genes and gene organization of the Ly49 region of the rat natural killer cell gene complex. *Eur J Immunol* 2005; **35**: 261–272.
37 Wilson MJ, Torkar M, Haude A *et al*. Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci USA* 2000; **97**: 4778–4783.
38 Carrington M, Norman P. *The KIR Gene Cluster*. National Library of Medicine (US), NCBI: Bethesda, MD, 2003.
39 Lander ES, Linton LM, Birren B *et al*. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**: 860–921.
40 Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**: 195–202.
41 Sonnhammer EL, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 1995; **167**: GC1–GC10.
42 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997; **25**: 4876–4882.
43 Felsenstein J. *PHYLIP (Phylogeny Inference Package) version 3.5c*. Department of Genetics, University of Washington, Seattle, 1993 (distributed by the author).