Immunogenetics (2000) 51:268–280 © Springer-Verlag 2000

ORIGINAL PAPER

Annalise M. Martin · Elizabeth M. Freitas
Campbell S. Witt · Frank T. Christiansen

The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster

Received: 25 November 1999 / Revised: 10 January 2000

Abstract. Natural killer (NK) immunoglobulin-like receptors (KIRs) are a family of polymorphic receptors which interact with specific motifs on HLA class I molecules and modulate NK cytolytic activity. In this study, we analyzed a recently sequenced subgenomic region on chromosome 19q13.4 containing eight members of the *KIR* receptor repertoire. Six members are clustered within a 100-kb continuous sequence. These genes include a previously unpublished member of the *KIR* gene family *2DS6*, as well as *2DL1*, *2DL4*, *3DL1*, *2DS4*, *3DL2*, from centromere to telomere. Two additional *KIR* genes, *KIRCI* and *2DL3*, which may be located centromeric of this cluster were also analyzed. We show that the *KIR* genes have undergone repeated gene duplications. Diversification between the genes has occurred postduplication primarily as a result of retroelement indels and gene truncation. Using pre- and postduplication Alu sequences identified within these genes as evolutionary molecular clocks, the evolution and duplication of this gene cluster is estimated to have occurred 30–45 million years ago, during primate evolution. A proposed model of the duplication history of the *KIR* gene family leading to their present organization is presented.

Key words *KIR* · Multigene family · Gene duplication · Evolution

A.M. Martin · E.M. Freitas · C.S. Witt · F.T. Christiansen (✉)
Department of Clinical Immunology and Biochemical Genetics, Royal Perth Hospital, P.O. Box X2213, Perth, WA 6001, Australia
e-mail: ftchrist@cyllene.uwa.edu.au
Tel.: +61-8-92242683
Fax: +61-8-92242920

E.M. Freitas · F.T. Christiansen
Department of Pathology, University of Western Australia, Nedlands, Western Australia, Australia

Introduction

Natural killer (NK) cells express a group of receptors which recognize polymorphic motifs on HLA class I molecules. The receptors are encoded by members of the *KIR* (killer immunoglobulin-like receptor) multigene family and belong to the immunoglobulin (Ig) superfamily. KIR molecules have two or three extracellular Ig domains, a stalk or linker, as well as a transmembrane and cytoplasmic domain (for a review see Lanier 1998). Based on the number of expressed Ig domains, the proteins have been subdivided into those which contain either two Ig domains (p50 and p58), which interact with HLA-C ligands (Colonna et al. 1993; Moretta et al. 1995), or three Ig domains (p70, p70Δ, and the disulfide-linked p140), which engage HLA-B or HLA-A molecules (Dohring et al. 1996; Litwin et al. 1994; Pende et al. 1996). The cytoplasmic domain of some *KIR* genes contains an immunoreceptor tyrosine-based inhibitory motif (ITIM) which recruits and activates the tyrosine phosphatase (PTP-1C or SHP-1) (Campbell et al. 1996; Fry et al. 1996). Engagement of these KIR receptors with HLA class I ligands delivers a negative signal to the NK cell via the phosphatase pathway (Campbell et al. 1996). Those KIR molecules with shorter cytoplasmic domains lack the ITIM motif and activate NK cytolysis (Moretta et al. 1995).

A number of closely related human *KIR* cDNA sequences have been reported (reviewed in Selvakumar et al. 1997a). As a consequence of the very high nucleotide similarity between the cDNA sequences, sequences which vary at 20 or more nucleotides have been proposed to represent different genes, while those which differ at less than 9 positions represent different alleles (Steffens et al. 1998). On the basis of this classification, the KIR receptors are thought to be encoded by 12 different genes (Steffens et al. 1998; Urhberg et al. 1997). Indeed, the use of fiber fluorescent in situ hybridization (FISH) has demonstrated the presence of 11 *KIR* sequences on Chromosome (Chr) 19q13.4 (Suto et al. 1998). Within the human population there are

several haplotypes which vary in gene content and number. Using sequence-specific primers (SSPs) to amplify various *KIR* genes, we and others have shown that one of two broad groups of haplotypes (A and B) is present in Caucasian and other ethnic populations (Urhberg et al. 1997; Witt et al. 1999). The A group of haplotypes commonly consists of six *KIR* genes (*2DL4*, *3DL1*, *3DL2*, *2DL1*, *2DL3*, and *2DS4*), while the B group of haplotypes commonly carries seven loci (*2DL4*, *3DL2*, *2DL2*, *2DS1*, *2DS2*, *2DS3*, and *3DS1*). Although the presence or absence of particular genes within a haplotype is known, the order and genomic organization of these genes has not been defined. To date only the genomic organization of *2DL4* and *2DL3* has been described (Selvakumar et al. 1997b; Wilson et al. 1997).

The organization and mechanisms of generation of various multigene families have been studied by a number of groups. The duplication of single or multiple genes has been previously proposed to explain the evolution of multigene families (Ohta 1991). For instance, the *HLA* class I region is organized as duplicated multigenic segments (Gaudieri et al. 1999; Kulski et al. 1999; Shiina et al. 1999). Such duplicated segments can be detected by dotplot analysis of continuous genomic sequences (Gaudieri et al. 1997; Mizuki et al. 1997). The presence of pre- and postduplication repetitive sequences such as Alus and long interspersed repetitive sequences (LINES or L1s) have contributed to the diversification of the segments. Indeed, the presence of various Alu subfamilies in these segments can be used as a molecular clock to explain the evolution and age of these duplications events (Freitas et al., in press; Kulski et al. 1999).

In this study, we analyzed continuous genomic sequence within Chr 19q13.4 carrying the *KIR* gene cluster. We report, for the first time, the arrangement of several *KIR* genes on a single haplotype and the genomic organization of *KIR2DL1*, *3DL1*, *2DS4*, *3DL2* and the truncated *2DS6*, as well as the previously described *2DL4* and *2DL3* genes. We also report possible new alleles of the *2DL1* and *2DL4 KIR* genes. Furthermore, we examined the duplication of the *KIR* genes within this cluster based on the pattern of pre- and postduplication retroelements and constructed a model to explain the chromosomal organization and evolution of these genes.

Materials and methods

Genomic sequences

The genomic sequences from two large contigs from the United States Department of Energy Joint Genome Institute were used (Ashworth et al. 1995). The first sequence was extracted from GenBank (accession number: AC006293) and the second (BC52946) was obtained from the Lawrence Livermore National Laboratories human Chr 19 sequence database (http://www-bio.llnl.gov/bbrp/genome/ genome.html).

Identification, classification, and characterization of the KIR genes

The *KIRCI* and *2DL3* genes on AC006293 were annotated by J.E. Lamerdin and co-workers (unpublished data), while the *KIR* genes on the BAC clone were identified by one of us (AM) using BlastN analysis (http://www.ncbi.nlm.nih.gov/BLAST). The nomenclature used for the identification of the *KIR* gene family is similar to that described by Long and co-workers (http://www.ncbi.nlm.nih.gov/prow/679664748g.htm). The exon-intron structure of the genes, as well as the presence of the putative pseudoexon 3, was deduced by comparison with the cDNA sequences for *3DL1* (GenBank accession number: L41269; Colonna and Samaridis 1995), and *2DL4* (GenBank accession numbers: AF003116-23; Selvakumar et al. 1997b). Retroelements were identified using the RepeatMasker program, version 2.0 (http://ftp.genome.washington.edu/cgi-bin/RM2; A.F.A. Smit and P. Green, unpublished data). The *KIR* genomic and cDNA sequences were aligned using ClustalW (http://www. genome.ad.jp/sit-bin/nph-clustalw).

Dotplot and phylogenetic analyses

Dotplot analysis of the BC52946 sequence from positions 1 to 120000 was performed using Dotter (Sonnhammer and Durbin 1995) and a window size of 21 nucleotides. Phylogenetic analysis of exons 1–5 and the first 500 nucleotides of intron 3 was performed using the neighbor-joining method with the genetic distances estimated according to the Jukes-Cantor algorithm (Genetics Computer Group, Madison, Wis.). The bootstrap values in 500 replications were estimated using the Molecular Evolutionary Genetic Analysis (MEGA) package (Kumar et al. 1993).

Amplification of 2DL4 in human and primate samples

Amplification of exons 7–8 of the *2DL4* gene (positions 1044–1177) was performed using the primer pair 5′-TCGCCAGACACCTGCATGCTG-3′ and 5′-TGTTCACTGTTCTGTGTCCC-3′. The thermocycling conditions included denaturation at 95 °C for 5 min, 5 cycles of 95 °C for 20 s, 64 °C for 45 s, and 72 °C for 90 s, 25 cycles of 95 °C for 30 s, 60 °C for 45 s, 72 °C for 90 s, followed by 1 cycle of extension at 72 °C for 10 min and 1 cycle at 4 °C for 1 min on a CR9600 Thermocycler (Corbett Research, Mortlake, Australia). The human DNA samples were obtained from Epstein Barr virus-transformed cell lines from the 10th International Histocompatibility Workshop and the Fourth Asian-Oceania Histocompatibility Workshop panels. The primate DNA was extracted from cell lines kindly donated by the ITRI-TNA Primate Centre (Rijswijk, The Netherlands) and the Departments of Cell Biology and Microbiology, Stanford University School of Medicine (Stanford, Calif.).

Results

Genomic organization of the KIR gene family clustered on a 100-kb continuous genomic sequence

A continuous genomic sequence from Chr 19q13.4 has made it possible to examine the genomic structure and content of the *KIR* multigene family. Six *KIR* genes are located within the first 100-kb sequence of BC52946. These genes were identified by their similarity to available *KIR* sequences using BlastN. Figure 1 depicts the order of the *KIR* genes which are tightly clustered and have relatively short (1–3 kb) intergenic sequences. The

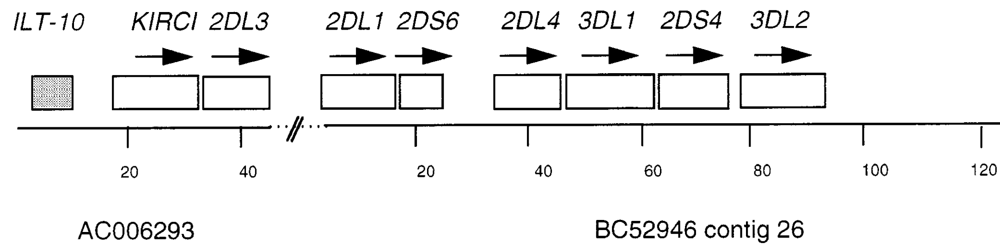

Fig. 1 Genomic organization of the *KIR* genes on Chr 19q13.4. Eight *KIR* sequences were located on two large contigs, AC006293 and BC52946. These clones were not concatenated as their sequence was derived from different human donors and lack significant overlapping sequence. The transcription direction of all *KIR* genes is indicated by an *arrow*. Note the presence of the immunoglobulin-like transcript 10 gene (*ILT-10*) centromeric to the *KIRCI* gene. No other *KIR* genes were located centromeric to *3DL2*

genes identified on the BAC clone were *KIR2DL1*, *2DS6*, *2DL4*, *3DL1*, *2DS4*, and *3DL2*, in that order from centromere to telomere. With the exception of *2DS6*, these genes are found on haplotype A (Urhberg et al. 1997; Witt et al. 2000). Two additional *KIR* genes, *KIRCI* and *2DL3*, are located on the 43-kb sequence of the cosmid AC006293. All the *KIR* genes have a centromeric to telomeric direction of transcription. In view of the fact that the BAC and cosmid clones were derived from different human donors (L.K. Ashworth, personal communication) and a lack of overlapping sequence, we were unable to concatenate these sequences. However, since the *2DL3* gene is characteristic of the A group of haplotypes (Witt et al. 2000) and the BAC clone has other genes belonging to this haplotype, we hypothesize that both clones are derived from an A haplotype. An immunoglobulin-like transcript 10 gene (*ILT-10*) was also identified and resides approximately 10 kb centromeric to *KIRCI*.

One of the new genes, *KIRCI*, lacks exon 6 (linker domain) as a result of a deletion from intron 5 to intron 6 (Fig. 2). *2DS6* is a truncated gene related to the *2D* gene family. Despite the lack of sequence for the cytoplasmic tail, its sequence in exons 1 and 3–5, is more closely related to the *2DS* genes. The nomenclature adopted here follows the last named *2DS5* cDNA sequence. The truncation of *2DS6* is also due to extensive deletions resulting in losses of exon 2 and exons 6–9, indicative of a pseudogene. The *2DL3* sequence described in this study is incomplete, as the cosmid ends within intron 6 of *2DL3*.

The KIR genes are closely related in their exon-intron structure

The exon-intron structure of the *KIR* genes identified in this study is shown in Fig. 2. It is apparent that they share a similar exon-intron structure of nine exons and eight introns, consistent with the notion of a common ancestral gene. All the genes have the classical AG-GT exon intron splice signal except for exon 3 of *2DL1* which has a CG splice site. An important feature of most of the *2D* genes is the presence of pseudoexon 3, which appears not to be expressed. Its lack of expression may be accounted by exon skipping or alternate splicing. As described for *2DL3* (Wilson et al. 1997), a premature stop codon occurs within the pseudoexon of *2DL1*, at codon 37 within the deduced first Ig domain. This may be responsible for the absence of this exon in mature cDNA transcripts. The mechanism for the absence of exon 3 in *2DS4* has yet to be determined. The most striking difference between all the genes is their length, ranging between 3.9 to 16.3 kb. This difference occurs primarily as a result of retroelement insertions and/or deletions (indels) within the introns of some of the genes. For example, *3DL2* has a unique 2.5-kb insertion within intron 6 which contains MST1D, (T)n, and LIPA3 sequences. Additionally, exons 6–9 are deleted in *2DS6*.

2DS6 and KIRCI are members of the 2D and 3D families respectively

Given that the *KIR* genes share a similar genomic structure, the degree of relatedness between the *KIR* genes was determined by aligning the derived cDNA sequences (Fig. 3). The cDNA sequences are highly conserved, with exons 3, 4, and 5 being the most divergent. This diversity may reflect the necessity of these exons to recognize a wide variety of epitopes on the highly polymorphic HLA class I molecules. There are relatively few lineage-specific substitutions. However, certain conserved nucleotide substitutions are shared by some or all of the genes within the *3D* lineage. For example, *KIRCI*, *2DL4*, *3DL2*, and *3DL1* share nucleotide substitutions at positions 172, 192, 217, 281, and 450 (Fig. 3). Similarly, the *2DL1*, *2DS4*, *2DL3*, and *2DS6* genes share nucleotide substitutions at positions 108, 519, and 601. On the basis of the observed nucleotide substitutions, the *2D* genes appear not to have diverged to the same extent as the *3D* family. Indeed, on the basis of the degree of shared similarity *2DS6* is closely related to members of the *2D* lineage, while *KIRCI* is related to the *3D* genes.

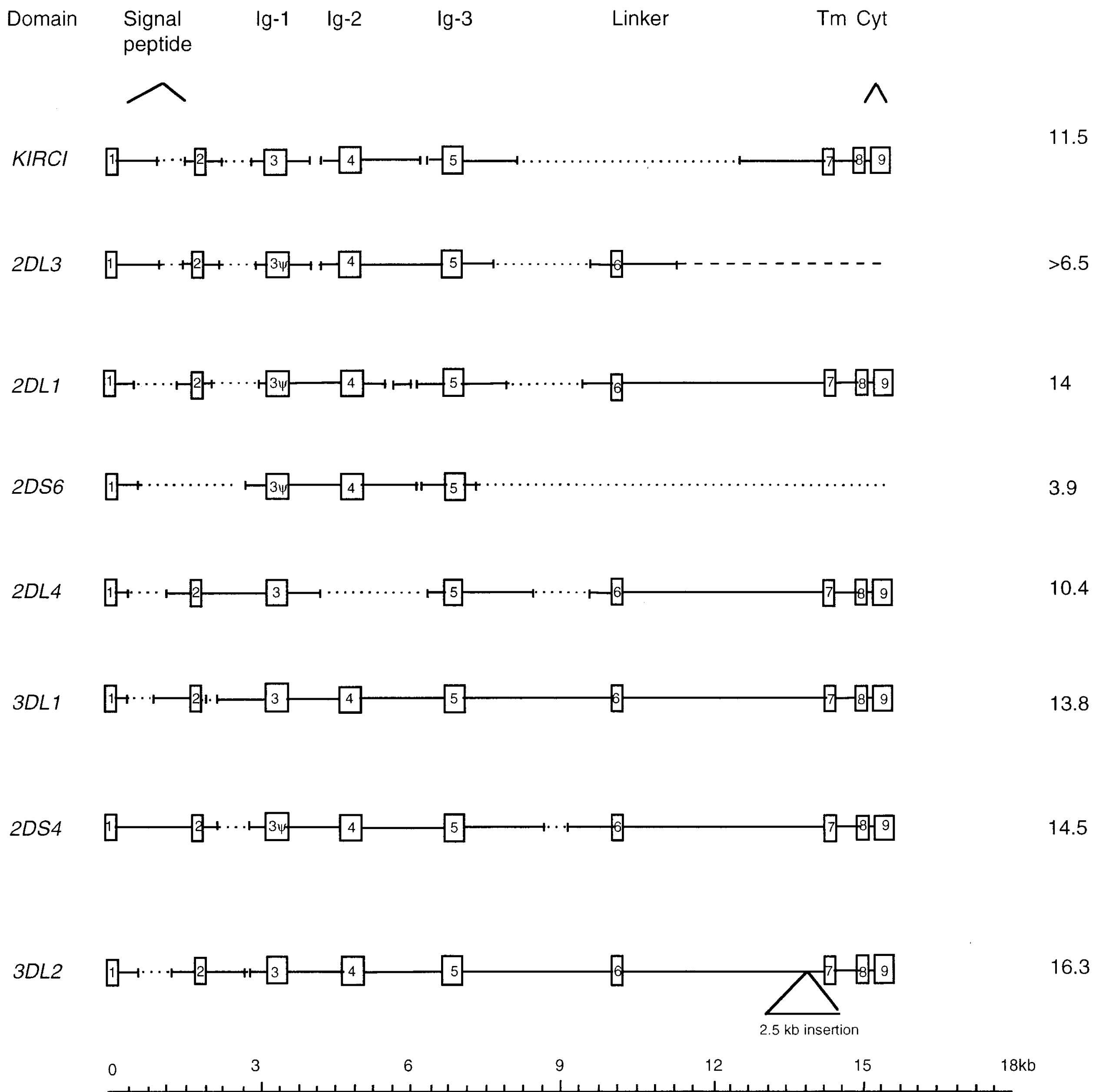

Fig. 2 Exon-intron structure of the *KIR* genes depicted according to their order from centromere to telomere. Exons are represented as *open numbered boxes*, deletions as *dotted lines*. The *dashed line* following exon 6 of *2DL3* indicates a lack of sequence due to the end of sequence coverage of AC006293. Note the presence of pseudoexon 3 (3ψ) in the *2DL1*, *2DL3*, *2DS4*, and *2DS6* genes and the approximately 2.5-kb insertion within intron 6 of *3DL2*

New alleles and polymorphisms

Comparison of the exonic sequences of the *KIR* genes with the previously reported cDNA sequences enabled us to identify new polymorphisms and potentially new alleles. The *2DL1* gene described in this study has a single G1039C substitution compared to that of a previously reported cDNA sequence (GenBank accession number: U24078). The *2DL4* gene also differs from the closely related cDNA sequence 15.212 (GenBank accession number: X97229), due to an A to G substitution at position 814 of exon 5. *2DL4* also has a single nucleotide deletion at position 1138 within exon 7 relative

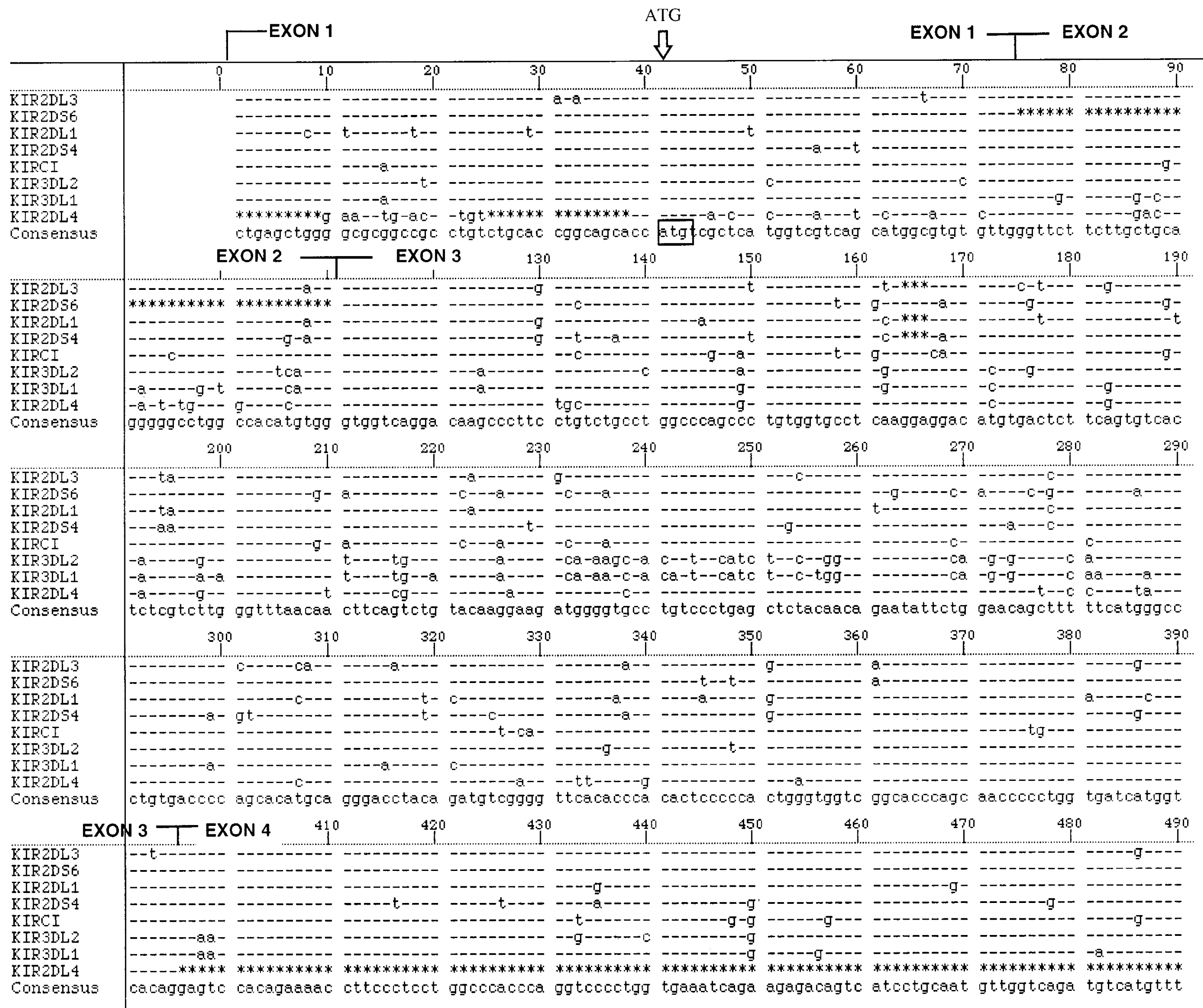

Fig. 3 Alignment of the deduced cDNA sequences of *KIR* genes. *Dashes* indicate identity, while *asterisks* represents deletions. The consensus sequence of the genes is also provided (*Consensus*). Exon boundaries were determined from the alignment of cDNA and the genomic sequences and detection of the AG-GT splice site signal

to all the *KIR* genes examined here. We have demonstrated this deletion in a number of unrelated individuals (C.S. Witt, A.M. Martin, F.T. Christiansen, unpublished data). Such a deletion would result in the introduction of a premature stop codon within exon 8 and a shortened cytoplasmic domain. Interestingly, the *2DL4* gene in this study spans only 10.4 kb, whereas that described by Selvakumar et al. (1997b) spans 12 kb. This size difference is attributable to indels within the introns. The derived *2DS4*, *3DL1*, and *3DL2* cDNA sequences are identical to their previously reported cDNA clones (GenBank accession numbers: AF002255, L76671, L41269, respectively). In addition, the sequence available for *2DL3* is identical to that of NKAT-2b (GenBank accession number: L76663). Confirmatory analyses will be required to determine whether the differences observed in *2DL1* and *2DL4* represent new alleles or sequencing errors.

Evolution of the KIR genes on this cluster

Examination of the genomic organization on the BAC clone reveals six closely linked *KIR* genes. Since these genes share a similar genomic organization and are closely related to one another, it is likely that these genes were derived from a common ancestral gene. To determine the nature of gene duplication and the evolution of the *KIR* gene family, we compared the genomic sequences of those genes on the BAC clone by dotplot analysis. Using a window size of 21 nucleotides, multiple lines of similarity were observed, corresponding to the locations of the *KIR* genes (Fig. 4A). Although these lines of similarity include intergenic se-

```
               500        510        520        530        540        550        560        570        580        590
KIR2DS6    ---------- ---------- --------a- -----t---- ---------- --c------- ---------- ---------- ---------- ----------
KIR2DL1    --a------- ---------- ---------- -----c---- ---------- ---------a -a-------- ---------- ---------- ------a---
KIR2DS4    ---------- ---------- --------a- -----ca--- ------a--- ---------- -a-------- ----t----- ---------- -----t----
KIRCI      ----g----- ---------- ---------c ac-g-----c -c-------- -g-----c-- -----c---- c--gt---c- --t-----at -----g----
KIR3DL2    ---------t ---------- ---------c -c-g-----c -c-ca----- -g-----c-- a--------- ---------- ---------- ----------
KIR3DL1    ---------t ---------a ---------c -c-------c -c-ca----- -g-----c-- a--------- ---------- ------t--- ----------
KIR2DL4    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
Consensus  gagcacttcc ttctgcacag agaggggatg tttaaggaca ctttgcgcct cattggagag ctccatgatg gggtctccaa ggccaacttc tccatcggtc

               600        610        620        630        640        650        660        670        680        690
KIR2DL3    ---------- a--------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -------c--
KIR2DS6    g-----c--- a--------- ---------- ---------- -------c-- --t------- ---------- ---------- ---------- ----------
KIR2DL1    g-----c--- a-----g--- ---------- ---------- ---------- ---------- -----g---- ---------- ---------- ----------
KIR2DS4    ---------c --t------- --a------- ---------- -------c-- ---------- ---------- ---------- ---------- ----g-----
KIRCI      ------ca-c --c------- ---------- ------tt-- ------c--- ----ta---- --g------- g--------- ---------- --------g-
KIR3DL2    --t------c --t------- --a------- ----t--t-- -------c-- ---------- ---------- ---------- -----c---- ----------
KIR3DL1    ---------t --c------- ---------- ---------- ---------- ---a------ ---------- ---------- --t--c---- --------g-
KIR2DL4    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
Consensus  ccatgatgca tgaccttgca gggacctaca gatgctacgg ttctgttact cactccccct atcagttgtc agctcccagt gaccctctgg acatcgtgat

       EXON 4 | EXON 5    710        720        730        740        750        760        770        780        790
KIR2DL3    ---------- ---------- ---------- ---------- ---------- ---t------ -----g---- ---------- ---------- ----------
KIR2DS6    t--------- -g--g----- ---------- ---------- c------t-- --a------- -----g---- ---------- ---------- ----------
KIR2DL1    --t------- ---------- ---------- --------t- ---------- ---t------ ------t--- ---------- ---------- ----------
KIR2DS4    --t------- ---------- ---------- ---------- ---------- ---------- ------t--- ---------- --------** **********
KIRCI      -gt------- ----g----- ---------- ---------- ---------- ---------- ------t--- ---------- ---------- -----tg-t-
KIR3DL2    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -t------t- ----------
KIR3DL1    --------c- ---------- ---------- ---------- -------a-- ---------- -----g---- ---------- -t-------- ----------
KIR2DL4    *****----- ---------- ----g--ta- -----g---- ---------- ---gc----- ---------- ---------- ---------a --------t-
Consensus  cacaggtcta tatgagaaac cttctctctc agcccagccg ggccccacgg ttcaggcagg agagaacgtg accttgtcct gcagctcccg gagctcctat

               800        810        820        830        840        850        860        870        880        890
KIR2DL3    ---------- ---------- ---------- ---------- -----t--t- -----g---- ---------- ---------- ---------- ----------
KIR2DS6    -----c---- -------a-- ---------- --t------- t----t---- ---------- --------t- ----c----- ----a----- ----------
KIR2DL1    ---------- ---------- ---a------ ---------- ---------- -----g---- ---------- ---------- ---t------ ----------
KIR2DS4    ********** ---------- ---a------ ---------- ---------- --------g- -gca------ ---------- ---------- ----------
KIRCI      -----t---- ---------- -----ca--- ---gg----- t-------a- ---g----tg -g------t- ---------- ----a----c ----------
KIR3DL2    -----c---- ----g----- ---a------ ---------- ---------g ---------- ---------a ---------- ---a------ ----------
KIR3DL1    ---------- ---------- --------ga ---------- ---------- --------g- ---------a ---------- ---a--t--c ----------
KIR2DL4    -----c---- ---------- ---------a ---------- t--------- ---------- -gca----t- ---------- ---------c --------t-
Consensus  gacatgtacc atctatccag ggagggggag gcccatgaac gtaggctccc tgcagtgccc aaggtcaacg gaacattcca ggccgacttt cctctgggcc

               900        910        920        930        940        950        960        970        980 EXON 5 | EXON 6
KIR2DL3    ---------- ------a--- ---------- ---------- ---------- --a------- ---------- ---------- ---------- ----------
KIR2DS6    ---------- ---------- ---------- ---------- ---------- ---------- -----g--ct t--------- ---------- ---------*
KIR2DL1    ---------- ---------- ---------- ---------- --a------- --a------- -------g-- ---------- ---------- ----------
KIR2DS4    ---------- ---------- ---------- ---------- -------g-- ---------- ---------- ------t--- ---------- -c--------
KIRCI      ---tg----- --------a- ---------- ---------- -----c-ctg ---c---c-- -----g--c- ---------- ----cc---- ----------
KIR3DL2    ---------- ---------- ---------- ---------- -----c-ctg ----g--t-- ---------- a--------- ---------- ----------
KIR3DL1    ---------- ---------- ---------- ---------- ----c----- ---------- -----g--c- ---------- ---------- ----------
KIR2DL4    ---------- -----a---- ---------- ---------- --a--ga--- ---------- -----g--c- ---------- ----c----- ----------
Consensus  ctgccaccca cggagggacc tacagatgct tcggctcttt ccgtgactct ccctacgagt ggtcaaactc gagtgaccca ctgcttgttt ctgtcacagg
```

Fig. 3 Continued

quences, it is not clear whether these genes arose from single gene duplications or segmental duplication. An investigation of the patterns of similarity between the family members within the dotplot reveals greater interruption between *2DL4* and other genes, indicative of an ancient or more divergent gene.

Relationship between the KIR genes

Phylogenetic analysis of the sequences of exons 1–5 for the *KIR* genes in this cluster was performed to examine the possible evolutionary relationships between the genes (Fig. 4B). These exons were selected, as they were present on all the *KIR* genes with the exception of the absence of exon 2 in *2DS6* and exon 4 in *2DL4*. From this analysis, the *KIRCI* and *2DL4* genes cluster together and appear to have been derived by duplication of a common ancestral gene. *2DL1* and *2DS4* are evolutionarily related, although *2DL3* is more closely related to *2DL1*, possibly reflecting a more recent duplication event. The *3DL1* and *3DL2* genes are also very closely related and form a cluster that is distinct from the other *KIR* genes. The phylogenetic relationship between the genes is similar to that reported by Lanier et al. (1997) and Valiante et al. (1997) whose analysis was based on available cDNA sequences. As the genomic sequence for these genes was available, it was possible to analyze the relationship between the genes for the first time, using intronic sequences which are less subject to selective pressure. The first 500 nucleotides of intron 3 from each gene were examined. In this analysis, a similar relationship was observed between the genes. The *3DL1* and *3DL2* genes remained clustered together, as did *2DL3* and *2DL1* (Fig. 4B). The *2DS4* and *2DS6* genes also maintained their relationship with *2DL3* and *2DL1*.

The presence of ancestral and postduplication retroelements can assist in determining the evolution of the KIR gene family

The *KIR* genes on this cluster can be classified into a number of groups based on the presence of shared and

```
                                                          EXON 6 ──┬── EXON 7
               1000       1010       1020       1030       1040       1050       1060       1070       1080       1090
KIR2DL3    ---------- ---------- ---------- ---------- ---g------ ********** ********** ********** ********** **********
KIR2DS6    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
KIR2DL1    ---------- ---------- ---------- ---------- ---------- --------c- ---------- a--------- ---------- ----------
KIR2DS4    ---------- ---------- ---------- ---------- ---------- ---------- ---------t ---------- ---------- -----aa---
KIRCI      ********** ********** ********** ********** ********** -----tg--- ----a----- ---------- ---------- ----------
KIR3DL2    ---------- -g-------- ---------- a--------- ------t-t- ---t-tg--- ---------t ---------- ---------- -------t--
KIR3DL1    ---------- -g-------- ---------- a--------- ------t-t- ---------- ---------- a--------- ---------- ----------
KIR2DL4    ---------t -g-------- ---------- ---------- -t------t- ---t-g---- ---------t -c-g-----a --ta------ --c-------
Consensus  aaacccttca aatagttggc cttcacccac tgaaccaagc tccaaaaccg gtaaccccag acacctgcac gttctgattg ggacctcagt ggtcatcatc
                                                              EXON 7 ──┬── EXON 8
               1100       1110       1120       1130                         1160       1170       1180       1190
KIR2DL3    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
KIR2DS6    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
KIR2DL1    -------*** ---------- ---------- ---------- ---------- ---------- --g------- --------t- ------a--- -------c--
KIR2DS4    -ct------- ---------- ---------- ---------- -----g---- ---------- ---------- ---------- ---------- ----------
KIRCI      -c---tg-t- ---------- ---------- ---------- -tg------- ---------- -t-------- ---------- ---------- ----------
KIR3DL2    -------t-c ---------- ---------- t--------- ---------- ---g------ ---------- ---------- ---g---g-- ----------
KIR3DL1    -------t-c ---------- ---------- ----t----- ---------- ---g------ ---------- ---------- ---------- -------cc-
KIR2DL4    -----t---- ----t-c--- ---------- ---------- -------*-- ---------- ---------a ---------- ---g--ac-- ----------
Consensus  ctcttcacca tcctcctctt ctttctcctt catcgctggt gctccaacaa aaaaaatgct gctgtaatgg accaagagcc tgcagggaac agaacagtga
      EXON 8 ──┬── EXON 9
                                 1220       1230       1240       1250       1260       1270       1280       1290
KIR2DL3    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
KIR2DS6    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
KIR2DL1    -t-------- ---------- ---------- ---------- ----a----- ---a------ ---------- ---------- ---------- ----------
KIR2DS4    ---------- t--------- -------a-- --------t- -------t-a ---------- -t-------- ---------- ---------- -----g----
KIRCI      ----g----- ---------- ---------- ---------- ---------- ---a------ ---------- ---------- ---------- ----------
KIR3DL2    -t--gc---- ---------- ---------- ---------- g--------- ---------- ---------t ---------- ----g----- ----------
KIR3DL1    --------ct -c--cc---- ---------g ---------- ---------- ---------- ---------- ---------- ---------- ----------
KIR2DL4    ----g----- ---------- ---------- ---------- ---------- ---------- --a------- ---------- ------g--- ----------
Consensus  acagcgagga ctctgatgaa caagaccctc aggaggtgac atacgcacag ttggatcact gcgttttcac acagagaaaa atcactcgcc cttctcagag

               1300       1310       1320       1330       1340       1350       1360       1370       1380       1390
KIR2DL3    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
KIR2DS6    ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
KIR2DL1    ---------- ---------- ---t--t--- ---*------ ---------- ------t--- ---------- ---------- ---------- ----------
KIR2DS4    ---------- ---------- --------a- ---*--ta-- ---------- ---------- ---------- -----t---t ---c------ ----------
KIRCI      ---------- ---------- ---------- ---a------ ---------- -------g-- ---------- -----t---t ---ct----- ----------
KIR3DL2    ---------- ----t----- ---------- ---*------ ---------- ---------- ---------- ---------- ---c------ ----------
KIR3DL1    ---------- -----t---- ------t-t- ---*------ ---------- ---a------ ---------- ---------- ---------- ----------
KIR2DL4    -ag-----g- ---t------ ---------- --g*t-ta-- ---------- ---------- --g-****-- ---------- ---------- ----------
Consensus  gcccaagaca cccccaacag ataccagcgt gtaacacgga acttccaaat gctgagccca gatccaaagt tgtctcctgc ccatgagcac cacagtcagg

               1400       1410       1420       1430       1440       1450       1460       1470       1480       1490
KIR2DL3    ********** ********** ********** ********** ********** ********** ********** ********** ***
KIR2DS6    ********** ********** ********** ********** ********** ********** ********** ********** ***
KIR2DL1    ---------c g--------- ---------- ---------- ---------- ---------- ---------- ---------- ---
KIR2DS4    ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---
KIRCI      ------t--- ---------- -------t-- ---------- ---------- ---------- ---------- -------c-- ---
KIR3DL2    t--------- g-t------- ---------- ---------- -----a---- -----a---a ---a------ ---------- ---
KIR3DL1    --------ac g--------- ---------- ---------- ------a--- ---------- ---------- ---------- ---
KIR2DL4    ------t--- ---------- ---------- ---------- --c--a-c-- --------ta ---------- ---------- ---
Consensus  ccttgagggg atcttctagg gagacaacag ccctgtctca aaaccgggtt gccagctccc atgtaccagc agctggaatc tga
```

Fig. 3 Continued

unique retroelements. The nature and location of ancestral and postduplication retroelements within the *KIR* genes are shown in Fig. 4C. The most notable feature is the presence of several ancestral retroelements (MLT1D, MSTB, MER70, and L1MA4) located in similar positions within each *KIR* gene. Furthermore, a MER2 element was also identified within the 3′ intergenic sequence of each intact *KIR* gene. Two copies of a simple $(TAGA)_n$ repeat are also observed within intron 4 of all the *KIR* genes, except *2DL4* which has a deletion in this region. Based on the evolutionary time of amplification and fixation of the MTL1D elements within the genome (Smit 1996), we propose that the ancestral *KIR* gene existed approximately 50–100 million years ago (mya).

Both *3DL1* and *3DL2* share a common Alu sequence belonging to the Sx subfamily located approximately 20 nucleotides upstream of the ancestral L1MA4 element. This is consistent with the phylogenetic analysis which indicates that these genes are closely related and evolved from a pathway distinct from the other *KIR* genes. The *2DL1*, *2DS4*, and *2DL3* genes have an Alu sequence belonging to the Sg subfamily located approximately 100 nucleotides 5′ of the L1MA4 element, but on the negative strand, indicative of a shared lineage. The *3D*-related genes can be subdivided into two categories on the basis of the type and number of Alus located within the ancestral L1MA4 element. For example, the *2DL4* and *KIRCI* genes have an AluSq sequence, while the *3DL1* and *3DL2* genes have an Sp and two Sx Alu sequences. Furthermore, the *2D* genes including *2DL1*, *2DL3*, and *2DS4* share the Sp and two Sx elements within the L1MA4 element similar to the *3DL1* and *3DL2* genes, indicating a common ancestry between the *3DL1* and *3DL2* genes and the members of the *2D* genes on this cluster. The presence of Alu sequences belonging to the S subfamily suggests that the *KIR* gene family evolved between 31 and 44 mya during primate evolution.

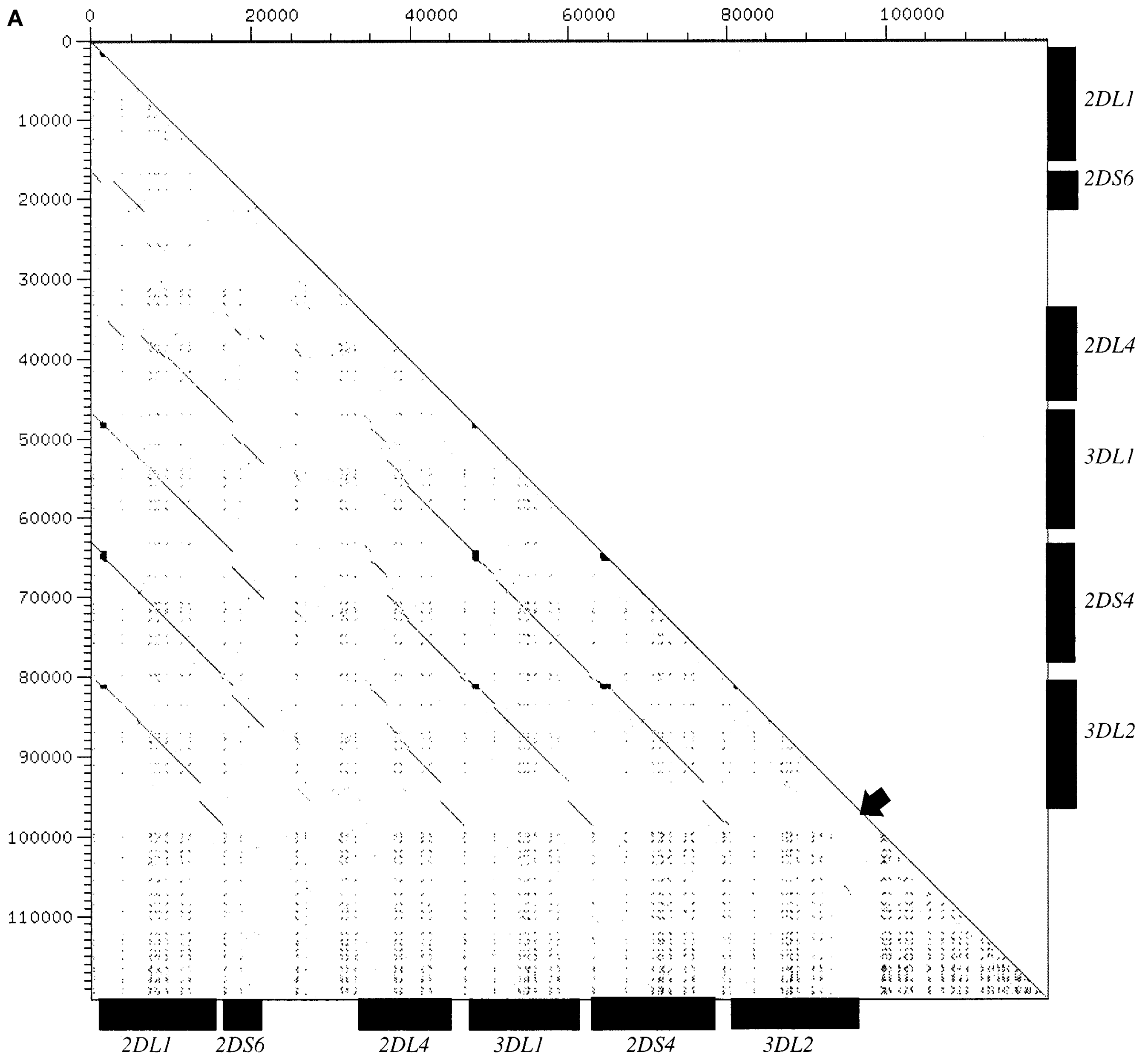

Fig. 4A–C The *KIR* genes on BC52946 are closely related. **A** Dotplot analysis comparing the first 120 kb of sequence from BC52946 against itself using a window of 21 nucleotides. Below the line of identity (*arrow*) there are multiple lines of similarity corresponding to the location of the known *KIR* genes (*black boxes*). Most of the interruptions between these lines represent gene truncations and retroelement indels, particularly between *2DS6* and *2DL4*. **B** Phylogenetic relationship of exons 1–5 and the first 500 nucleotides of intron 3 of the *KIR* genes showing that the *3D* and *2D* genes belong to distinct lineages. The phylogenetic trees were generated using Jukes-Cantor as the genetic distances, and a neighbor-joining algorithm. The bootstrap values obtained using 500 replications and the *bar* representing 1 nucleotide substitution are also depicted. The clustering of *3DL1* and *3DL2* suggests duplication from an immediate common ancestor. Similarly, *2DL4* and *KIRCI* may share a common ancestor, while *2DL3* and *2DL1* are duplication products of a common ancestor. **C** Schematic representation of the repeat elements within the *KIR* genes and their flanking sequence is indicative of their evolution from a common progenitor. The presence of repeat elements within the *KIR* genes (MLT1D, MER70, MSTB) and their flanking sequence (MER2) indicates their evolution from a common progenitor. *KIRCI* and *2DL4* may be duplication products of an immediate ancestor with an L1MA4 element containing an AluSq insertion. The *2D* and *3D* genes have diverged from an ancestor containing an L1MA4 element with an Sp and two Sx Alu sequences

Proposed evolutionary model for the KIR gene family

While the exact mechanism leading to the current arrangement of *KIR* genes is unknown, we propose a model of the evolution of the *KIR* gene family based on sequence homology, phylogenetic analysis, and on the presence of paralogous and postduplication retroelements (Fig. 5). The relative age of radiation and fixa-

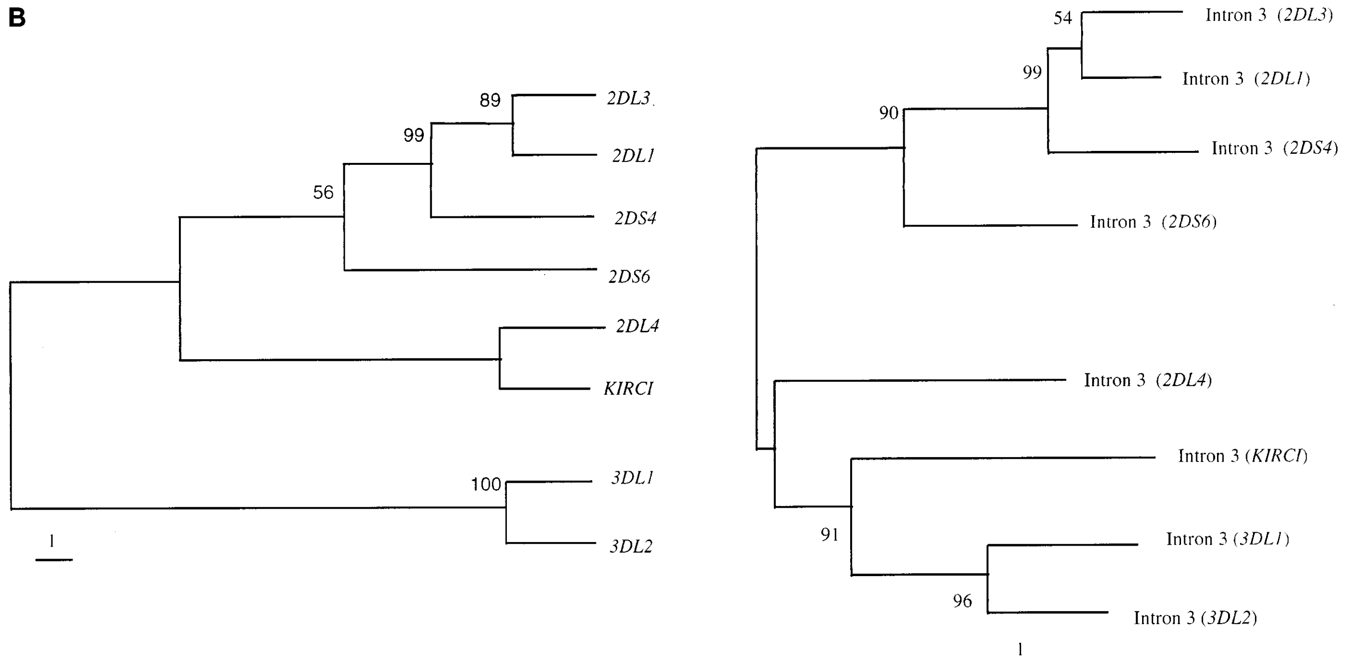

Fig. 4B

tion of the Alu subfamilies was used to date some of the duplication events. The ancestral *KIR* gene contained the MLT1D, MER70, MSTB, and L1MA4 retroelements, which were inserted approximately 50–100 mya. The initial gene duplication of the *KIR* progenitor occurred prior to the insertion of the AluSq sequence into *2DL4* approximately 44 mya. Subsequently, approximately 37 mya, AluSp and two AluSx elements were inserted into the L1MA4 element of the *2D*/*3D* progenitor. The *2D*/*3D* ancestral gene then underwent a gene duplication event prior to the insertion of the AluSx sequence upstream of the L1MA4 element in the *3D* progenitor. Approximately 31 mya, an Alu belonging to the Sg subfamily inserted upstream of the ancestral L1MA4 element of the *2D* progenitor. A tandem duplication of the *2DS* and *3D* ancestral genes resulted in the formation of the present day *2DS6*, *3DL1*, *2DS4*, and *3DL2* genes. The available data do not make clear whether this tandem duplication occurred as a segment or as independent gene duplication events. A further duplication of *2DS4* resulted in *2DL1*, with a translocation of *2DL4* between *2DS6* and *3DL1* giving rise to the present arrangement of the *KIR* cluster. The model does not try to account for the evolution of *KIRCI* and *2DL3*, as these genes may not be present on the same haplotype and may have evolved independently. However, based on the phylogenetic analysis and the paralogous retroelements, *KIRCI* and *2DL4* are predicted to be duplication products of a common ancestor, with *2DL3* probably a duplication product of the *2DL1* ancestor. The model presented here is provisional and will be refined as more sequence and haplotype data become available.

Preliminary evidence suggests that 2DL4 emerged during primate evolution

To obtain further information on the evolutionary history of the *KIR* genes, we examined human and nonhuman primates for the presence of *2DL4*, which is probably the most divergent gene in this cluster, by amplifying exons 7–8 using *2DL4*-specific primers (Fig. 6). *KIR2DL4* was found in all of the human subjects examined, in one of the chimpanzees (R88-15166), and as a weak product in the other chimpanzee species (R88-15165). No amplification was observed in the macaque species. Furthermore, the chimpanzees showed quantitative differences in the amplification of *2DL4*, which may reflect polymorphisms within the primer sites. While the absence of amplification in the macaque suggests the absence of this gene among the older nonhuman primates, we cannot exclude the possibility of nucleotide substitution(s) within the primer sequences. The use of degenerate primers may help distinguish between these two possibilities. As *2DL4* is proposed as one of the oldest *KIR* genes, these data support the hypothesis that the *KIR* gene family evolved during primate evolution, possibly during the divergence of the Old and New World monkeys.

Discussion

We show for the first time the genomic organization of eight *KIR* genes. Most of the genes on the two contigs

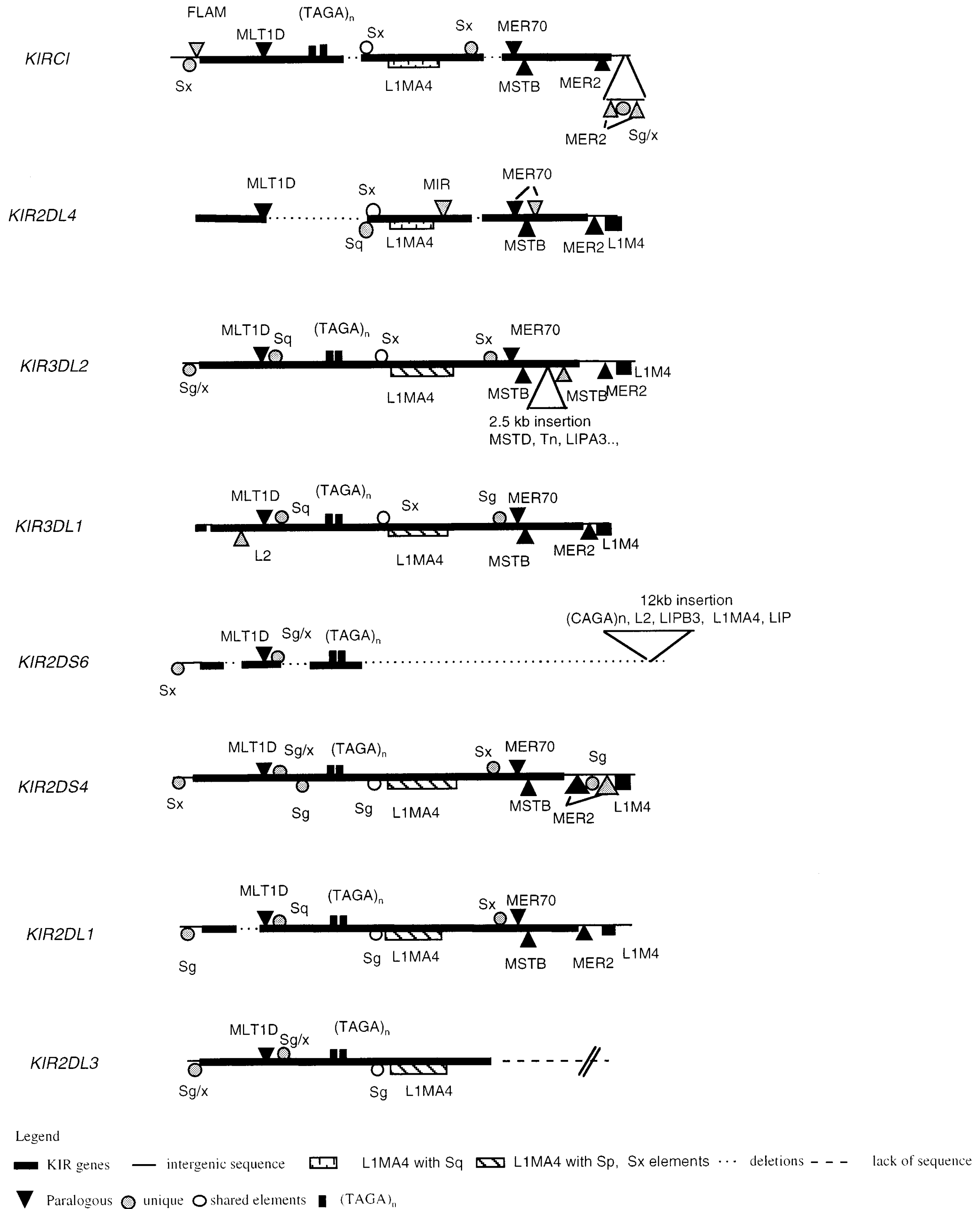

Fig. 4C

are present on our previously predicted A group of haplotypes which are the most common haplotypes within the Caucasian population. We were unable to detect additional *KIR* genes within the 100- to 200-kb sequences at the telomeric end of the cluster despite careful searches. The genes present immediately centromer-

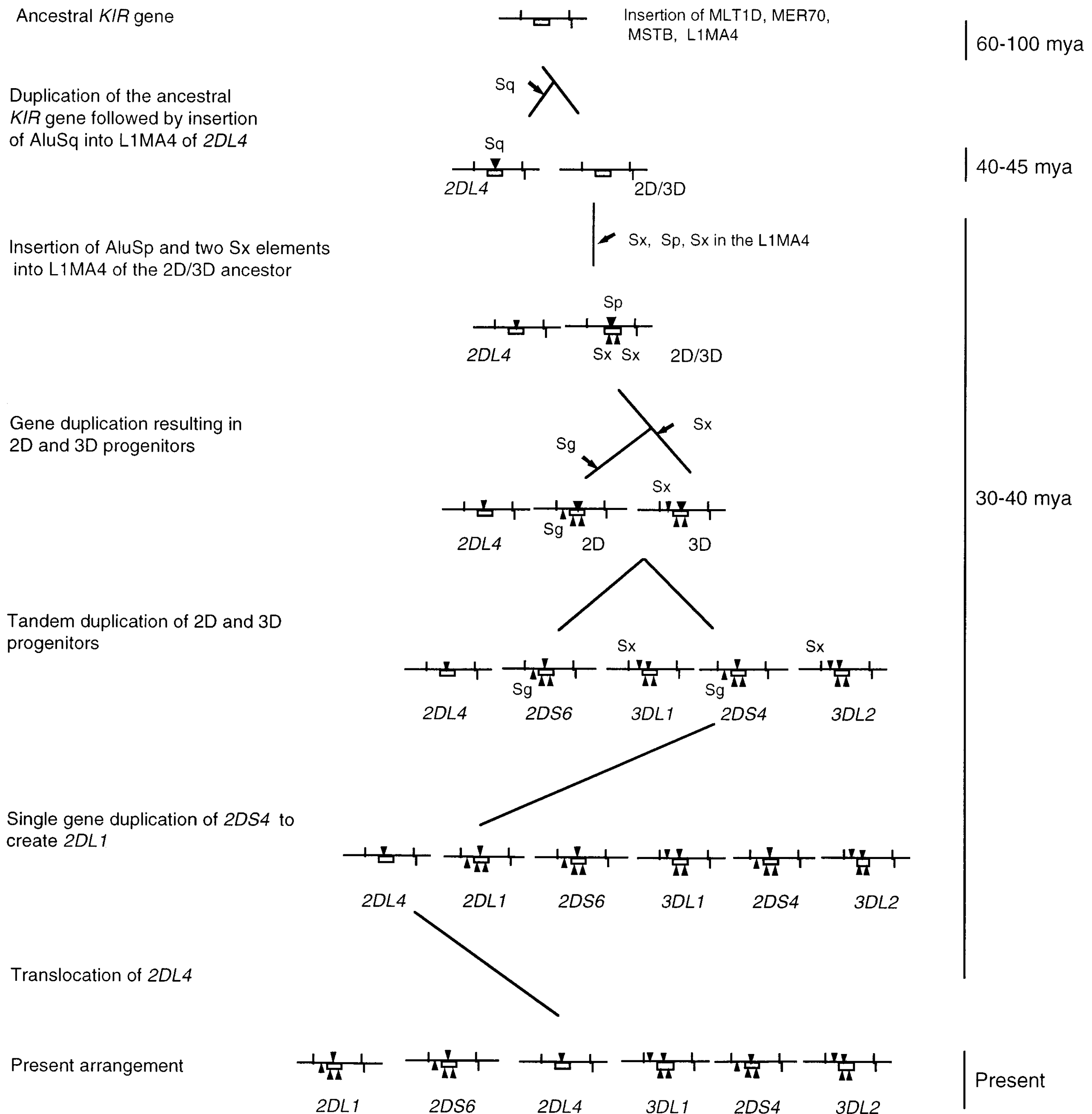

Fig. 5 Model of the evolution of the *KIR* genes present on BC52946. The *solid horizontal lines* represents *KIR* genes, with *vertical lines* representing the ancestral retroelements. Differential insertion of various Alu subclasses (*triangles*) into the ancestral L1MA4 element helps distinguish the lineages

ic to this BAC cluster have yet to be determined. However, we predict that *KIRCI* and *2DL3* may be located centromeric of the cluster of *KIR* genes on the BAC clone, as *2DL3* has been reported to be characteristic of the predicted A haplotypes. The presence of two additional genes not previously recognised on the A haplotypes (*KIRCI* and *2DS6*) can be attributed to the inability to detect them using the currently available SSP primers. The failure to detect their cDNA sequences could be due to their being pseudogenes.

Variable numbers of *KIR* genes have been reported to constitute a haplotype. Eleven *KIR* sequences were detected on a single chromosome using fibre FISH (Suto et al. 1998). Furthermore, at least six genes on the A group of haplotypes and seven on B group haplotypes have been detected, including the presence of a recombination event between haplotypes (Urhberg et al. 1997; Witt et al. 2000). Hence, haplotypes consisting of variable combinations of the *KIR* gene repertoire, including truncated or pseudogenes and as yet undes-

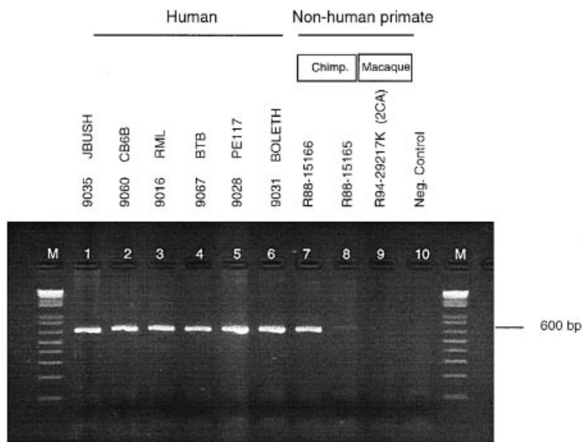

Fig. 6 The *2DL4* gene is present in human subjects, with quantitative differences among the chimpanzee samples. Amplification of exons 7–8 was performed using primers and conditions as described in Materials and methods. The amplified products were analyzed on a 2% agarose gel in 0.5×Tris borate EDTA buffer (Sambrook et al. 1989). A weakly amplified product of the expected length was observed in one of the chimpanzees (R88-15165), while no product was observed in the macaque [R94-29217K (2CA)] (*M* molecular-weight ladder)

cribed genes, are predicted to be generated by gene duplication and recombination events.

The organization of the B group of haplotypes is not known. It shares some genes with the A haplotypes, including the *2DL4* and *3DL2* loci, with the other genes (*2DL2*, *2DS2*, *2DS1*, *2DS3*, and *3DS1*) being characteristic for this group. Whether the *2DS6* or *KIRCI* genes identified here are also present on the B haplotypes is not known, but studies using gene-specific primers on individuals homozygous for this haplotype are in progress.

The nucleotide sequence of *2DL4* on this haplotype differs from the previously reported cDNA sequences. A deletion of C1138 occurs within exon 7 of *2DL4* which results in a nonameric poly(A) sequence at the exon 7 boundary. We have observed the same sequence in a number of randomly selected Caucasian individuals (C.S. Witt, A.M. Martin, F.T. Christiansen, unpublished data). This deletion alters the reading frame and is predicted to generate a premature stop codon resulting in a membrane-bound protein with a truncated cytoplasmic domain. The fact that this deletion is present at a high frequency within a panel of 50 Caucasians suggests a functional significance.

The difference in genomic structure of *2DL4* (*KIR103*) genes on this haplotype and that described by Selvakumar and co-workers (1997b), is primarily in the length of introns 2, 4, 5, and 8 which are approximately 0.8, 0.8, 2.5, and 0.1 kb on this haplotype compared to 1, 1, 3.2, and 0.5 kb, respectively, as reported by Selvakumar and co-workers (1997b). The differences occur within the introns and probably reflect retroelement indels. These differences suggest that there are at least two *2DL4* alleles present in the current human population which are evolving independently. The genomic organization of *2DL3* is similar to that described by Wilson and co-workers (1997). We have also observed a similar stop codon at codon 37 in pseudoexon 3 within the deduced amino acid sequence of *2DL3*. The contig used in our study may also contain a new allele of *2DL1* which differs from the cI-47.11 sequence at position 1039 if the polymorphism at this site is independently confirmed. Furthermore, the nomenclature of *KIRCI* needs to be updated. The current nomenclature of the *KIR* genes is based on the number of expressed Ig-like extracellular domains. How this classification should apply to pseudogenes is unclear. Since *KIRCI* has three Ig-like exons and clusters with the *3D* genes by phylogenetic analysis, it probably should be considered a *3D* gene.

We estimated the age of the duplication of the *KIR* genes based on the ages of the radiation and fixation of the AluS subfamily having occurred 30–45 mya (Mighell et al. 1997; Shen et al. 1991), during the divergence of the Old and New World monkeys. The inability to amplify *2DL4* from the older macaque species is consistent with this model. Furthermore, the presence of *KIR* sequences in a number of nonhuman primate species using RFLP analysis (Valiante et al. 1997) is consistent with our model, and suggests that the *KIR* genes are closely related and have evolved relatively recently.

Acknowledgements We wish to acknowledge the staff of the United States Department of Energy Joint Genome Institute for providing the sequence data used in this study, whose work is under the auspices of the Department of Energy under Contract no. W-7405-ENG-48. We also thank Professor Marius Giphart, Dr. Silvana Gaudieri, and Dr. Yurek Kulski for their critical review and helpful suggestions, and D. R. Bontrop and Professor A. Hughes for kindly providing the primate material used in this study. This work was supported by the National Health and Medical Research Council (Australia) and the Medical Research Foundation, Royal Perth Hospital, Perth, Western Australia. This is publication number 9923 of the Department of Clinical Immunology, Royal Perth Hospital.

References

Ashworth LK, Batzer MA, Brandiff B, Branscomb E, Jong P de, Garcia E, Garnes JA, Gordon LA, Lamerdin JE, Lennon G, Mohrenweiser H, Olsen AS, Slezak T, Carrano AV (1995) An intergrated metric physical map of human chromosome 19. Nat Genet 11:422–427

Campbell KS, Dessing M, Lopez-Botet M, Calla, M, Collona M (1996) Tyrosine phosphorylation of a human killer inhibitory receptor recruits protein tyrosine phosphatase 1C. J Exp Med 184:93–100

Colonna M, Samaridis J (1995) Cloning of immunoglobulin-superfamily members associated with HLA-C and HLA-B recognition by human natural killer cells. Science 268:405–408

Colonna M, Bowellino G, Falco M, Ferrara GB, Strominger JL (1993) HLA-C is the inhibitory ligand that determines dominant resistance to lysis by NK1- and NK2-specific natural killer cells. Proc Natl Acad Sci USA 90:12000–12004

Dohring C, Scheidegger D, Samaridis J, Cella M, Colonna M (1996) A human killer inhibitory receptor specific for HLA-A. J Immunol 156:3098–3101

Freitas EM, Gaudieri S, Zhang WJ, Kulski JK, Bockxmeer F van, Christiansen FT, Dawkins RL (in press) Duplication and diversification of the apolipoprotein CI (ApoCI) genomic segment in association with retroelements. J Mol Evol

Fry AM, Lanier LL, Weiss A (1996) Phosphotyrosines in killer cell inhibitory receptor molecules of NKB1 are required for negative signalling and for association with protein tyrosine phosphatase 1C. J Exp Med 184:295–300

Gaudieri S, Giles KM, Kulski JK, Dawkins RL (1997) Duplication and polymorphism in the MHC: Alu generated diversity and polymorphism within the *PERB11* gene family. Hereditas 127:37–46

Gaudieri S, Kulski JK, Dawkins RL, Gojobori T (1999) Different evolutionary histories in two subgenomic regions of the major histocompatibility complex. Genome Res 9:541–549

Kulski JK, Gaudieri S, Martin A, Dawkins RL (1999) Coevolution of PERB11 (MIC) and HLA class I genes with HERV-16 and retroelements by extended genomic duplication. J Mol Evol 49:84–97

Kumar S, Tamura K, Nei M (1993) MEGA: molecular evolutionary genetic analysis. Comput Applic Biosci 10:189–191

Lanier LL (1998) NK cell receptors. Annu Rev Immunol 16:359–393

Lanier LL, Corliss B, Phillips JH (1997) Arousal and inhibition of human NK cells. Immunol Rev 155:145–154

Litwin V, Gumperz JE, Parham P, Phillips JH, Lanier LL (1994) NKB1: a natural killer cell receptor involved in the recognition of polymorphic HLA-B molecules. J Exp Med 180:537–543

Mighell AJ, Markham AF, Robinson PA (1997) Alu sequences. FEBS Lett 417:1–5

Mizuki N, Ando H, Kimura M, Ohno S, Miyata S, Yamazaki M, Tashiro H, Watanabe K, Ono A, Taguchi S, Sugawara C, Fukuzumi Y, Okumura K, Goto K, Ishihara M, Nakamura S, Yonemoto J, Kikuti YY, Shiina T, Chen L, et al. (1997) Nucleotide sequence analysis of the HLA class I region spanning the 237-kb segment around the HLA-B and -C genes. Genomics 42:55–66

Moretta A, Sivori S, Vitale M, Pende D, Morelli L, Augugliaro R, Bottino C, Moretta L (1995) Existence of both inhibitory (p58) and activatory (p50) receptors for HLA-C molecules in human NK cells. J Exp Med 182:875–884

Ohta T (1991) Multigene families and evolution of complexity. J Mol Evol 33:34–41

Pende D, Biassoni R, Cantoni C, Verdiani S, Falco M, Donato C di, Accame L, Bottino C, Moretta A, Moretta L (1996) The natural killer cell receptor specific for HLA-A allotypes: a novel member of the p58/p70 family of inhibitory receptors that is characterised by 3 immunoglobulin domains and is expressed as a 140 kDa disulphide linked dimer. J Exp Med 184:505–518

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbour Press, Cold Spring Harbor, NY

Selvakumar A, Steffens U, Dupont B (1997a) Polymorphism and domain variability of human killer cell inhibitory receptors. Immunol Rev 155:183–196

Selvakumar A, Steffens U, Palanisamy N, Chaganti RSK, Dupont B (1997b) Genomic organization and allelic polymorphism of the human killer cell inhibitory receptor gene *KIR*103. Tissue Antigens 49:564–573

Shen RM, Batzer MA, Deininger PL (1991) Evolution of the master Alu gene(s). J Mol Evol 33:311–320

Shiina T, Tamiya G, Oka A, Takashima T, Yamagata E, Kikkawa E, Iwata M, Tomizawa M, Okuwaki N, Kuwano Y, Watanabe K, Fukuzumi Y, Itakura S, Sugawara C, Ono A, Yamazaki M, Tahiro HL, Ando A, Ikemura T, Soeda E, et al. (1999) Molecular dynamics of MHC genesis unravelled by sequence analysis of the 1,796,938 bp HLA class I region. Proc Natl Acad Sci USA 96:13282–13287

Smit AFA (1996) Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res 21:1863–1872

Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:GC1–GC10

Steffens U, Vyas Y, Dupont B, Selvakumar A (1998) Nucleotide and amino acid sequence alignment for human killer cell inhibitory receptors (*KIR*). Tissue Antigens 51:398–413

Suto Y, Ishikawa Y, Kasahara M, Kasai F, Yabe T, Akaza T, Juji T (1998) Gene arrangement of the killer cell inhibitory receptor family of human chromosome 19q13.4 detected by fibre-FISH. Immunogenetics 48:235–241

Uhrberg M, Valiante NM, Shum BP, Shilling HG, Lienert-Weidenbach K, Corliss B, Tyan D, Lanier LL, Parham P (1997) Human diversity in killer cell inhibitory receptor genes. Immunity 7:753–763

Valiante NM, Lienert K, Shilling HG, Smits BJ, Parham P (1997) Killer cell receptors: keeping pace with MHC class I evolution. Immunol Rev 155:155–164

Wilson MJ, Torkar M, Trowsdale J (1997) Genomic organization of a human killer cell inhibitory receptor gene. Tissue Antigens 49:574–579

Witt CS, Dewing C, Sayer DC, Uhrberg M, Parham P, Christiansen FT (1999) Population frequencies and putative haplotypes of the killer cell inhibitory receptor (*KIR*) sequences or evidence for recombination. Transplantation 68:1784–1789