Genes and Immunity (2005) 6, 71–83

www.nature.com/gene

FULL PAPER

Direct sequence comparison of two divergent class I MHC natural killer cell receptor haplotypes

AP Makrigiannis[1,2,3], D Patel[1], M-L Goulet[1], K Dewar[4] and SK Anderson[5]

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

The murine Ly49 gene family encoding natural killer cell receptors for class I MHC is an example of a rapidly evolving cluster of immune response genes. Determining the genomic sequence of the 129S6/SvEvTac (129S6) Ly49 cluster and comparing it to the known sequence of the C57BL/6 (B6) region provided insight into the mechanisms of Ly49 gene evolution. 129S6 contains 20 Ly49, many of which are pseudogenes and 40% of the genes have no counterpart in the B6 genome. The difference in gene content between these two strains is primarily the result of distinct patterns of gene duplication. Phylogenetic analyses of individual exons showed that Ly49 genes form distinct sub-families and an ancestral haplotype can be surmised. Dotplot analysis supports limited allelism in the two haplotypes; however, large regions of variation punctuate these islands of co-linearity. These variable regions contain a high concentration of repetitive elements that are predicted to contribute to the dynamic evolution of this cluster. The extreme variation in Ly49 haplotype content between mouse strains provides a genetic explanation for the documented differences in natural killer cell phenotype, and also indicates that differences in natural killer cell function observed between B6 and 129-derived gene-targeted mice should be interpreted with caution.
Genes and Immunity (2005) **6,** 71–83. doi:10.1038/sj.gene.6364154
Published online 20 January 2005

Keywords: *Ly49; natural killer cell; haplotype; comparative genomics; strain-129*

Introduction

Complete sequencing of the human and mouse genomes have provided valuable data for the study of common and divergent aspects of the immune system in these two mammalian species. While there are many differences in the presence or expression patterns of many immune-related genes,[1] the general structure and function of the immune system is similar in the two species. Mammals, in general, appear to have a comparable make-up of immune genes. For example, sequencing of the cat MHC class II region (FLA) and a comparison to the human HLA and mouse H-2 class II regions showed remarkable similarities in the varieties of genes despite approximately 80 million years of evolution since the divergence of the most recent common ancestors.[2] However, there are some exceptions. For example, comparison of the human and mouse MHC class I gene maps shows that there is an ancestral framework of non-MHC genes in which class I-like gene subregions are located.[3] These class I-encoding subregions then independently expanded in the two species. Independent post-speciation divergence of class I MHC during mammalian evolution may explain the differential usage of killer cell Ig-related receptor (KIR) *vs* Ly49 class I MHC receptors in primate and rodent natural killer (NK) cells, respectively. The *Ly49* family has been selectively expanded in the rodent and horse lineages,[4,5] while only single genes have been maintained in primates, dogs, cats, pigs, and cattle.[6,7] In contrast, *KIR* have expanded in primates, dog, pigs, and cows.[7–9] Identification of conserved *KIR* or *Ly49* among species or within species haplotypes may provide valuable clues as to the importance of particular genes in immunity.

Human NK cells express the *KIR* family of genes. KIRs enable human NK cells to detect target cells that may have aberrant MHC expression due to infection or transformation. A single *KIR* haplotype can contain a variety of activating and inhibitory *KIR* genes, although

Correspondence: Dr AP Makrigiannis, Laboratory of Molecular Immunology, Institut de Recherches Cliniques de Montréal, Rm. 1340, 110 avenue des Pins Ouest, Montréal, QC, Canada H2W 1R7.
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 2 September 2004; revised 8 October 2004; accepted 11 October 2004; published online 20 January 2005

there appears to be framework *KIR* genes present in all haplotypes.[10] Despite this restriction, *KIR* haplotypes vary widely in the numbers and types of genes they contain as evidenced by human *KIR* region sequencing and mapping studies.[11] Instead of *KIR*, mouse NK cells express the *Ly49* family of genes. The type I Ig-related KIR monomers and type II lectin-like Ly49 homodimers are a remarkable example of convergent evolution of function and share many characteristics, including: (1) variegated NK (and some T) cell expression,[12] (2) MHC ligands,[13] (3) activating and inhibitory counterparts, and (4) identical adaptor and downstream signaling partners for both activating and inhibitory signaling functions.[14]

Both the human *KIR* and mouse *Ly49* genes are tightly clustered in a head to tail fashion with the mouse *Ly49* genes found on mouse chromosome 6. Recent gene-mapping studies comparing the C57BL/6 (B6), BALB/c and 129S6/SvEvTac (129S6) inbred mouse strains have shown that mouse *Ly49* haplotypes are highly plastic.[15–17] Strain-129S6 possesses the largest known *Ly49* haplotype with at least 19 family members, while the smallest is BALB/c with eight *Ly49* genes. The small number of BALB/c *Ly49* are, in large part, due to a missing 200 kb segment, which in B6 mice contains *Ly49d* and *h*.[17] The total lack of these genes in BALB/c mice provides a genetic explanation for their deficiency in NK cell-mediated CHO tumor cell killing and early resistance to murine cytomegalovirus, which are dependent on Ly49D, and H respectively.[18–20] Thus, a fifth characteristic shared by *KIR* and *Ly49* gene families is extreme haplotype-specific variation in gene number resulting in differential NK cell function.

Recent gene-mapping studies have shown that 'framework' genes may exist for *Ly49* haplotypes in mice as they do for human *KIR*. In order to compare the B6 and 129S6 *Ly49* haplotypes to a third strain, the BALB/c *Ly49* physical gene map was recently deduced.[17] Comparison of the three haplotypes showed that three gene pairs are conserved: *Ly49a* and *c*, *Ly49g* and *i*, and *Ly49 e* and *q*. However, the complete exonic sequences for the 129S6 and BALB/c *Ly49* are known from cDNAs for approximately half of the genes, while the remaining predicted genes have only been identified from exon/intron fragments. Mouse strain-129 is useful for many animal models, but its greatest application is the production of gene-targeted mice through 129-derived embryonic stem cells. Knowledge of significant differences in the genome of strain-129 relative to B6 mice is important in order to avoid background effects due to close linkage of target and non-target genes, which backcrossing cannot overcome. To fully elucidate the divergent 129S6 *Ly49* haplotype and gain insight into the processes involved in NK cell receptor gene evolution, this gene cluster was sequenced.

Results

Sequencing of the *Ly49* gene cluster from the 129S6 inbred mouse strain

In a previous report,[15] a physical map of the strain-129 *Ly49* region was deduced using *Ly49*-containing BAC clones from a 129S6-derived genomic library. This study provided an overlapping contig of sequential BAC clones for the sequencing of this gene region. End-sequencing assisted in the identification of the minimum number of overlapping BACs needed to cover the region. BACs 201k3, 264k14, 147o20, 52o15, and 17e15 were selected, shotgun cloned, and sequenced to approximately 7x coverage. The individual BAC sequences were assembled from these shotgun clone reads. Some gaps were present in each BAC assembly, but for the most part these were predicted to be small as subclones spanned them. Size analysis by pulsed-field gel electrophoresis of BACs minus the vector showed that the actual BAC sizes were very similar to the lengths of the assembled BAC sequences (data not shown). In one case (201k3), the PFGE-estimated size showed that there is approximately 16 kb missing, but the missing sequence is spread over multiple gaps. For all other BACs, the differences in the two size determinations were not significant. As the minimum length known for a *Ly49* gene with all exons is 18 kb, it is highly unlikely that unknown *Ly49* remain undiscovered in any of the sequence gaps.

The BAC clone assembled sequences were then merged to generate a sequence of the entire gene cluster. A representation of the gene content and order is shown in Figure 1. The total length of the deduced sequence for this region was 695 043 bp and contained 19 *Ly49* genes. In order of appearance the *Ly49* found in this region of the 129S6 genome are: α, *o*, i_2, *p*, *pd*, *g*, i_1, *ui*, *u*, *r*, ec_1, q_3, *t*, *s*, *lr*, ec_2, q_2, *v*, and *e*. No $Ly49q_1$ or *Ly49b* gene was present in the strain-129 regional sequence. Similar to the B6 *Ly49*, all strain-129 *Ly49* genes were in the same orientation and the average gene length was 24 kb. The open reading frame-detection software GENSCAN[21] detected only *Ly49* genes and repetitive element-associated motifs (reverse-transcriptase, gag, transposase, etc). A percent identity plot (PIP) performed against the B6 region showed that most regions have highly similar counterparts in the 129S6 and B6 *Ly49* haplotypes.

The sequence of the entire strain-129 *Ly49* region confirmed earlier mapping studies and resolved the order of $Ly49i_2/p$ and $Ly49q_2/v$, which were previously only mapped to the same BAC clones.[15] The BAC sequencing also resulted in the identification of a new strain-129 gene, $Ly49q_3$. $Ly49q_3$ is very similar in exonic sequence to $Ly49q^{B6}$ and $Ly49q_2^{129S6}$. Also like $Ly49q_2$, $Ly49q_3$ is missing exons 6 and 7. Upstream of *Ly49o*, a pseudogene was found that was a strain-129 orthologue of the previously described α gene fragment from the B6 *Ly49* cluster. The strain-129 and B6 α orthologues share a common location, very high sequence identity, and the presence of only exons 1, 2, 4 and 5.[16] The average gene density (including pseudogenes) over the entire 695 kb (α to *Ly49e*) of the strain-129 *Ly49* cluster is one gene per 37 kb. For comparison, the gene density for B6 (595 kb; α to *Ly49q*, but not counting *Ly49l*-exon 7 as a gene) and BALB/c (300 kb; α to *Ly49q*) are one gene per 40 and 34 kb, respectively.

The 129S6 *Ly49* region is highly populated with repetitive elements, especially LINE1 and to a lesser extent with LTR and SINE motifs (Table 1). Repetitive elements constitute almost 50% of the total sequence of both the B6 and 129S6 clusters with a similar make-up of motifs. This is greater than the total for the mouse genome, which is approximately 39%.[22] In both mouse *Ly49* regions, LINE1 make up 36% of the sequence, which is significantly higher than the 19% known for the total mouse genome.[22] Although found almost

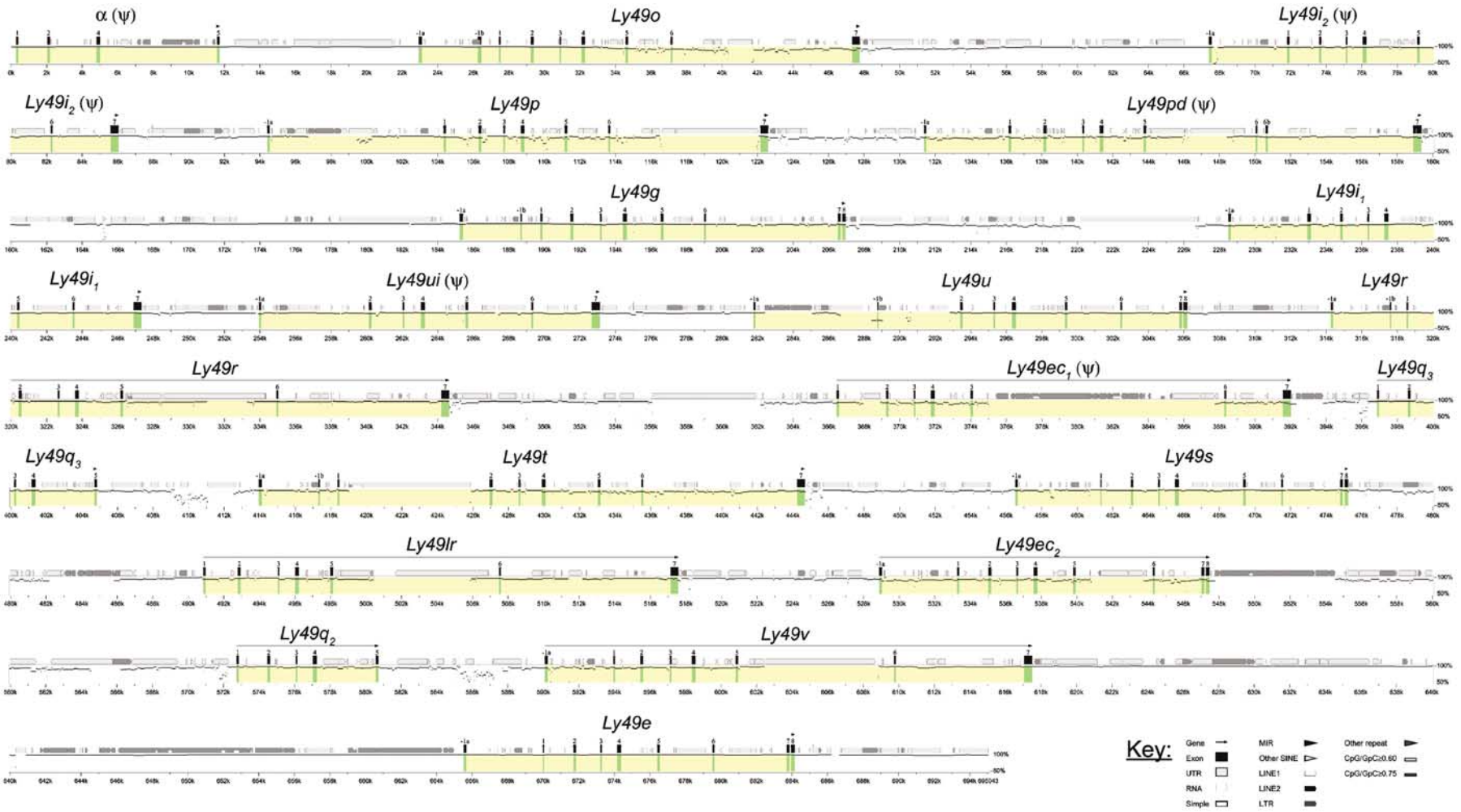

Figure 1 Genomic organization of the 129S6 *Ly49* cluster. BAC subclones were sequenced and compiled to generate the complete 129S6 *Ly49* cluster sequence. $Ly49q_1$ was not present in the BACs sequenced. Advanced Pipmaker was used to construct a scale diagram of the exons and a PIP *vs* the B6 *Ly49* sequence (single coverage). Genes and exons are marked in yellow and green, respectively. The length of the sequence in kilobases is shown underneath the PIP. Arrows show gene orientation, and the gene name is above the arrow. Obvious pseudogenes are denoted with a ψ. The functional status of $Ly49q_2$, q_3, *lr*, and ec_2 is unclear. The locations of various kinds of repetitive elements were revealed with RepeatMasker and are shown (see key). Accession numbers for individual BAC sequences are given in the Materials and methods.

Table 1 Summary of repeats in the 129S6 and B6 *Ly49* clusters[a]

Type of element	*Number of elements*		*Length occupied (bp)*		*Percentage of sequence (%)*	
	129S6[b]	*B6*[c]	*129S6*	*B6*	*129S6*	*B6*
SINEs						
Alu/B1	89	84	10 263	9922	1.48	1.73
B2-B4	36	29	4882	3780	0.70	0.66
IDs	4	2	284	144	0.04	0.03
MIRs	1	0	70	0	0.01	0.00
Total	130	115	15 499	13 846	2.23	2.42
LINEs						
LINE1	320	270	245 747	203 714	35.36	35.57
L3/CR1	3	1	161	92	0.02	0.02
Total	323	271	245 908	203 806	35.38	35.59
LTR elements						
MaLRs	30	26	15 646	12 771	2.25	2.23
ERVL	0	1	0	493	0.00	0.09
ERV_classI	20	18	14 018	14 117	2.02	2.46
ERV_classII	38	32	44 008	37 208	6.33	6.50
Total	88	77	73 672	64 589	10.60	11.28
Small RNA	7	8	427	497	0.06	0.09
Simple repeats	243	205	15 521	13 367	2.23	2.33
Low complexity	134	106	6587	5086	0.95	0.89

[a]As detected by Repeatmasker.
[b]The region scanned for repeats in the 129S6 cluster was from the beginning of the α gene fragment to the end of *Ly49e* (695 kb).
[c]The region analyzed for the B6 cluster was from the beginning of the α gene fragment to exon 6 of *Ly49q* (561 kb).

everywhere, LINE1 elements were especially concentrated between genes, between exons 5 and 6, 6 and 7, and sometimes between exon −1a and 1. *Ly49t* was unique in having a large LINE1 element between exons 1 and 2. The large intergenic region between *Ly49v* and *e* contained an unusually high number of LTR elements and is highly similar to the LTR-rich region between *Ly49x* and *e* of the B6 haplotype. This region has been hypothesized to act as a regulation barrier for *Ly49e* and *q*, which have different cell-type expression patterns compared to the other *Ly49*.[16] The average GC content of the 129S6 and B6 *Ly49* regions is 37.84 and 37.93%, respectively. When repeats are taken out, the GC content of both *Ly49* regions is ∼35%.

Transcriptional control elements of 129S6 *Ly49*

Studies of *Ly49* promoter elements have revealed three distinct sites of transcript initiation, Pro1, Pro2, and Pro3.[23,24] Pro1 is a bidirectional promoter complex that is active early in NK cell development, and the ability of this element to function as a probabilistic switch provides a potential explanation for the selective expression of Ly49 proteins in subsets of NK cells.[25] Pro2 is a major site of transcript initiation detected in mature NK cells, and Pro3 is an additional promoter detected 5′ of the first coding exon (exon 2). A comparison of the Pro1 and Pro2 regions of the 129S6 *Ly49* genes is shown in Figure 2. Sequence alignments reveal that there are two classes of Pro1 element, an inhibitory type and an activating type. The Pro1 element may not be functional in the activating *Ly49* genes, since there have been no Pro1 transcripts detected from any of the activators. Although the Pro1 elements of the inhibitory *Ly49* genes are generally conserved, the Pro2 promoters can be classified into three distinct inhibitory groups and one activator group: group 1 = *Ly49g, o, t,* and *v*; group 2 = Ly49q_1, q_2, and q_3; group 3 = *Ly49e*, ec_1, ec_2, i_1, i_2, and *s*; group 4 = *Ly49r, lr, p, pd,* and *u*.

The Pro1 element present in the B6 *Ly49j* gene was found to have only a weak forward promoter activity, providing a potential explanation for the expression of Ly49J on only a minor subset (5%) of NK cells as compared to the significantly larger subset (35%) expressing the highly related Ly49I protein.[25] The weak forward activity of the *Ly49j* Pro1 element was associated with a 19 bp deletion adjacent to the forward TATA element. Comparison of the B6 *Ly49j* Pro1 element with the 129S6 genes reveals that the $Ly49i_1$ Pro1 is nearly identical to the *Ly49j* element. This suggests that $Ly49I_1$ expressing NK cells would be expected to represent only a small subset in the 129S6 strain.

A Pro1 element and exon −1a was identified in the majority of the strain-129 *Ly49* genes with the exception of the *Ly49q* series and the *Ly49lr* gene (Figure 1). No regions with homology to Pro1 were detected in the *Ly49q* genes, and the Pro2 region of these genes is distinct from the other *Ly49* genes, which may account for the distinct expression pattern of the Ly49Q receptor.[26] The *Ly49lr* gene is highly related to the *Ly49d* and *Ly49r* genes, and the Pro2 region of these genes shares 99% nucleotide identity, suggesting that *Ly49lr* has a potentially active adult promoter; however, transcripts of

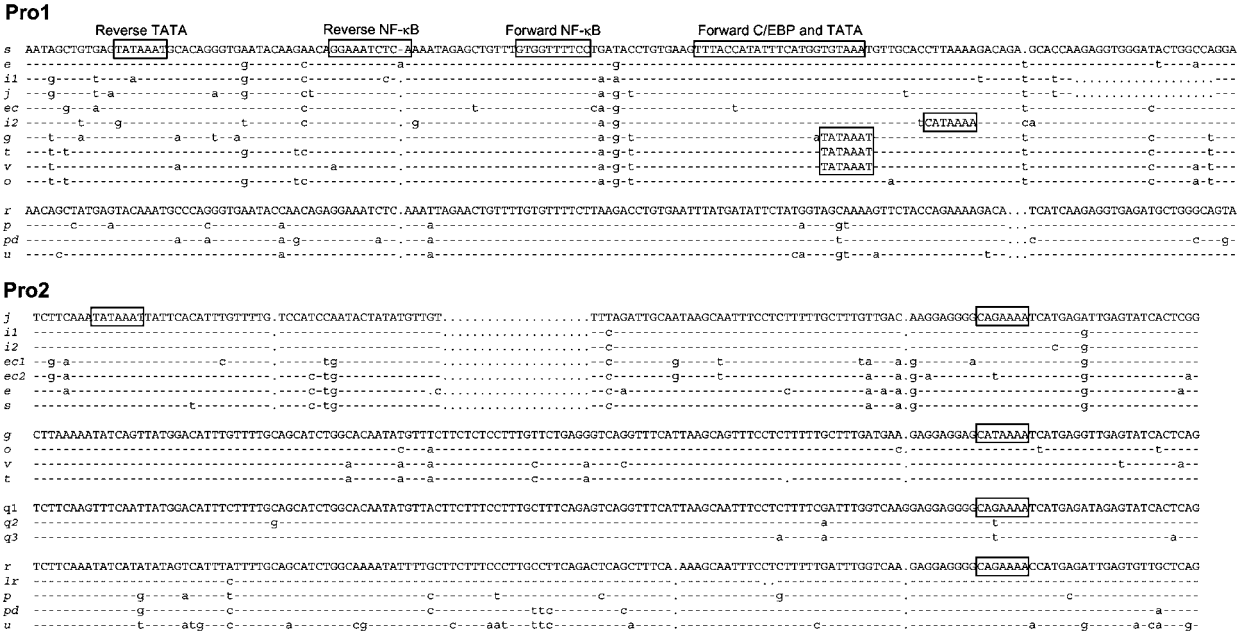

Figure 2 Comparison of the Pro1 and Pro2 promoter elements. The Pro1 and Pro2 regions of the Ly49 genes were aligned and separated into homologous groups (>90% homology within a group). The sequence of one member of each group is shown, and only nucleotide differences are shown for other members. The principle features of the Pro1 region are boxed and labeled,[25] and TATA-related elements are boxed in the Pro2 alignments.

Ly49lr could not be detected by RT-PCR (data not shown). The Pro1 region of *Ly49lr* is absent as a result of a 1660 bp deletion that deletes the majority of exon −1a and the upstream region containing Pro1. Deletion of the *Ly49a* Pro1 region was shown to abolish expression in a transgenic mouse model suggesting that the lack of detectable *Ly49lr* transcripts may be due to the deletion of the Pro1 element.[27]

Analysis of new 129S6 *Ly49* exonic sequences

Sequencing of the 129S6 *Ly49* region revealed the full exon sequences of 19 *Ly49* and allowed the functional status of these genes to be determined. Previously, based on exon 3–4 fragments, *Ly49lr, pd*, and *ui* were predicted to be activators as they contained an arginine codon in their transmembrane domain.[15] In this study, sequencing of exon 2 confirmed that these *Ly49* did not have intact ITIM sequences. All other 129S6 *Ly49* described here had ITIM (and no transmembrane arginine codon) (Table 2). However, $Ly49ec_1$, i_2, *pd*, *ui*, and α contain major defects in their exonic sequences including small deletions and insertions resulting in frameshifts, early in-frame stop codons, and missing exons indicating they are likely pseudogenes (Table 2). *Ly49lr* exon sequence was intact, but no transcript could be detected by RT-PCR (data not shown). $Ly49ec_2$ also had no obvious exonic anomalies, but RT-PCR resulted in only $Ly49ec_1$ transcripts being detected due to the very high level of sequence identity between these genes (data not shown). Therefore, the functional status of $Ly49ec_2$ is unknown and $Ly49ec_1$ appears to be a transcribed pseudogene. $Ly49q_2$ and q_3 transcripts were detected in splenocytes, but as two of three exons coding for the ligand-binding domain are missing the functional status of these genes is unknown. A new final exon (exon 8) was found for *Ly49g, u, s,* ec_2, and *e* (Figure 1). This exon is non-coding, and unlike all other *Ly49* introns, intron 7 is very small with a length of 80–100 bp and has also been reported for some B6-*Ly49*.[16]

In order to better understand the relationships among the new *Ly49*, a sequence comparison of all 129S6 *Ly49* cDNAs was performed. The cDNA-coding region of all new genes was also included and was assembled from putative exonic sequence based on known splicing boundaries in highly related genes. The cDNA for *Ly49b* from strain-129, which is not present in the sequenced cluster, has been deposited in GenBank (Accession AF395446). We had previously shown that an exon 3–intron 3–exon 4 gene fragment with very high sequence similarity to $Ly49q^{B6}$ existed in the 129S6 genome downstream of $Ly49e^{129S6}$ and had been assigned as $Ly49q_1^{129S6}$.[15] In this study, RT-PCR of strain-129 splenic RNA, using primers capable of amplifying $Ly49q^{B6}$, was performed and allowed the identification of the full coding region sequence of $Ly49q_1^{129S6}$. The strain-129 allele of *Ly49q* is highly similar to both the B6 and BALB/c alleles with approximately 98% identity at the protein level (Figure 3a). In addition to the spleen, $Ly49q_1^{129S6}$ mRNA could be detected in the liver, lung, kidney, and IL-2-activated NK cells (Figure 3b). Weak expression in the heart was consistently detected and may have been due to blood leukocytes that are known to express *Ly49q*.[26]

A phylogenetic analysis of all assembled and aligned, full length strain-129 and B6 *Ly49* cDNAs (including known and putative genes and *Ly49l* from BALB/c) was performed. A bootstrapped consensus tree was generated that displayed the evolutionary relationship among the different members of the *Ly49* family (Figure 4). The major subfamilies are displayed in their own color. These include the activating D-like (*Ly49d, r, lr, x, p, pd*),

Table 2 Characteristics of 129S6 *Ly49* genes

Ly49	*Arginine codon in TM*[a]	*ITIM in IC*[b]	*Observed sequence anomalies*	*Transcript*
b	No	Yes	None	Yes
e	No	Yes	None	Yes
g	No	Yes	None	Yes
i_1	No	Yes	None	Yes
o	No	Yes	None	Yes
p	Yes	No	None	Yes
q_1	No	Yes	None	Yes
r	Yes	No	None	Yes
s	No	Yes	None	Yes
t	No	Yes	None	Yes
u	Yes	No	None	Yes
v	No	Yes	None	Yes
ec_2	No	Yes	None	NT[c]
ec_1	No	Yes	No exon −1a; single bp deletion in exon 3; seven bp insertion in exon 7	Yes
i_2	No	Yes	In-frame stop in exon 4	NT
lr	Yes	No	No exon −1a	No
pd	Yes	No	In-frame stop in exon 2	NT
q_2	No	Yes	No exons 6 and 7; 'GTG' start codon	Yes
q_3	No	Yes	No exons 6 and 7	Yes
ui	Yes	No	No exon 1; in-frame stop in exons 2 and 4; single bp insertion in exon 6	NT
α	(No exon 3)	Yes	No exons 3, 6, and 7; no start codon	NT

[a]TM = transmembrane.
[b]IC = intracytoplasmic.
[c]NT = not tested.

activating H-like (*Ly49h, u, ui, k, n*), activating L-like (*Ly49l, m*), inhibitory A-like (*Ly49a, o, v*), inhibitory G-like ($Ly49g^{129S6/B6}$, *t*), inhibitory C-like (*Ly49c, i,* i_1, i_2, *j*), inhibitory E-like ($Ly49e^{129S6/B6}$, *f, s,* ec_1, ec_2), inhibitory Q-like ($Ly49q^{B6}$, q_1, q_2, q_3), and inhibitory B-like ($Ly49b^{129/B6}$) subfamilies. Although, it is possible to further subdivide the larger families (for example: *Ly49f* and *s vs* $Ly49e^{129S6/B6}$, ec_1, and ec_2), we determined this level of subdivision was adequate for the evolutionary analyses presented here. The carbohydrate recognition domains (CRD) of some pairs of activating and inhibitory sub-groups are highly similar, probably reflecting gene-conversion or recombination events and so are labeled in similar shades (for example: D and A, H and I, and G and L subfamilies). Phylogenetic analysis of individual exons shows that exon 2 (cytoplasmic domain) and exon 3 (transmembrane) clustering is similar, but different from the clustering for exons 4 (stalk), 5, 6, and 7 (CRD) (Figure 4). This suggests that recombination events took place between exons 3 and 4 for the formation of 'hybrid' *Ly49*. Bootstrap analysis strongly supports the groupings shown here. A bootstrapped consensus tree created with the CRD-coding exons was very similar to the individual trees shown for exons 5–7 (data not shown). The phylogenetic polarization of the exons coding for the CRD may have implications for the types of ligands with which these two greater Ly49 subfamilies (A/D/G/L *vs* H/C/E) interact.

Comparison and evolution of the 129S6 and B6 *Ly49* haplotypes

Phylogenetic analysis also showed that relative to the B6 *Ly49* haplotype, the 129S6 *Ly49* cluster expanded through a significantly different series of events. Relative to B6, the 129S6 haplotype development consists of two extra duplications of a D-like gene resulting in *Ly49p* and *pd*, two extra duplications of an E-like gene ($Ly49ec_1$ and ec_2), two extra duplications of a Q-like gene ($Ly49q_2$ and q_3), and single duplications of a G-like gene (*Ly49t*), and an A-like gene (*Ly49v*). In further contrast to the B6 haplotype, 129S6 has one less duplication of the H-like and I-like *Ly49*, and has no L-like members.

When the *Ly49* gene-sizes and gene-types of the B6 and 129S6 haplotypes were displayed graphically (Figure 5) we observed the following: (1) the numbers of genes and the size of *Ly49* haplotypes can vary widely, (2) the types of genes (activators *vs* inhibitors, and different subfamily representation) in these haplotypes likewise differ, and (3) there are regions of conserved gene content between the two haplotypes. These regions form a framework that can be used to show the evolutionary steps of major events that led to the present day B6 and 129S6 haplotypes from a most recent common ancestral haplotype (Figure 6). While the order of *Ly49* region modifications cannot be verified unless mice with these intermediary haplotypes are found, the model follows the most likely routes based on coding region nucleotide identity. The model is also supported by some gene duplications that have inherited obvious landmarks such as the missing exons 6 and 7 of $Ly49q_2$ and q_3, which indicate these as being derived from each other *vs* $Ly49q_1$. It is interesting to note that $Ly49q_2$ and q_3 appear to have been part of separate block duplications with $Ly49ec_2$, and ec_1, respectively. $Ly49ec_1$, and ec_2 are much closer in sequence identity to each other than to their most likely common ancestor, *Ly49e* (see Figure 4). It is not obvious which block duplication from *Ly49q* and *e* first occurred.

The graphical comparison in Figure 5 also indicates that *Ly49pd* and *Ly49l* are not allelic as previously hypothesized, and this conclusion is supported by total cDNA and individual exon analysis (Figure 4). However,

a

```
Q1-129    MSEQEVTYSTVRFHKSSGLQNQVRPEDNQGSREAGHKECSIPWHLIVIAFGILCVLLLVI   60
Q-B6      MSEQEVTYSTVRFHKSSGLQNQVRPEDNQGSREAGHKECSIPWHLIVIAFGILCVLLLVI
Q-BALB    MSEQEVTYSTVRFHKSSGLQNQVRPEDNQGSREAGHKECSIPWHLIVIAFGILCVLLLVI
          ************************************************************

Q1-129    VAVLVTNILQYKQEKHELQETLNCHHNCSTMQSDINAKEEMLRNMPLECSTGDDLLKSLN  120
Q-B6      VAVLVTNILQYKQEKHELQETLNCHHNCSTMQNDINAKEEMLRNMPLECSTGDDLLKSLN
Q-BALB    VAVLVTNILQYKQEKHELQETLNCHHNCSTMQSDINAKEEMPRNMPLECSTGNDLLKSLN
          ********************************.******** **********:*******

Q1-129    REQKRWYSETKSVLNSSKHPGGSLEIHWFCYGIKCYYFIMNKKGWHKCKQICEHYSLSLL  180
Q-B6      REQKRWYSETKSVLNSSKHPGGSLEIHWFCYGIKCYYFIMNKKGWRKCKQICEHYSLSLL
Q-BALB    REQKRWYSETKSILNSSKHPGGSLEIHWFCYGIKCYYFIMNKKGWRKCKQICEHYSLSLL
          ************:*********************************:**************

Q1-129    KIDAEDELKFLQLQVTPDSYWIGFSFDKKSEKWTWIENGTSKYALNMSTYNVKSGECVFL  240
Q-B6      KIDAEDELKFLQLQVTPDSYWIGFSFDKKSEKWTWIENGTSKYALNMSTYNVKSGECVFL
Q-BALB    KIDAEDELKFLQLQVTPDSYWIGFSFDKKSEKWTWIENGTSKYALNMSKYNVKSGECVFL
          ***********************************************.************

Q1-129    SKTRLENNKCEHVYPCICEKRLDKFPDSLPNNS  273
Q-B6      SKTRLENNKCEHVYPCICEKRLDKFPDSLPNNS
Q-BALB    SKTRLENNKCEHFYPCICEKRLDKFPDSLPNNS
          ************.*******************
```

b

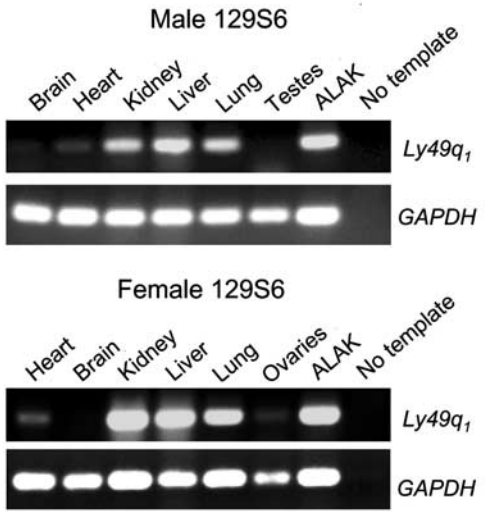

Figure 3 Cloning and characterization of $Ly49q_1^{129S6}$. The complete coding sequence for $Ly49q_1^{129}$ was amplified from 129S6 splenic cDNA, cloned, and sequenced. (**a**) ClustalX alignment of $Ly49Q_1^{129S6}$ amino-acid sequence with Ly49Q of B6 and BALB/c mice. An asterisk (*) indicates identity, while a semi-colon (:) and period (.) indicate strong and weak conservation, respectively. (**b**) Total RNA was isolated from the indicated tissues of male and female 129S6 mice, reverse transcribed, and amplified with primers specific for $Ly49q_1$ or *GAPDH*. A representative experiment is shown.

it is unclear if the exon 7 fragment of *Ly49l* in the B6 haplotype is a remnant of a complete *Ly49pd* or *Ly49l*-like gene. In contrast, similar comparisons show that $Ly49x^{B6}$ and $Ly49lr^{129S6}$ are most likely allelic. As a *Ly49x*-like gene fragment has been identified in BALB/c mice,[17] *Ly49x* may also be a 'framework' gene (ie present in all haplotypes), but this will depend on the elucidation of the complete $Ly49x^{BALB}$ genomic sequence.

In the PIP shown in Figure 1, when single coverage (nonlinear) analysis is used most regions of the 129S6 *Ly49* cluster have a very high match to the B6 haplotype showing that both haplotypes are clearly related. The relationship, however, is only partially linear as shown by Dotplot analysis comparing the B6 and 129S6 region sequences (Figure 7a). The flanking sequences are clearly related with a strong match evident from α to *Ly49c* in the B6 and α to $Ly49i_2$ in the 129S6 haplotype. The similarity ends with *Ly49p* and *pd*, which are clearly different from *Ly49m* and *j*. The B6-*Ly49g*, *i*, and *h* genes also match well to their 129S6 counterparts. After the *Ly49h* gene the sequence colinearity falls off dramatically and reappears at the other end with *Ly49e*. Limited sequencing of a BAC containing the $Ly49e^{129S6}$–$Ly49q_1^{129S6}$ intergenic region and the complete $Ly49q_1^{129S6}$ gene suggests that this area is highly conserved relative to the B6 cluster (data not shown). The B6 and 129S6 *Ly49* are highly homologous to each other in the coding and noncoding regions as shown by the plethora of *Ly49*-length lines throughout the Dotplot. Similar results are found when the 129S6 cluster sequence is compared to itself (Figure 7b). Only short stretches (single gene lengths) of colinearity are seen, even among obvious block duplications such as the *Ly49q* and *e* triplication. This may be due to constant assault from transposons on intergenic and intron regions, and implies that the *Ly49q*/*e* block duplications are relatively old.

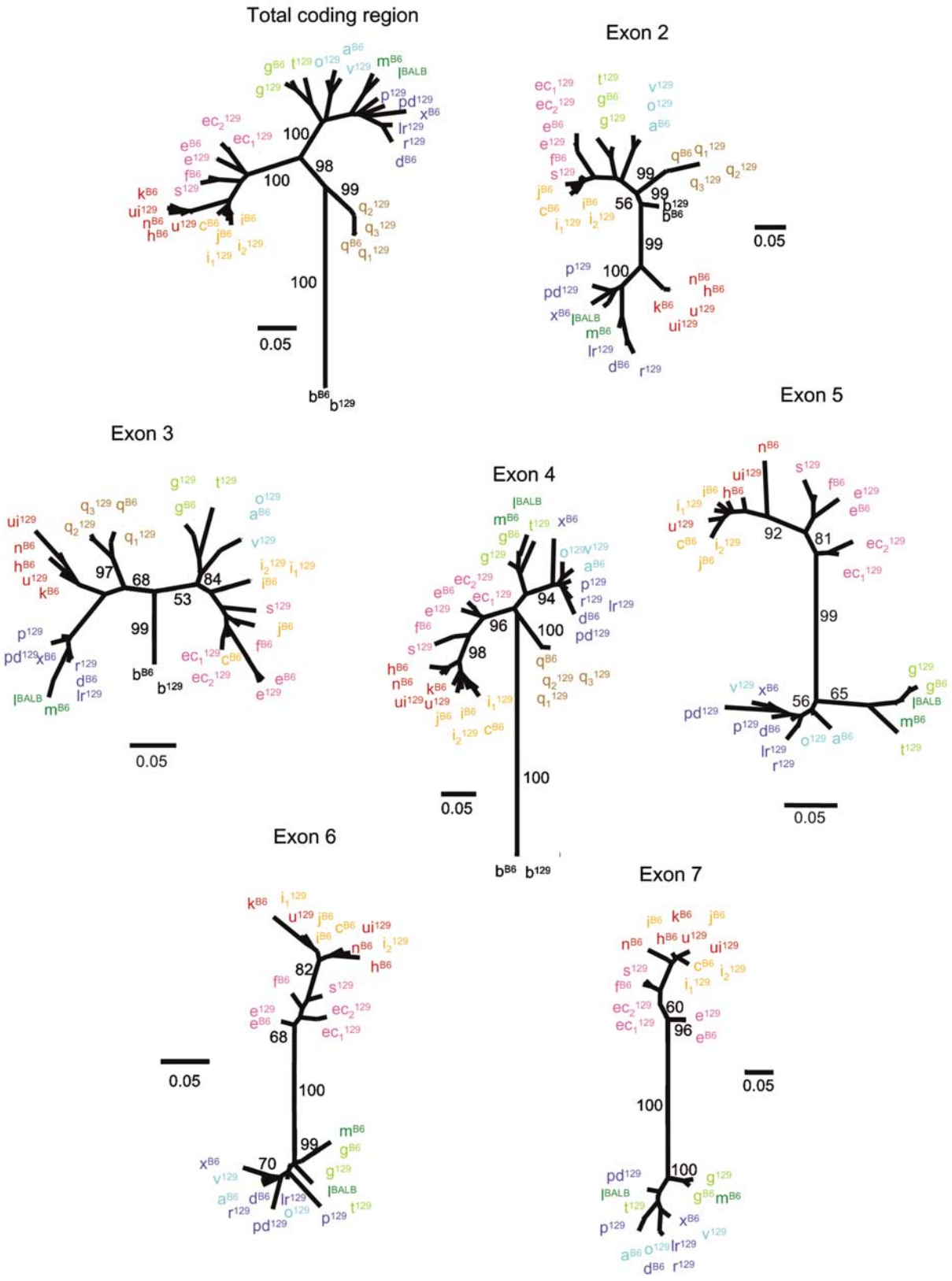

Figure 4 Phylogenetic analysis of all strain-129 and B6 *Ly49* cDNAs by exon. The total nucleotide coding region and individual exons from all strain-129 and B6 *Ly49* were aligned with ClustalX. Phylip was used to perform bootstrap analysis (1000 replicates) of the different cDNA alignments. Bootstrap values for major branchings are shown. A radial tree with the best supported branching was generated using TreeExplorer. *Ly49b* and *q*-related genes were omitted from exon 5–7 analyses to simplify tree shape. Different *Ly49* subfamilies grouped according to total cDNA clustering are labeled in different colors to emphasize the hybrid nature of different genes. The scale-bar underneath each tree represents the percent divergence of the various sequences.

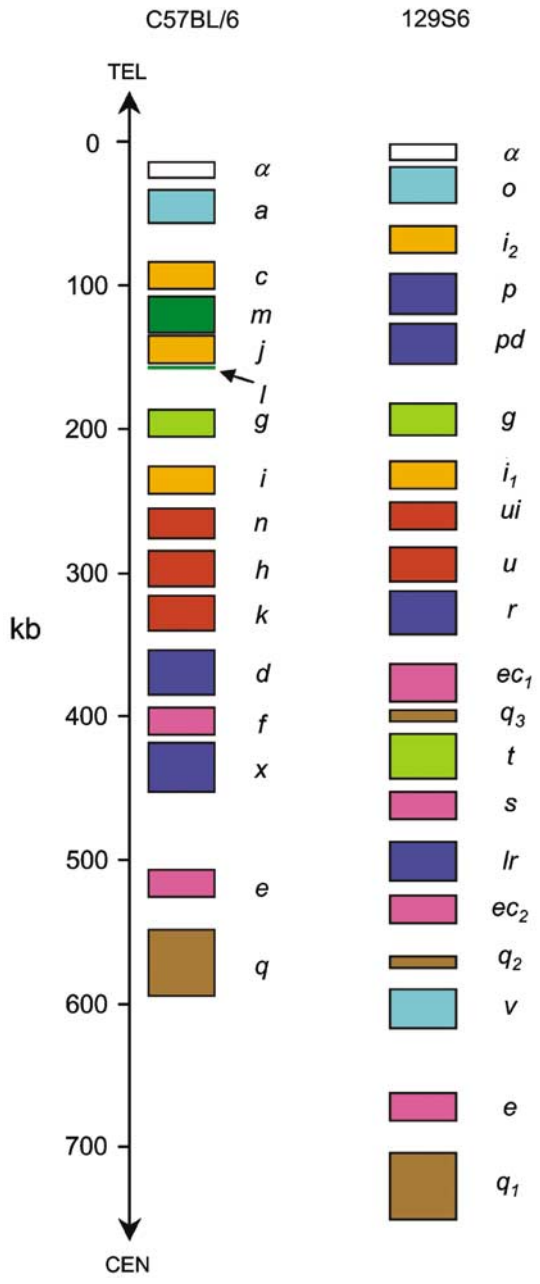

Figure 5 Scale diagram of the 129S6 and B6 *Ly49* regions. Rectangles represent the length of genes from the first (−1a, if known, or 1) to last exons (7 or, in some cases, 8; $Ly49q_2$ and q_3 end at exon 5). The B6 *Ly49* map was constructed as described in the Methods. *Ly49l* in the B6 map contains only exon 7. The colors for different genes were based on the subfamily phylogeny of Figure 4. The intergenic distance from $Ly49e^{129}$ to $Ly49q_1^{129}$ and size of $Ly49q_1^{129}$ was based on the B6 haplotype. The scale bar is in kilobases (kb) and shows the relative locations of the telomere (TEL) and centromere (CEN) of mouse chromosome 6.

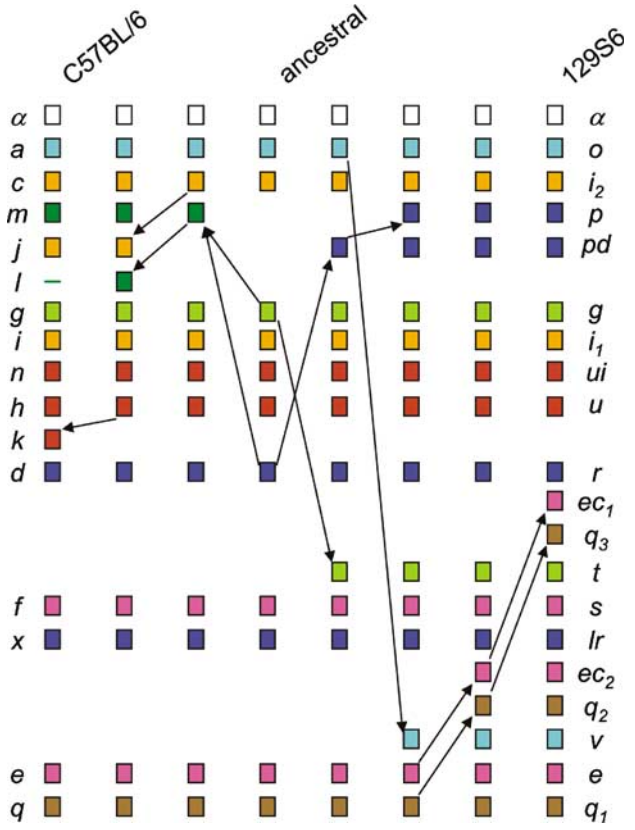

Figure 6 Hypothetical derivation of 129S6 and B6 *Ly49* haplotypes from an ancestral haplotype. A possible history of duplications that led to the 129S6 and B6 *Ly49* haplotypes from an ancestral haplotype containing all common genes of the two haplotypes is shown. The B6 and 129S6 haplotypes and associated gene names are shown on the left and right, respectively. Gene sizes and intergenic distances are not drawn to scale. The colors for different genes were based on the subfamily phylogeny of Figure 4. Arrows indicate likely duplications over time resulting in intermediate haplotypes. The creation of *Ly49m* is shown as a recombination event between *Ly49g* and *d*.

Discussion

To gain insight into the complexity of *Ly49* genomics and evolution, the *Ly49* cluster was sequenced from a previously assembled BAC contig map of this region of the inbred 129S6 mouse strain.[15] Strain-129 was originally chosen for the study of NK cell receptor diversity due to extreme differences compared to B6 mice in NK cell surface receptor expression and functional phenotypes. Strain-129 is especially important for its general use in the construction of gene knockout mice. Based on *Ly49* RFLP studies, cDNA cloning, gene mapping, and FACs analysis with haplotype-differentiating anti-Ly49 antibodies, the 'strain-129' *Ly49* haplotype is present among the following inbred mouse strains: 129P3, 129X1, 129S6, C57L/J, C57BR/cdJ, SJL/J, and FVB/N.[28,29,15] The 129S6 *Ly49* gene cluster spans 695 kb and contains 19 genes not including $Ly49q_1$ and *b*. This cluster includes five obvious and up to nine potential pseudogenes. The large number of pseudogenes containing all exons and intact promoter regions is intriguing. One possible explanation for their retention could be that they are involved in generating new genes by gene conversion, as this phenomenon has been observed for immunoglobulin variable gene segments and proposed for the human class I MHC region where 96/224 genes are not predicted to be expressed or translated.[30,31]

The *Ly49* haplotypes of two different inbred mouse strains, B6 and 129S6, have now been sequenced. Allelic regions can be identified between B6 and 129S6

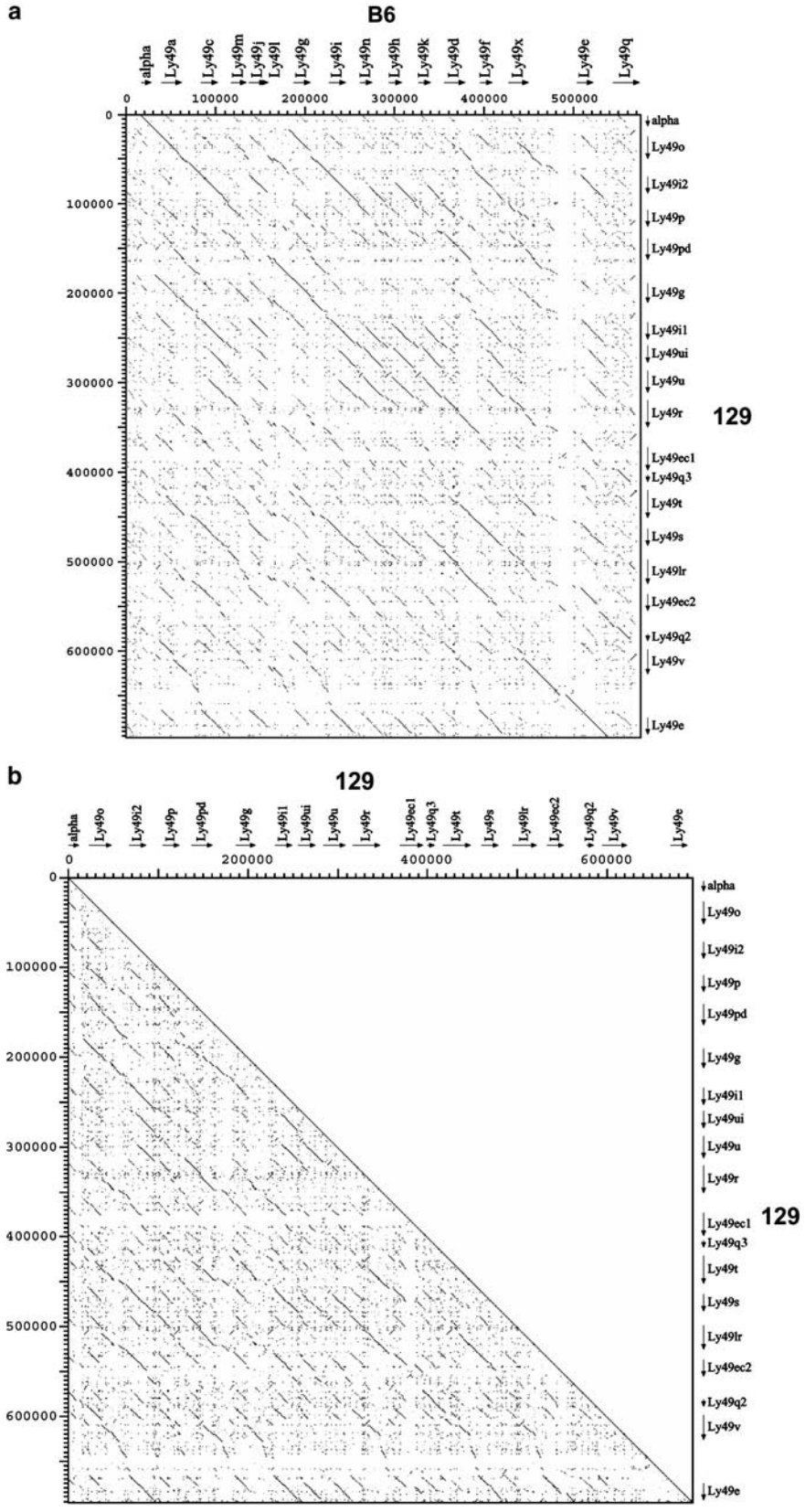

Figure 7 Dotplot analysis of the inbred mouse strain-129S6 *Ly49* region. The similarity of the 129S6 *Ly49* sequence (695 kb) was ascertained against (**a**) the B6 *Ly49* sequence (horizontal, 561 kb) and (**b**) itself using Dotter. Arrows indicate the location and orientation of genes. The scale in base pairs (bp) is indicated horizontally and vertically. Diagonal lines indicate regions of high sequence identity, which may be the result of gene duplication within a haplotype (129 *vs* 129) or conservation between haplotypes (129 *vs* B6).

haplotypes, but are surrounded by noncontiguous sequence (Figure 7a). The extreme polymorphism in the intergenic region between *Ly49* is again reminiscent of the class I MHC in humans where the most polymorphic gene loci had the highest variation in surrounding noncoding sequence.[30] In addition, the 129S6 haplotype has more 'complete' genes (all exons) and more functional transcribed genes than the B6 cluster. Both 129S6 and B6 haplotypes appear to be much larger in terms of numbers of genes and total size than the BALB/c *Ly49* haplotype, which contains eight potential genes spread over 300 kb.[17] The reason for the large number of inhibitory *Ly49* has been hypothesized to allow a greater diversity in individual NK cell subsets, which in turn would be more likely to contain populations that can react to specific downregulation of MHC isoforms. If this is true, one would predict that NK cell surveillance for specific MHC-downregulation resulting from infection or transformation would be most efficient in 129S6 (8 inhibitory, NK cell-expressed Ly49), followed by B6 (7), and then in BALB/c (4). However, for the comparison to be meaningful all these *Ly49* haplotypes would have to be on the same genetic background to eliminate non-*Ly49* immune contributions.

Including $Ly49q_1^{129S6}$, there are 11 confirmed functional genes within the main 129S6 cluster, and 12 when *Ly49b* is included. $Ly49q_1^{129S6}$ is very similar to the B6 allele and is easily detectable in hematopoietic (spleen and liver) and nonhematopoietic tissues (lung and kidney) by RT-PCR (Figure 3a, b). Recent studies have shown that $Ly49q^{B6}$ is expressed on the cell surface of granulocytes and macrophages and it can be induced with IFN-γ.[26] We were able to amplify $Ly49q_1^{129S6}$ from mRNA isolated from adherent IL-2 activated NK cell cultures, although whether $Ly49q_1^{129}$ is present on these NK cells or contaminating granulocytes in the culture is unclear. Relative to B6 and BALB/c, there are two duplications of the *Ly49q* gene in the 129S6 genome, $Ly49q_2$ and q_3. The functionality of $Ly49q_2$ and q_3 is in doubt as they are missing exons 6 and 7, which are essential as they code for most of the ligand-binding domain.

The *Ly49* cluster is densely populated by repetitive elements, especially the LINE1 class of non-LTR retrotransposons (Figure 1 and Table 1). Mobile elements facilitate genome evolution in a number of ways that can be destructive or constructive.[32] For example, LINE1 insertions may help to promote unequal homologous recombinations that lead to tandem gene duplications. This is seen for the H-like subfamily in B6 where *Ly49n, h,* and *k* are clustered together on the chromosome. Similarly, in the 129S6 cluster *Ly49p* and *pd* share high sequence homology and are found together and are most likely the result of tandem duplication. Introduction of LINE1 elements into introns may also promote early transcript termination due to strong poly(A) signals in LINE1. In addition, LINE1 promoters can affect neighboring gene transcription rates. Thus, it is likely that the plethora of LINE1 elements in the *Ly49* region has facilitated rapid evolutionary change for this gene family. Finally, a recent report has found an increased concentration of LINE1 motifs surrounding mono-allelically expressed genes.[33] In agreement with this report, several Ly49 have been shown to be expressed in single NK cells from either or both alleles in a probabilistic fashion.[34,35]

In addition to complete gene duplication, it is possible that unequal homologous recombination at repetitive elements such as LINE1 also facilitates recombination between exon groups (such as exons 2, 3 and 4–7) and has allowed the evolution of 'hybrid' *Ly49*. For example, *Ly49u* appears to be a product of the 5′ half of *Ly49r* and the 3′ half of $Ly49i_1$, with the B6 alleles being *Ly49h, d,* and *i*, respectively. The hypothetical haplotype derivation shown in Figure 6 proposes that the *Ly49r*-like and *Ly49i*-like chimerization took place before the divergence of the B6 and 129S6 haplotypes. However, the recombination that led to *Ly49m* in B6 seems to have taken place after this evolutionary split. The 5′ and 3′ regions of *Ly49m* are closely related to the same regions of *Ly49d* and *g*, respectively, and so *Ly49m* may also be the result of recombination between activating and inhibitory *Ly49* genes. The possible evolution of the B6 and 129S6 *Ly49* clusters shown in Figure 6 agrees with that of Wilhelm *et al*[16] in the kinds of duplications and recombinations that took place. However, in the present model the *Ly49t* and *q/e* duplications are specific to the 129S6 derivation and were never a part of the evolutionary history of the B6 haplotype. In the former model, the $Ly49ec_1$ and q_3 second duplication would have occurred before the *Ly49lr* and *v* recombination. It was previously hypothesized that a recombination between *Ly49lr* and *v* gave rise to the *Ly49x* gene in B6 mice, thus deleting $Ly49ec_2$ and q_2;[15] however, new exonic data in the present report suggest that *Ly49x* and *lr* are possible alleles. The model in Figure 6 also differs from that of Wilhelm *et al* in that it does not take into account possible gene conversion events.

While the 129S6 *Ly49* haplotype contains the most *Ly49* known for any one haplotype, it may not be the largest. The NOD mouse expresses a unique assortment of *Ly49*, including the most known functional activating *Ly49*.[36] The number of functional activating *Ly49* appears to be proportional to the total number of genes. The ratio of functional activating *Ly49* to total *Ly49* genes for the known haplotypes is: BALB/c, 1/8; B6, 2/14; 129S6, 3/19. With four functional activating *Ly49*, the NOD *Ly49* haplotype may be larger than 129S6. The recent description of the genomic organization of the first rat *Ly49* haplotype (strain: BN/SsNHsd/MCW) with 36 complete or almost complete genes over 1.8 Mb supports the possibility of larger murine *Ly49* haplotypes.[37] The complete sequence characterization of the *Ly49* cluster from 129S6 and B6 mice has yielded considerable information on the potential repertoire of class I MHC receptors, their evolution, and elements controlling their expression. Finally, the extreme genetic polymorphism found in this immunologically important region demands caution in the interpretation of gene-knockout studies involving NK cell function when the mutant mouse is derived from a 129-ES cell and the identity of the *Ly49* cluster has not been shown, even after many backcross generations. Future investigation of other distinct murine *Ly49* genotypes will help to further define the evolution and function of this interesting gene cluster.

Materials and methods

Animals and cell culture

129S6 mice were bred and maintained at the IRCM animal care facility in accordance with institutional

guidelines. Plastic-adherent IL-2 activated NK cells were prepared from splenocytes as previously described.[34]

cDNA cloning and tissue RT-PCR

Total RNA was extracted from mouse plastic adherent lymphokine-activated killer cells and various mouse organs using TRIzol reagent (Invitrogen, Montreal, QC, Canada). Complementary DNA was produced using Superscript First Strand cDNA synthesis kit (Invitrogen). For full-length cloning and expression studies, $Ly49q_1^{129S6}$ (GenBank Accession: AY642595) was amplified from cDNA as previously described.[17] *GAPDH* cDNA was amplified with forward primer 5′-ACTCACGGCAAATT CAACGGC and reverse primer 5′-ATCACAAACATG GGGGCATCG. PCR cycling parameters used were: 94°C 30 s, 57°C 30 s, 72°C 1 min, 30 cycles.

Identification of BAC clones for sequencing

The 129S6 *Ly49* gene cluster was sequenced using a sequence-ready minimal BAC tiling path that had been identified in a prior study from a 129S6/SvEvTac in-bred mouse genomic BAC library (RPCI-22; BACPAC Resources, Oakland, CA, USA).[15] The following BAC clones were selected for complete sequencing: 201k3, 264k14, 147o20, 52o15, and 17e15.

BAC clone sequencing

BAC plasmid DNA from each BAC clone was purified, physically sheared into 2–4 kb fragments, end-repaired, and cloned into pUC19 as previously described.[38] Subclones were bidirectionally sequenced using M13-forward and M13-reverse universal primers using Big Dye Terminators (Applied Biosystems, Foster City, CA, USA). Approximately 7-fold redundant sequencing was performed using AB13730XL instruments (Applied Biosystems). Each BAC was individually assembled using CONSED version 13.0 (http://www.phrap.org/consed/consed.html),[39] and sequence contigs were ordered and oriented using subclone read-pair information. The BAC clone assembled sequences are available under the following accessions: 201k3 (AY686474), 264k14 (AY686475), 147o20 (AY686476), 52o15 (AY686477), and 17e15 (AY686478). Merging the BAC clone sequences into a single contiguous sequence of 695 kb resulted in a regional sequence. The 129S6 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 using the following GenBank Accession files: AC087336 (α exon 1—*Ly49l* exon 7; 165 870–1 (reverse)), AC090127 (*Ly49g* exon 6—*Ly49f* exon 6; 1–212 404), AC090563 (*Ly49d* exon 7—*Ly49q* exon 6; 160 734–35 273 (reverse)). The gap between *Ly49l* and *g* was filled with sequence from Ensemble (mouse chromosome 6, May 2004 release, 130 973 035–130 907 915 (reverse)), which is not in GenBank. The location of exon 7 of $Ly49q^{B6}$ was also found in Ensemble. Dotplot comparison of 129S6 *vs* B6 and itself was performed using Dotter (available from http://www.cgb.ki.se/cgb/groups/sonnhammer/Dotter.html).[40] 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 129S6 *vs* B6 *Ly49* sequence was constructed using Advanced Pipmaker with a setting of single coverage (available at http://pipmaker.bx.psu.edu/pipmaker/).[41] To facilitate Dotter and Pipmaker analysis, some small gaps in noncoding, but highly conserved, areas were filled using the B6 sequence. This sequence constituted less than 1% of the total 129S6 regional sequence and does not appear in the GenBank submission. The B6 *Ly49* cluster sequence and the 129S6 *vs* B6 PIP (Figure 1) are also available as supplementary material at http://www.ircm.qc.ca/microsites/amakrigiannis/en/.

Detection of open reading frames was performed with GENSCAN (available at http://genes.mit.edu/GENSCAN.html).[21] 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 using PHYLIP (Phylogeny Inference Package) version 3.5c as previously described (available at http://evolution.genetics.washington.edu/phylip.html).[17,43] Phylograms were constructed with TreeExplorer version 2.12 (K Tamura, unpublished; available from http://evolgen.biol.metrou.ac.jp/TE/TE_man.html).

Acknowledgements

We gratefully acknowledge Gary Levesque for BAC subcloning and subclone sequencing, Carole Dore and Xiaolan Zhang for BAC end sequencing, Vince Forgetta for bioinformatics support, Etienne Rousselle for performing pulse-field gel electrophoresis, and Brian Wilhelm for assistance with sequence analysis and critical reading of the manuscript. 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 Mestas J, Hughes CC. Of mice and not men: differences between mouse and human immunology. *J Immunol* 2004; **172**: 2731–2738.
2 Yuhki N, Beck T, Stephens RM, Nishigaki Y, Newmann K, O'Brien SJ. Comparative genome organization of human, murine, and feline MHC class II region. *Genome Res* 2003; **13**: 1169–1179.
3 Amadou C, Kumanovics A, Jones EP, Lambracht-Washington D, Yoshino M, Lindahl KF. The mouse major histocompatibility complex: some assembly required. *Immunol Rev* 1999; **167**: 211–221.
4 Gagnier L, Wilhelm BT, Mager DL. Ly49 genes in non-rodent mammals. *Immunogenetics* 2003; **55**: 109–115.
5 Takahashi T, Yawata M, Raudsepp T *et al*. Natural killer cell receptors in the horse: evidence for the existence of multiple transcribed LY49 genes. *Eur J Immunol* 2004; **34**: 773–784.
6 Mager DL, McQueen KL, Wee V, Freeman JD. Evolution of natural killer cell receptors: coexistence of functional Ly49 and KIR genes in baboons. *Curr Biol* 2001; **11**: 626–630.
7 McQueen KL, Wilhelm BT, Harden KD, Mager DL. Evolution of NK receptors: a single Ly49 and multiple KIR genes in the cow. *Eur J Immunol* 2002; **32**: 810–817.

8 Vilches C, Parham P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu Rev Immunol* 2002; **20**: 217–251.
9 Volz A, Wende H, Laun K, Ziegler A. Genesis of the ILT/LIR/MIR clusters within the human leukocyte receptor complex. *Immunol Rev* 2001; **181**: 39–51.
10 Carrington M, Norman P. *The KIR Gene Cluster.* National Library of Medicine (US), NCBI: Bethesda, MD, 2003.
11 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.
12 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.
13 Karlhofer FM, Ribaudo RK, Yokoyama WM. MHC class I alloantigen specificity of Ly-49+ IL-2-activated natural killer cells. *Nature* 1992; **358**: 66–70.
14 McVicar DW, Burshtyn DN. Intracellular signaling by the killer immunoglobulin-like receptors (KIR) and Ly49. *Science's STKE* 2001. http://stke.sciencemag.org/cgi/content/full/OC_sigtrans;2001/75/re1.
15 Makrigiannis AP, Pau AT, Schwartzberg PL, McVicar DW, Beck TW, Anderson SK. A BAC contig map of the Ly49 gene cluster in 129 mice reveals extensive differences in gene content relative to C57BL/6 mice. *Genomics* 2002; **79**: 437–444.
16 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.
17 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.
18 Idris AH, Smith HR, Mason LH, Ortaldo JR, Scalzo AA, Yokoyama WM. The natural killer gene complex genetic locus Chok encodes Ly-49D, a target recognition receptor that activates natural killing. *Proc Natl Acad Sci USA* 1999; **96**: 6330–6335.
19 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.
20 Scalzo AA, Fitzgerald NA, Wallace CR *et al.* The effect of the Cmv-1 resistance gene, which is linked to the natural killer cell gene complex, is mediated by natural killer cells. *J Immunol* 1992; **149**: 581–589.
21 Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997; **268**: 78–94.
22 Waterston RH, Lindblad-Toh K, Birney E *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; **420**: 520–562.
23 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.
24 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.
25 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.
26 Toyama-Sorimachi N, Tsujimura Y, Maruya M *et al.* Ly49Q, a member of the Ly49 family that is selectively expressed on myeloid lineage cells and involved in regulation of cytoskeletal architecture. *Proc Natl Acad Sci USA* 2004; **101**: 1016–1021.
27 Tanamachi DM, Moniot DC, Cado D, Liu SD, Hsia JK, Raulet DH. Genomic Ly49A transgenes: basis of variegated Ly49A gene expression and identification of a critical regulatory element. *J Immunol* 2004; **172**: 1074–1082.
28 Yokoyama WM, Kehn PJ, Cohen DI, Shevach EM. Chromosomal location of the Ly-49 (A1, YE1/48) multigene family. Genetic association with the NK 1.1 antigen. *J Immunol* 1990; **145**: 2353–2358.
29 Makrigiannis AP, Pau AT, Saleh A, Winkler-Pickett R, Ortaldo JR, Anderson SK. Class I MHC-binding characteristics of the 129/J Ly49 repertoire. *J Immunol* 2001; **166**: 5034–5043.
30 The MHC Sequencing Consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* 1999; **401**: 921–923.
31 Haino M, Hayashida H, Miyata T *et al.* Comparison and evolution of human immunoglobulin VH segments located in the 3′ 0.8-megabase region. Evidence for unidirectional transfer of segmental gene sequences. *J Biol Chem* 1994; **269**: 2619–2626.
32 Kazazian Jr HH. Mobile elements: drivers of genome evolution. *Science* 2004; **303**: 1626–1632.
33 Allen E, Horvath S, Tong F *et al.* High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc Natl Acad Sci USA* 2003; **100**: 9940–9945.
34 Makrigiannis AP, Rousselle E, Anderson SK. Independent control of Ly49g alleles: implications for NK cell repertoire selection and tumor cell killing. *J Immunol* 2004; **172**: 1414–1425.
35 Held W, Kunz B. An allele-specific, stochastic gene expression process controls the expression of multiple Ly49 family genes and generates a diverse, MHC-specific NK cell receptor repertoire. *Eur J Immunol* 1998; **28**: 2407–2416.
36 Kane KP, Silver ET, Hazes B. Specificity and function of activating Ly-49 receptors. *Immunol Rev* 2001; **181**: 104–114.
37 Wilhelm BT, Mager DL. Rapid expansion of the Ly49 gene cluster in rat. *Genomics* 2004; **84**: 218–221.
38 Lander ES, Linton LM, Birren B *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; **409**: 860–921.
39 Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**: 195–202.
40 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.
41 Schwartz S, Zhang Z, Frazer KA *et al.* PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res* 2000; **10**: 577–586.
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.* Distributed by the author. Department of Genetics: University of Washington, Seattle, 1993.