CALL FOR PAPERS | Comparative Genomics

Comparative analysis of the paired immunoglobulin-like receptor (PILR) locus in six mammalian genomes: duplication, conversion, and the birth of new genes

Michael D. Wilson,1 Joseph Cheung,2 Duane W. Martindale,1 Stephen W. Scherer,2,3 and Ben F. Koop1

1Centre for Biomedical Research, Department of Biology, University of Victoria, British Columbia;
2Department of Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario; and 3Department of Molecular and Medical Genetics, University of Toronto, Ontario, Canada

Submitted 16 November 2005; accepted in final form 17 August 2006

Wilson MD, Cheung J, Martindale DW, Scherer SW, Koop BF. Comparative analysis of the paired immunoglobulin-like receptor (PILR) locus in six mammalian genomes: duplication, conversion, and the birth of new genes. *Physiol Genomics* 27: 201–218, 2006. First published August 22, 2006; doi:10.1152/physiolgenomics.00284.2005.—Many aspects of the immune system are controlled by homologous cell surface receptors that mediate inhibitory and activating pathways. The paired immunoglobulin-like receptor (PILR) locus at 7q22 encodes both PILRA, an inhibitory receptor, and PILRB, its activating counterpart. Mouse Pilrb1 is a novel immune system regulator, and its ligand CD99 participates in the recruitment of T-cells to inflamed tissue. We characterized the PILR locus in six mammalian genomes and investigated the structure and mRNA expression of human PILRB. Synteny at the PILR locus is conserved in the human, chimpanzee, dog, mouse and rat genomes. The absence of the PILR locus in opossum and chicken genomes suggests it arose after the divergence of placental and nonplacental mammals. In humans, a Williams-Beuren syndrome-related segmental duplication has created a complex chimeric transcript representing the predominantly expressed form of PILRB. Unlike PILRA, PILRB transcripts were detected in a wide variety of tissues including cells of the lymphoid lineage. In the mouse genome, a second activating gene, Pilrb2, and six pseudogenes were found. Extensive gene duplications in the rat genome have resulted in at least 27 Pilrb genes and or pseudogenes. Abundant gene duplication events involving novel CD99-related genes were also detected in the rat genome. In addition to duplication, we show that gene conversion has played a persistent role in the evolution of the PILR genes. Overall, we demonstrate that the PILR locus is dynamically evolving via multiple evolutionary mechanisms in several mammalian genomes.

PILRB; CD99; segmental duplication; poliovirus receptor-related immunoglobulin domain containing; 7q22

HOMOLOGOUS PAIRS OF LEUKOCYTE cell surface receptors regulate many immune responses by mediating either activating or inhibitory signals (29). The genomic organization of these highly similar activating and inhibitory receptor pairs give insight into their evolution and explain how they can bind common ligands yet illicit different physiological responses. Many paired receptor genes are rapidly evolving through gene duplication and possibly gene conversion (20, 66, 67, 70). This rapid evolution is thought occur in response to evolutionary pressure provided by pathogens (4).

Paired receptors are found in two major gene families: the c-type lectin family and the immunoglobulin-like superfamily (IgSF) (29). In both families, inhibitory and activating receptors both have an extracellular ligand binding domain, a transmembrane region, and a cytoplasmic domain. While the similarity in the extracellular domain of inhibitory and activating receptors is readily detectable, inhibitory receptors are distinguished by a long cytoplasmic tail with an immunoreceptor tyrosine-based inhibitory motif (ITIM). The ITIM-bearing cytoplasmic tail can recruit proteins such as tyrosine phosphatases, which in most cases downregulate immune responses (reviewed by (29)). In contrast, the activating receptors typically have a truncated cytoplasmic domain and a charged amino acid residue in the transmembrane spanning region. The charged residue interacts with molecules that can often activate cellular processes. These molecules include the immunoreceptor tyrosine-based activation motif (ITAM)-bearing molecules such as TYROBP (formerly DAP12), CD3ζ, and FcγRI; as well as the non-ITAM-bearing molecules such as DAP10, SLAMF6, and CD244 (reviewed by (29)).

Insight into the evolution and function of paired receptors comes from the study of inhibitory and activating receptors expressed in natural killer (NK) cells (reviewed by Refs. 30, 59). For example, rapid evolution through gene duplication and loss has been well described for the killer cell lectin-like receptor, subfamily A genes (KlrA; formerly called Ly49) in rat (20, 68) and mouse (67). From a functional perspective, the loss of a specific activating KlrA gene renders mice susceptible to mouse cytomegalovirus infection (MCMV) (31). This susceptibility is due to a MCMV-encoded KlrA receptor ligand (53) that can bind to inhibitory and activating receptors in MCMV-sensitive and -resistant mouse lines respectively (5). These findings have led to the hypothesis that the evolution of the KlrA locus (and possibly other paired receptor loci) could be driven by pathogens (4). In this model, pathogens evolve ligands for inhibitory receptors, and in response the host creates activating receptors through duplication, mutation, and gene conversion of inhibitory receptors (3).

In mammals, the majority of the IgSF paired receptors reside at the leukocyte receptor cluster (LRC) on human chromosome 19q13.3–13.4 and the orthologous locations on mouse chro-
mosomes 7 and 10 (34). Paired receptors are also found at other chromosomal locations. They include: the *Mair-I* and *Mair-II* in mouse and the orthologous human loci NKIR/IREM and IREM2 (71); human TREM1 (1); the signal regulatory proteins (SIRPs) (27); and the paired immunoglobulin-like receptor (PILR) genes PILRA and PILRB (18, 37), the focus of this study.

The human PILR locus at chromosome 7 contains two genes of the IgSF known as PILRA and PILRB (18, 37). PILRA is predominantly found as a transmembrane protein with a single variable (V) Ig-like extracellular domain and a cytoplasmic tail containing two ITIM motifs. PILRA was first characterized using a yeast two-hybrid system with the immune-system regulator PTPN6 (protein tyrosine phosphatase, nonreceptor type 6; formerly called SHP-1) as bait (37). PILRA was later shown to recruit PTPN11 (protein tyrosine phosphatase, nonreceptor type 11 formerly called SHP-2) to a greater extent than PTPN6 (18). Mouse models have revealed that both PTPN6 and PTPN11 are important for hematopoiesis (46, 60). In humans, PTPN6 (9) and PTPN11 (reviewed by Ref. 57) are involved in leukemia. Human PILRA is expressed in myeloid cells including monocytes, macrophages, granulocytes as well as monocyte-derived dendritic cells, but is not detected in lymphocytes (B, T or NK cells) (18). PILRA is also overexpressed in IL-10-induced myeloid dendritic cells, which are thought to play a role in counteracting T-cell activation (65). Before this study, the mRNA expression of human PILRB was thought to be the same as that of PILRA (37).

Like other paired receptors, the activating PILRB is distinguished from its inhibitory counterpart by having a truncated cytoplasmic tail and a charged amino acid in its transmembrane domain. The binding of Pilrb1 (a mouse ortholog of PILRB) to its ligand [an ortholog of human CD99 (43)] activates NK and dendritic cells (52). This mouse C99 molecule (GenBank: BAD12394 and NP_079860) also participates in trans-endothelial migration (TEM) of lymphocytes in vitro and in vivo, and can specifically recruit T-cells to inflamed skin (10). CD99 is now considered a potential drug target for treating inflammation (10). Overall, the PILR gene products are now considered novel regulators of the innate and adaptive immune systems (52).

A recent comparative genomic study of the Siglec gene cluster in five mammalian genomes revealed rapid evolution by multiple mechanisms, possibly due to an evolutionary arms race between host and pathogen (2). Gene duplication and gene conversion have been suggested to play a role in the evolution of the human PILR genes (37, 52); however, these ideas have not yet been explored by evolutionary analyses and statistically based gene conversion detection methods. Two chromosome 7 sequence assemblies (21, 49), combined with high-throughput sequence data for additional mammalian genomes, provide a reliable starting point for understanding the evolutionary mechanisms that shaped the mammalian PILR locus. Here we characterize the genomic structure and expression of human PILRB, sequence the mouse Pilr locus, and perform an evolutionary analysis of the PILR loci in the human, mouse (*Mus musculus*), chimp (*Pan troglodytes*), dog (*Canis familiaris*), opossum (*Monodelphis domestica*) and rat (*Rattus norvegicus*) genomes.

**MATERIALS AND METHODS**

Isolation and characterization of clones. BAC clone 139n8 was obtained from the CITB-CJ7-B mouse BAC library derived from the 129Sv/J strain (Research Genetics, Huntsville, AL) and physically linked to the *Aceh* locus on mouse chromosome 5 as previously described (69). BAC 493b1 was obtained from the same 129Sv/J library using a 139n8-specific probe using primers 5′-TCTCTCTCTCTCAGCTTCCAGCCTTCAG-3′ and 5′-GAGGACAGGGCATTTTGGTGTT-3′.

DNA sequencing. Genomic libraries were made for BACs 139n8 and 493b1 as previously described (69). Random clones from each sub-library were then run on ABI 373 or 377 automated DNA sequencers using fluorescently labeled primers (Amersham and ABI). An additional library for 493b1 was ligated into HincII digested pUC19 (Invitrogen), and individual plasmids were prepared and sequenced on an ABI3700 machine. All gaps, low quality regions, and regions only sequenced on one strand were filled using PCR with custom primers. An in silico digest of the assembled 493b1 sequence using BglII, BamHI, and EcoRI conformed to the actual restriction digest of the 493b1 BAC. BAC 139n8 overlapped 493b1 by 35 kb and a composite assembly with 493b1 and 139n8 sequences produced two ordered contigs that are congruent with publicly available mouse C57BL/6 genomic sequence. The 139n8 sequence was submitted as phase II draft sequence under accession number DQ055128.

Sequence analysis. The Phred/Phrap software suite was used for base-calling and contig assembly, and Consed was used for sequence finishing analysis including visual inspection, editing, and in silico digests (17, 19). Repeat elements were characterized using Repeat-Masker2 (A. F. A. Smit and P. Green unpublished; http://www. repeatmasker.org/). Further repetitive sequence analysis and graphical display of inverted repeats in the mouse genomic sequence was performed using Miropeats (44). The Miropeat results displayed here were generated using “threshold 1,000,” which shows only repeat regions that meet the criteria of having nucleotide matches minus mismatches totaling greater than 1,000. Intra- and interspecies comparisons were performed using the local alignment algorithm BLASTZ, implemented by PipMaker (50); http://bio.cse.psu.edu/). We compared finished human chromosome 7q22 sequence to paralogous regions on chromosome 7 as well as the orthologous regions in mouse (*Mus musculus*), chimp (*Pan troglodytes*), dog (*Canis familiaris*), opossum (*Monodelphis domestica*) and rat (*Rattus norvegicus*) genomes (see supporting data for a summary of the genomic sequences and their sources; the online version of this article contains supplemental material). Relevant regions were identified, and coordinates of these regions were extracted from the sequences. These discreet regions were then used for multiple alignments and further evolutionary analysis. The files used for making these comparisons, as well as the alignments and annotations, are provided in the supporting material. An interactive graphical display of these files can be viewed locally using the Laj software [http://bio.cse.psu.edu/; (69)] or through our web site (http://web.uvic.ca/~bioweb/laj.html).

Multiple alignments for the purpose of gene conversion analysis were first done using CLUSTALW followed by visual inspection and manual editing when necessary (58). We constructed phylogenetic trees using neighbor-joining analysis with both Jukes-Cantor and Kimura 2-parameter distances. Changing the distance matrix did not change conclusions from this study, and so only Jukes-Cantor distances are shown. Gaps were removed from multiple alignments before analysis. Neighbor joining trees and linearized trees were produced using MEGA3.0 (28). Gene conversion analysis was performed using GENECONV [http://www.math.wustl.edu/~saywer; (48)] as well as using the CODEDOUBLE method implemented by Drouin et al. (16). The probability of recombination events in sequence alignments was assessed using the Hidden Markov Model Method (HMM) (23) implemented by TOPALi v0.23 (36).

Expression studies, cloning, and partial cDNA's. Human semi-quantitative expression studies of PILRB mRNA were carried out on...
normalized human blood fraction multiple tissue cDNA panels (MTC; Clontech). The purity of the human blood fractions was assessed by the manufacturer to be: greater than 95% for the CD4+, CD19+, and CD8+ cells; and greater than 98% for the mononuclear cells. Semi-quantitative PCR analysis was performed using the HotStarTaq (Qiagen) enzyme with the manufacturer’s reaction buffer in a 50 μl volume. The cycling parameters were as follows: 95°C, 15 min for Taq activation, and eight cycles at 94°C, 30 s; 65°C, 1 min; 72°C, 1 min; followed by 30 to 33 cycles of 94°C, 30 s; 58°C, 1 min; 72°C, 1 min; and a final extension of 5 min at 72°C. The normalized cDNA in the human blood fractions MTC panel (Clontech) were amplified with primer pairs: JTV-101F 5'-GTACAGGTAAGGCTATCAGGCT-3', PILRB-2449F 5'-CTCCTGGACAGAATACTACAGACT-3', PILRB-3436R 5'-TCAGTGAGGTCACTCATCTCC-3', 987 bp predicted product and G3PDH control primers (Clontech) #5409-1 (983 bp product). PCR products generated using primers PVRIG-964F and PILRB-2878R, as well as PVRIG-964F and PILR-2257 (5'-GATCTGGATGAGGCGACAGGTGTC-3') were gel purified and cloned using TOPO TA Cloning kit (Invitrogen). Plasmids from 48 randomly selected colonies were sequenced in both directions as above. Sequences of 5' untranslated region (UTR) variants were deposited in GenBank under accession numbers DQ851871–DQ851885.

Northern blot. A comparison between PILRB and PILRA expression was performed using Northern blot hybridization. Probe PILRB was designed to hybridize to both PILRA and PILRB and was generated by PCR off of a human PILRB cDNA clone (IMAGE clone 1636712, accession number A1017695) using primer PILRB-2449F and PILRB-2878R. The 429-bp product was gel purified before labeling. The PILRA/B probe is 97.2% identical to the PILRA gene (12 bp substitutions over the 429 bp) that allows the probe to hybridize both PILRA and PILRB. A PILRB-specific probe was generated by PCR off of the same human PILRB cDNA clone (IMAGE clone 1636712) using primers PILRB-3193F 5'-GAGAAAGGATGTTG-3' and PILRB-3436R. The 241-bp product derived from the last exon (and 3'-UTR) of PILRB was gel purified before labeling. Both probes were randomly labeled with [α-32P]dCTP using the RediPrime random primer labeling kit (Amersham). Each probe was hybridized overnight at 65°C to commercially prepared Northern blots of human total RNA (human 12-lane multiple tissue Northern blot (MTN) # 7780-1, BD Biosciences] using the supplied ExpressHyb hybridization solution (Clontech). After hybridization, the blots were washed twice in 2× SSC, 0.1% SDS at room temperature for 15 min each, followed by two 15-min washes in 0.1 × SSC, 0.1% SDS performed at 65°C. Autoradiography was performed using Kodak Biomax film with an intensifying screen for 18 h or a phosphorimager system.

RNA blot. Human RNA blots (Clontech) containing 51 human poly-A+ RNA samples that have been normalized to the mRNA expression levels of eight different genes were hybridized with the same two probes used for the Northern blots. The PILRA/B probe was hybridized for overnight using ExpressHyb at 65°C. The blot was washed according to the manufacturer’s protocol. Imaging was done using the phosphorimager system. The RNA blot was stripped, checked for radioactivity, and then reprobed with the PILRB specific probe.

RESULTS

Comparative genomic analyses. PILRA is found next to Zcwpw1 in the human, chimp, dog, mouse, and rat genomes. All placental mammalian genomes share conserved synteny in the region spanning CYP3A4 through to ZCWPW1. The mouse genome has undergone an inversion with respect to human, dog, and chimp genome sequences. This inversion has brought the Cyp3a13 gene into the PILRA locus (Fig. 1). There is no evidence for such an inversion within the rat genome.

Orthologs of STAG3 and ZCWPW1 are assembled in a completely sequenced, gap-free, opossum contig that is orthologous to human chromosome 7q22. The Stag3/Zcwpw1 opossum sequence corresponds to the best human/ opossum chain found over the entire opossum genome (26). The region between Stag3 and Zcwpw1 is 15 kb in length, GC-poor (39.1%), and repeat-rich (72% repetitive elements). Most of the repetitive elements consist of LINEs as well as Monodelphis domestica-specific repeats (54% and 46% respectively). We could not identify candidate opossum orthologs for PVRIG or the PILR genes in this region (Fig. 2), or in other regions. While putative homologs of STAG3 and ZCWPW1 exist in the chicken genome (Gallus gallus), they do not reside together in the current assembly. Like with the opossum genome, we did not identify candidate chicken orthologs for the PILR genes.

PILRB is a mosaic transcript created through gene duplication, segmental duplication, and “exonification” of repetitive elements. The human PILRB gene (AF16108) was previously cloned and described to contain four exons; the first three of which are homologous to PILRA (37). Currently, cDNA and EST data support the existence of longer variants of PILRB that contain a number of noncoding upstream exons. This cDNA evidence resulted in the annotation of two additional isoforms of PILRB. The longest isoform (labeled variant 1 by NCBI; Fig. 3) is supported by a full-length cDNA (AK056412). Variant 1 encodes the same product as the previously described PILRB gene (now labeled as variant 3) (37). An additional isoform (variant 2) shares most exons with variant 1; however, an alternative splicing event in the coding region results in an early frameshift mutation that disrupts the PILRB open reading frame (ORF). Alignments of human cDNA and expressed sequence tag (EST) sequences to the PILRB genomic locus show that more than three PILRB isoforms exist (Fig. 3).

According to cDNA and EST information, PILRB transcripts can be as large as 3.5 kb and contain an additional 14 noncoding exons upstream of the previously described PILRB coding region. We analysed these exons and determined that they are homologous to several distinct evolutionary sources (Fig. 4A). The most dramatic event shaping the human and chimp PILRB locus is the insertion of a large segmental duplication that is ~100 kb and shares 96% identity with segmental duplications that flank the region commonly deleted in Williams-Beuren syndrome (WBS). These segmental duplications, also called duplicons, are responsible for mediating the 1.55-Mb disease-causing deletion as well as large-scale genomic inversions at 7q11 (42, 62). Two additional WBS-related duplicons are found on chromosome 7 [a 7q11.22 duplication and a 7q22 CUTL1-associated duplicon; (Fig. 2)]. While deletions of 7q22/7q11 and 7q22 have been observed (see http://chr7.org for details), no direct evidence has been reported for recombination between these duplicons. However, analysis of PILRC cDNA data from public databases suggests that the insertion of this duplication in front of PILRB has affected its structure and expression (Fig. 3).

Comparative genomic analysis show this 7q11-related segmental duplication inserted in front of PILRB (Fig. 1 and Fig. 2). Specifically, this insertion occurred between what
appears to be the remnants of a prior tandem duplication of stromalin antigen 3 (STAG3) and a novel gene we call PVRIG. We named PVRIG (poliovirus receptor-related immunoglobulin domain containing) based on the homology between its second exon and the variable Ig domain of the polio virus receptor (PVR/CD155) and polio virus receptor-like genes (PVRL1/HVEC, PVRL2/HVEB, PVRL3, and PVRL4). While we could only detect this Ig-like homology using PSI-BLAST, it was clear that the conserved regions correspond to the sequence in the PVRL/HVEC and PVRL2/HVEB gene products that bind to glycoprotein D of the herpes simplex virus (12); see supporting material for further analysis of PVRIG.

Evidence of this prior STAG3-PVRIG duplication consists of regions paralogous to: STAG3 exon 3 through exon 4; STAG3 intron 9; and a 13-kb region that extends from intron 33 through to PVRIG (Fig. 4A). The duplicated STAG3 gene has been previously referred to as STAG3L4 (45). Additional STAG3-like genes are found within the duplications that flank the WBS region. This suggests that the segmental duplication at the PILR locus is the progenitor of the segmental duplications that flank the WBS region.

Segmental duplication brings a strong, bidirectional promoter into the vicinity of PILRB. The 3’ end of the 100-kb segmental duplication contains the PMS2L1/JTV1L genes that are incorporated into PILRB variant 1 (Fig. 3). The ancestral PMS2 gene resides at 7p22 and possesses a bidirectional promoter that can drive its own expression, as well that of PILRB and PILRA variants. The presence of PMS2 in exon 1 supports the hypothesis that PILRB was created from a duplication of PILRA.

Expression patterns of human PILRB. Experiments using RTPCR with PILRA and PILRB specific primers suggested that PILRA and PILRB are co-expressed in all of the same tissues.

Fig. 1. Dot plot display of human chromosome 7q22 (x-axis) vs. the orthologous regions at chimpanzee chromosome 6, dog chromosome 6, rat chromosome 12, mouse chromosome 5 and opossum scaffold 16810 (y-axis). Diagonal line shows statistically significant local alignments between human and the other mammalian sequence. WBS, Williams-Beuren syndrome; PILR, paired immunoglobulin-like receptor.

Fig. 2. Multiple sequence comparison of the CYP3A/PILR locus at 7q22 to related segmental duplications on chromosome 7 and orthologous regions of various species. Alignment of the CYP3A/PILR locus to WBS duplincs 1, 2, and 3, 7q11.22 duplinc, 7q22 CUTL duplinc, orthologous regions from chimp, dog, mouse, rat and opossum. Alignments were performed using MultiPipMaker (http://bio.cse.psu.edu/). Aligned regions are shown in light gray and strongly aligned regions are shown in dark gray (gap-free alignments of at least 100 bp and 70% nucleotide sequence identity). RepeatMasker (http://www.repeatmasker.org/) was used to mask human repeats before the alignments were made.

In addition to JTV1 exon, the complex PILRB transcript consists of regions paralogous to: STAG3 exon 3, STAG3 exon 4, and a novel exon from within the intragenic region found 3’ of STAG3 exon 34. PILRB exons 5 and 6 are derived from LTR and SINE elements respectively. The next six exons are homologous to PVRIG. The PVRIG-like region of PILRB is followed by exon 13, which has been derived almost entirely from a LINE1 element. The last 17 bp of exon 13 is homologous to sequence found 1.5 kb 5’ of PILRA exon 1. PILRB exon 14 is homologous to the last half of PILRA exon 2 and is flanked by sequence that corresponds to the putative promoter region for PILRA (Fig. 4A).

The PILRB exons encoding the extracellular region (exons 15 and 16) are ~96% identical to PILRA exons 1 and 2. PILRB exon 17 contains the transmembrane region with the charged lysine residue and is homologous to PILRA exon 3. Finally, the coding region and majority of the 3’-UTR of the last exon are derived from a long terminal repeat (LTR) repetitive element (Fig. 4A).

The last intron of PILRB contains sequence homologous to PILRA intron 3, exon 4 and intron 4 (Fig. 4A). Both the splice acceptor and donor sites have mutated in the region paralogous to PILRB exon 4 (ag/g and g/gt have mutated to ac/g and g/ct). These mutations could explain why this exon has not been detected in any PILRB isoforms. The presence of PILRA exon 4 remnants in the intronic sequence of PILRB supports the hypothesis that PILRB was created from a duplication of PILRA.
According to cDNA and EST information in GenBank, PILRB transcripts around 3.5 kb should exist. Since the previously reported PILRA/PILRB RTPCR data were not explicitly presented, and the Northern blot results did not include data for bands greater than 2.5 kb (37), we repeated the human Northern blot and RTPCR experiments. We used the same human multiple tissue Northern (MTN) panels (Clontech) and a PILRA/B probe that was designed to cross hybridize with both PILRB and PILRA (Fig. 4B). As previously shown (18, 37), the PILRA/B probe gave a distinct band at 1.4 kb for peripheral blood leukocyte, and less intense bands for lung and spleen. However, contrary to previous results (37), strong signals were observed for bands between 3 and 5 kb for several tissues (Fig. 4B). Northern blots were also probed with a PILRB-specific 3′/H11032-UTR probe (Fig. 4C). This probe was designed from the last exon of PILRB, which shares no homology with PILRA. The same overall expression pattern for the 3–5 kb size range was observed; however, the strong 1.4-kb band for the peripheral blood leukocyte diminished substantially (Fig. 4C). The diffuse signal around 1.4 kb in the liver may represent smaller PILRB transcripts; however, this band did not show up in the PILRA/B probed blot.

While the 3′-UTR bears homology to an LTR repetitive element, it has sufficiently diverged and appears to be specific for human PILRB only. For example, the only significant alignments (e-value of less than one) found by BLAST in the entire human EST and nonredundant database come from imperfect matches to positions 130–170 of the probe. While it is possible this probe gave some nonspecific signal, the 3–5 kb band remained. This suggests that the 3–5 kb signal is not a hybridization artifact.

The same probes and experimental strategy were used to probe the RNA poly-A/H11001 blots that contain samples from 44 adult human and 7 fetal tissues. The PILRA/B and PILRB probes showed ubiquitous expression with similar signal patterns for both probes. A more intense signal was given by the PILRA/B probe for the peripheral blood leukocyte, which supports our Northern blot results (see Supplemental Materials for the RNA blot data).

The expression of PILRB was also investigated using PCR on first-strand cDNA from several blood cell fractions. We utilized primers corresponding to three regions of PILRB variant 1 (Fig. 4A). The expression pattern of the PILRB coding region matched the expression pattern obtained by primer sets amplifying the 5′ noncoding exons (Fig. 4D). While all blood fractions tested showed PILRB expression, it appears that resting CD8+, CD4+, and CD19+ cells have higher expression levels than activated cells. Primer sets specific to the PVRIG-like region of PILRB amplified multiple bands suggesting that alternative splicing events are common. At least 5 alternative GenBank transcripts are found in the PVRIG-like region of PILRB (Fig. 3). PCR products from PVRIG-964F and PILRA/B-2878R (Fig. 3) as well as PVRIG-964F and PILR-2257 primers were gel purified, cloned and sequenced. Fifteen clones that obeyed canonical splice rules when aligned to the human PILRB genomic sequence were obtained (DQ851871–DQ851885); 10 of them showed unique splicing patterns. This shows that splicing through the PVRIG-like region is highly variable and helps explain the range of transcripts observed on both the Northern blot and PCR analyses.

Our semiquantitative PCR results for the expression of PILRB in lymphoid tissues are supported by microarray experiments found in the gene atlas database at the Genomics Institute of the Novartis Research Foundation [GNF; http://symatlas.gnf.org; (54)]. Three records for PILRB were generated from the GNF1H microarray experiment (55). Closer inspection of the sequence of the reporters used for these hybridization experiments show that two of the three reporters were actually derived from PILRA (219788_at and 222218_s_at) and the third reporter (220954_at) was amplified from PILRB variant 1 (exon 16 and exon 17).
Fig. 4. Exon homology and expression analysis of human PILRB. A: homology of the 18, PILRB variant 1 exons. Exons are labeled as colored boxes and triangles based on their evolutionary relationship; the actual primer locations for mRNA expression studies are indicated with black arrows. B: human Northern blot hybridized with a PILRA/PILRB cross-hybridizing probe. C: Northern blot hybridized with a PILRB-specific probe and a β-actin-positive control. The tissues on the Northern blot in order of left to right are: brain, heart, skeletal muscle, colon, thymus, spleen, kidney, liver, small intestine, placenta, lung, and peripheral blood leukocyte. D: PCR analysis of PILRB expression using normalized first-strand cDNA from a blood tissue panel (Clontech). The locations of the primers used for amplification are shown in A. G3PDH primers were used as a template control. ORF, open reading frame; LTR, long terminal repeat.
Consistent with our data and the literature, the expression data for the PILRA probes (219788_at and 222218_s_at) reveal high expression (greater than 10-fold higher than the global median for each gene over all the tissues tested) for whole blood, peripheral blood (PB)-CD14+ monocytes, and BM-CD33+ myeloid cells. The PILRB reporter (220954_at) showed that PB-CD19+ B cells, PB-CD4+ T-cells, PB-CD56+ NK cells, and PB-CD8+ T cells were all significantly above the threefold global median. In particular, PB-CD8+ cells were expressed at 10-fold higher than the median. The PILRB reporter (220954_at) appears to be reasonably PILRB-specific as whole blood, PB-CD14+ monocytes, and BM-CD33+ myeloid cells expression were all below the threefold median mark. Together these mRNA expression experiments suggest that unlike PILRA (18), PILRB mRNA is expressed in lymphocytes. Overall, these experiments show that PILRB mRNA expression is different than PILRA expression. Characterization of the PILRB protein will be essential to determine if the altered PILRB mRNA expression is physiologically relevant.

Genomic sequencing and analysis of the mouse Pilr locus reveals a second activating Pilrb gene called Pilrb2. The mouse region orthologous to the PILR locus was identified on BAC 493b1. 493b1 was sequenced and assembled into a single contig of 194414 bp. The assembly consisted of 2,911 template reads with an average error estimation of 0.34 per 10 kb (GenBank accession number AY823670). Eight genes were annotated on 493b1: 2010007H12Rik, Bip11, Zwpl1, Pilra, Pilrb1, Pilrb2, Cyp3a13, and Gjel. In addition to these eight genes, five unprocessed pseudogenes related to the Pilr genes were found: Pilr-ps1, Pilr-ps2, Pilr-ps3, Pilr-ps4 and Pilr-ps5. The annotation of pseudogenes was manually performed by inspection of non-reciprocal local alignments obtained by comparing 493b1 BAC against itself. Regions corresponding to splice donor and acceptor sites were noted (see supporting material and AY823670). The organization and duplicated regions of the mouse PILR locus is shown in Figs. 5 and 6C.

The most striking feature of the 493b1 mouse sequence is the recent tandem duplication involving a ∼15.3-kb segment. The 15.3-kb duplicated segment, which contains Pilrb1 and Pilr-ps5, aligns with 97.2% identity to an adjacent 13.7-kb segment. The adjacent 13.7-kb segment contains a previously unreported Pilrb2 gene and Pilr-ps4. A predicted Pilrb2 transcript can be deduced by comparing Pilrb1 to the duplicated genomic sequence. Pilrb1 and the predicted Pilrb2 share 93.5% nucleotide identity and 87.5% amino acid identity. One full-length Pilrb2 transcript from a bone cDNA library exists (AK036467). This sequence differs slightly from the predicted Pilrb2 transcript through the utilization of a splice acceptor site in the region homologous to intron 3 of mouse Pilra (Fig. 5). This splice site is conserved between Pilrb1 and Pilrb2 and its utilization leaves the region homologous to exon 4 of Pilra within its 3′-UTR. Compared with our predicted Pilrb2 transcript, AK036467 translates into a product with two substitutions and one extra amino acid at the COOH terminus end. The functional analysis performed by Shiratori et al. (52) was with Pilrb1.

The transmembrane region of Pilra, Pilrb1 and Pilrb2, which ultimately distinguishes the mouse inhibitory receptors from their activating counterparts, is encoded in exon 3. Global nucleotide alignments of exon 3 show low percent identity between Pilra and Pilrb1 and Pilra and Pilrb2 (38.3% and 45.3% respectively) and high percent identity between Pilrb1 and Pilrb2 (97.4%). The 97.4% identity is similar to the
Fig. 6. Duplication analysis of the mouse PILR locus. A: i: duplication analysis using Miropeats with stringent parameters (threshold 1,000) connects relatively long regions of DNA sharing high percent identity. Colored arcs highlight the tandem duplication in the mouse Pilr locus and the black lines connect inverted repeats comprising one LINE1 element upstream of Pilra to two related LINE1 elements that flank a pseudogene (Pilr-ps6) located over 400 kb away. ii: Known genes in this region are displayed in purple (+ strand) and orange (−strand) for the coding regions, the pink and light-orange areas represent UTR regions. Arrows indicate transcript orientation and gene names are given below. B: a magnification of A, highlighting the recent tandem duplication of the mouse Pilr locus and the location of the LINE1 element. i: Green and yellow lines highlight the high similarity between the first two exons of Pilra and the first two exons of Pilrb and Pilrb2. The red lines highlight the recent tandem duplication that gave rise to two copies of Pilrb. The black lines show inverted repeats. C: transcriptional map and dot plot analysis of the mouse Pilr locus compared with itself using the BLASTZ local alignment algorithm.
percent identity seen between the 13.5-kb tandem duplication containing the Pilrb genes. Multiple sequence alignments of exon 3 from Pilra, Pilrb1, Pilrb2 and Pilr-ps1 highlight several deletion/insertions as well as substitution mutations. These mutations ultimately result in a frame shift that creates a stop codon in exon 4. The stop codon is created from a TAA sequence in exon 4, which is conserved in all three Pilr genes. Presumably, these mutations resulted in a Pilrb1/Pilrb2 ORF that lacks an ITIM sequence and has a charged lysine in the transmembrane region. The noncoding role of exon 4 in Pilrb1 and Pilrb2 supports the hypothesis that activating receptors initially arose from inhibitory receptors.

The pseudogenes Pilr-ps2, Pilr-ps4 and Pilr-ps5 all contain homology to exons 6 and 7. The ITIM domain in exon 6 is abolished in all 3 pseudogenes and none of these exons are in the correct order or orientation to considered intact remnants of any Pilr gene or Pilrb-related pseudogene (Pilr-ps1, Pilr-ps3 and Pilr-ps6).

The recent tandem duplication in mouse BAC 493b1 was collapsed in the May 2004 mouse genome assembly. To facilitate a larger scale analysis, the problematic portion of the assembly was replaced with the 493b1 sequence and the resulting 683 kb region containing the Pilr locus was compared with itself. A sixth Pilr-like unprocessed pseudogene (Pilr-ps6) was identified over 400 kb away from the Pilr locus in the inverse orientation with respect to Pilra. Pilr-ps6 resides between a Pvirg-like sequence and two fragments of Cyp3a-like pseudogenes (Fig. 6). The region between the Pilr-ps6 and Cyp3a-like pseudogenes corresponds to the break of synteny observed between the human and mouse genomic sequences. This implies that the 22-kb region between Pilr-ps6 and Cyp3a-like pseudogene is the site of an inversion breakpoint. Within this 22-kb region, a long (6.2 kb) LINE1 element is found. This LINE1 element is 93% identical to another LINE1 element found in the inverse orientation 400 kb away between Pilra and Pilrb1 and may have been responsible for mediating this inversion (Fig. 6, A and B). Furthermore, the orthologous regions in dog, chimp and rat genomes all lack CYP3A genes in the PILR region (Fig. 1).

Evolution of the mouse Pilra, Pilrb1 and Pilrb2 genes. The mouse Pilr locus underwent a recent tandem duplication giving rise to Pilrb1 and Pilrb2. To estimate the time of this duplication, we used the methods and rat/mouse divergence times previously described for dating gene duplications at the rat Ktra receptor locus (20). We computed the Kimura distances for the 13.5-kb alignment of the Pilrb1 and Pilrb2 blocks (0.31 ± 0.002) as well as the distances between rat Pilra and mouse Pilra exon 3 (0.214 ± 0.037). We chose exon 3 as there is no evidence for gene conversion within this portion of Pilra. Using the estimated divergence time of 33 million years ago (MYA) for the mouse and rat lineages (39), we estimate the duplication of Pilrb1 block ((0.31+0.214)×33 MYA) to have occurred roughly 4.8 ± 0.9 MYA.

In addition to recent duplication, inspection of the alignments between the mouse Pilr genes suggested that gene conversion played a role in making discreet regions of the anciently duplicated and paralogous Pilra and Pilrb1/Pilrb2 genes more similar than the equivalent regions of the recently duplicated Pilrb1 and Pilrb2 genes. To look for gene conversion, we focused our analysis on a 3-kb repetitive element-free region that begins upstream of each Pilr gene and extends through to the middle of intron 2. In addition to the phylogenetic analyses, two computational methods were used to look for gene conversion events. The computer program GENCONV gave significant predictions for intron 1 gene conversion events between Pilra and Pilrb1 as well as between Pilra and Pilrb2. Nonoverlapping gene conversion events between Pilra and Pilrb2 were detected in the upstream genomic region (see supporting information for more details). Significant gene conversion events were also detected using the “HMM analysis” program (23) implemented by the TOPali (36). HMM analysis can help infer mosaic structures between four sequences and it gives the probability that a given topology changes at a given position in the alignment (Fig. 7). The HMM prediction supports the gene conversion predictions made by GENCONV for the upstream genomic and intron 1 regions and suggests that recombination events also occurred in coding regions between Pilra and Pilrb1/Pilrb2. For example, it is only in intron 2 where we see the expected (recombination free) topology with orthologous rat and mouse Pilra grouping and the recently duplicated mouse Pilrb1 and Pilrb2 grouping (Fig. 7).

Extensive Pilrb gene duplications have occurred in the rat genome. The rat WGS assembly suggests that extensive gene duplication events have shaped the Pilr locus; however, the exact genomic structure remains unclear as it contains many gaps and duplications, which is indicative of a problematic assembly. High-throughput gene predictions of the rat genome, performed by NCBI using the GNOMON method (http://www.ncbi.nlm.nih.gov/genome/guide/gnomon.html), suggest that the rat genome has 18 copies of Pilrb and two copies of Pilra. Many of these predictions span large genomic regions (>1 MB) and contain many gaps. Such regions are unlikely to represent the true gene structures, but rather represent chimeric collections of exons. For example, the two identical copies of Pilra are separated by a large gap and are positioned in opposing orientation. Furthermore, overlapping regions of each rat Pilra contig are nearly identical and contain the ortholog of Zcwpw1, which is next to Pilra in all of the placental mammalian genomes investigated to date. For this study we only analyze one copy of rat Pilra (XM213737) and assume that the second copy is the result of an assembly artifact.

To obtain a better estimate of the total copy number of rat Pilr transcripts (not just the ones found in automated predictions) we searched the rat genome with the BLAT algorithm (25) using three distinct exon 3 sequences from rat Pilrb predictions and the exon 3 sequence from rat Pilra as in silico probes. We then extracted all nonoverlapping segments that spanned the appropriate size (~300 bp, including flanking genomic DNA). We initially detected the two Pilra and 44 nonoverlapping Pilrb exon 3 segments in the RGC v3.1 assembly, 40 of which spanned the entire length of exon 3. Of these, one Pilra and 27 Pilrb segments were nonidentical (Table 1, top). Of the 27 sequences, 20 contained the full open reading frame with the lysine residue in the transmembrane region. Of the remaining 7 sequences, 5 contained nonsense mutations. Interestingly, the remaining two exon 3 sequences contained a threonine (ACA) instead of the conserved lysine (AAA) residue in the transmembrane region but appeared to compensate for this by having an arginine (AGA) nine amino acids downstream in place of the conserved glycine (GGA or GGG). Phylogenetic analysis grouped these two sequences...
with the mouse Pilr-ps1 pseudogene. This suggests that this pseudogene (or gene) existed before the divergence of the mouse and rat lineages.

We also used the methodology of Hao and Nei (20) to estimate the rate of gene expansion by duplication. We produced a linearized tree and calibrated the molecular clock at 33 MYA for the divergence of mouse and rat Pilra (20). From this calibration we estimate that the expansion from a single sister Pilrb gene occurred just before the estimated time of the mouse and rat divergence (34 MYA). From this rate of gene expansion by duplication can be calculated as \((27-1)/34=0.76\) per gene per MY for the rat and \((4-1)/34=0.088\) per gene per MY for the mouse. In comparison, Hao and Nei found that the rapidly evolving nature of the PILR genes in several mammalian species, we asked whether CD99 and its related genes were evolving by similar mechanisms. Like the PILR genes, CD99 and its orthologs are highly divergent (43). In addition to CD99, the human genome contains two paralogous CD99-like genes, CD99L2 (56) and XG, as well as the pseudogene CD99L1 (43). Mouse orthologs of CD99 [Cd99 (43, 52)] and CD99L2 [Cd99L2 (56)] have been established, while XG orthologs have not (43). Using conserved synteny and sequence identity, we can clearly identify orthologs of CD99 and CD99L2 in dog, opossum and chimp genomes. As observed for the mouse, the rat genome appears to have one putative CD99 ortholog, and no XG ortholog (Table 1, bottom). At least 28 additional unique high-throughput GNOMON rat cDNA predictions are annotated as similar to the CD99-related gene Mic2 like 1. Mic2 like 1 (Mic2l1; NM_134459.1), a proposed rat ortholog of CD99L2 (56), was originally characterized as a putative single pass transmembrane protein differentially expressed in the ventral medullary surface of the rat brain (51). We found a second putative Mic2l1 rat gene on chromosome 5.
Table 1. Summary of PILR- and CD99-related gene content in 6 mammalian genomes

<table>
<thead>
<tr>
<th>Species</th>
<th>No. of Genes</th>
<th>PILR-related genes</th>
<th>CD99-related genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>2</td>
<td>PILRA and PILRB</td>
<td>CD99, XG, CD99L2 + CD99L1 pseudogene</td>
</tr>
<tr>
<td>Chimp</td>
<td>2</td>
<td>PILRA and PILRB</td>
<td>CD99, XG and CD99L2 + CD99L1 pseudogene</td>
</tr>
<tr>
<td>Dog</td>
<td>2</td>
<td>PILRA and PILRB</td>
<td>CD99, XG and CD99L2</td>
</tr>
<tr>
<td>Mouse</td>
<td>9</td>
<td>PILra, Pilrb1, and Pilrb2 + 6 pseudogenes</td>
<td>C99 and C99L2</td>
</tr>
<tr>
<td>Rat</td>
<td>∼28*</td>
<td>Pilra- + Pilrb-related sequences (∼27*)</td>
<td>C99, Mic2l1, Mic2l2 (Chr5) + Mic2l1-like (∼20*)</td>
</tr>
<tr>
<td>Opossum</td>
<td>0</td>
<td></td>
<td>CD99, XG, and CD99L2</td>
</tr>
</tbody>
</table>

PILR, paired immunoglobulin-like receptor. *Conservative estimate based on genomic draft sequence; includes nonidentical genes and pseudogenes.

that shares 80% nucleotide identity over its entire ORF. Further analysis of the 28 GNOMON predictions at the nucleotide and protein level reveals that the similarity between Mic2l1 and the Mic2l1-like transcripts is found primarily in exons 2–6. This region of Mic2l1 encodes the Cd99-related putative extracellular region (51). Two of the Mic2l1-like transcript predictions also encode a NIDO domain upstream of the Cd99-related sequence. These predicted transcripts were assembled at rat chromosomes 1, 3, and 16. As would be expected from a locus that is evolving rapidly through gene duplication, several Mic2l1-like transcripts are clustered together in the genome assembly. Suggestive of a problematic assembly, many predictions have yet to be assigned a chromosomal location; the genes that are assigned are in gap-rich regions. Some of these 28 nonidentical Mic2l1-like predictions may contain exons spanning large, gap containing distances and may represent chimeric transcript predictions or splice variants. As an independent estimate of Mic2l1-like copy number, we used the BLAT algorithm and rat Mic2l1 exons 4–5 to detect Mic2l1-like genes. Exons 4 and 5 are separated by ∼2 kb and should represent a single transcript if they are assembled in a gap-free contig. Twenty-five nonoverlapping sequences containing the entire exon 4 and exon 5 sequences that met the criteria of being spaced 1,900–2,400 bp apart with no assembly gaps were identified. Genomic DNA from these regions was extracted and aligned. Twenty-two unique sequences (defined as less than 99% identical) including Mic2l1 and its chromosome 5 paralog remained. The average pairwise percent identity was ∼95% (alignment gaps excluded). Overall, this should be a conservative estimate for the total number of Mic2l1-like genes and pseudogenes (Table 1, bottom). While more experimental work is needed to assess the copy number and verify their genomic locations, it is clear that the Mic2l1-like genes, like rat Pilrb genes, have rapidly expanded in the rat lineage by gene duplication.

Evolution of the PILR locus in other mammalian genomes.

Phylogenetic analysis of regions homologous to intron 1, exon 2 and exon 3 were performed (Fig. 8). The phylogenetic tree derived from intron 1 and exon 2 alignments show a close clustering of paralogous rodent sequences, while the exon 3 phylogeny shows the expected evolutionary relationship; with mouse and rat orthologs clustering together. While the topology of the human and chimp trees do not change, the distance between paralogs is much longer for exon 3 than it is for intron 1. This is explained by gene conversion events that homogenized the noncoding intron 1 sequences.

To test the hypothesis that gene conversion played a role in the evolution of exon 2 we used the CODouble method (16). This method is based on the assumption that in the absence of gene conversion, one would not expect a large fraction of synonymous sites in two duplicated genes to be the same within each species and different between species (8). We compared exon 2 sequences from all inter-species combinations of mouse Pilra/Pilrb1, mouse Pilra/Pilrb2, chimp PILRA/PILRB, human PILRA/PILRB, dog Pilra/Pilrb, and rat Pilra/Pilrb. Rat Pilrb exon 2 was taken from the prediction XM_222062.2. Except for the comparison of human and chimp PILR genes that contained no co-doubles, the null hypothesis of no gene conversion was rejected for all comparisons at each of the 2-fold, 3-fold and 4-fold degenerate sites ($P = 0.00000$ for each comparison). In contrast, the same analysis using exon 3 produced no evidence of gene conversion as no co-doubles were found between any of the species.

DISCUSSION

All placental mammals studied have one PILRA gene and at least one PILRB gene. In contrast, the chicken and opossum draft genome sequences do not contain any obvious orthologs for the PILR genes or the neighboring Ig-like PVRIG gene. Since opossum orthologs of STAG3 and ZCWPW1 are assembled in a gap-free contig that shares synteny with human chromosome 7q22, it is possible that the PILR locus originated after the divergence of eutherian (placental) and metatherian (marsupial) mammals. While it appears that the PILR locus is specific to placental mammals, it is also possible that the PILR and PVRIG genes were both lost in the opossum lineage.

On the basis of the conserved synteny of the PILR locus in all placental genomes characterized to date, we hypothesize that the PILR locus arose before the mammalian radiation by seeding the repeat-rich region between STAG3 and ZCWPW1 with two immunoglobulin-like domain containing genes (PVRIG and PILRA/PILRB). Evidence of an ancestral gene family of Ig-like receptors that existed before the divergence of...
mammals and birds comes from studies of the chicken Ig-like receptor (CHIR) (15, 41). Evidence of paired receptor loci that encode a V-set Ig domain with an antigen receptor-like joining (J) motif, includes the signal regulatory proteins (SIRPs) and novel immune type receptors (NITRs), which are found in mammals and bony fish respectively (64). Like the NITR genes, the PILR and PVRIG genes contain a single exon that encodes the V domain and a portion of the consensus J sequence. The existence of a classical Ig-like locus encoding V-set Ig domain is also supported by the proximity of the
Ig-like polio virus receptor (PVR) gene (19q13.31) to the LRC locus (19q13.42). We propose a basic model for the evolution of the PILR locus starting at the point where PVRIG, PILRA and PILRB are present (Fig. 9).

Like many other paired receptors, duplication has played an important role in the evolution of the PILR locus. Our analysis supports the hypothesis that activating receptors evolved from inhibitory receptors by gene duplication. Furthermore, the high similarity of the paralogous extracellular domains of activating and inhibitory receptors within each mammalian species can be explained by gene conversion, which may have occurred several times during the mammalian radiation. The high divergence of the transmembrane domain encoded by exon 3 of PILRA and the homologous region of PILRB could be explained by relaxed selection pressure, which led to mutations that gave rise to the charged amino acid in the transmembrane region. Interestingly, this charged amino acid occurs in a slightly different position in rodents and nonrodents. This raises the possibility that the original duplication of PILRA to PILRB occurred around the time of the mammalian radiation and the fixation of exon 3 mutations that gave rise to the charged amino acid occurred in separate lineages.

The PILR locus is a good example of a region where duplication events and breaks of synteny occur together (6, 7). The mouse Pilr locus is inverted with respect to other mammalian genomes. The inversion is supported by the presence of Cyp3a13 in the mouse Pilr locus and the location and orientation of the Pilr-ps6 pseudogene, which is next to a Cyp3a-like pseudogene, 400 kb away. Repetitive sequence may have played a role in the inversion of the mouse Pilr locus as long (6 kb) and highly similar (93% identical) LINE1 elements are found in inverse orientation, close to where one would predict the inversion breakpoints to be (Fig. 6). However, if this were the inversion breakpoint, it would imply that the Pilrb1/Pilrb2 progenitor gene was created after the inversion, presumably through the duplication of Pilra. This would make Pilr-ps6 the true ortholog of human PILRB. The genomic organization of the PILR locus in other mammalian genomes, the presence of Pilrb2 next to Cyp3a13, as well as the high divergence between exon 3 of Pilrb1 and Pilra, do not support the idea that Pilrb1 has been recently created from Pilra. It is more likely that the inversion occurred between the Pilrb2 and Cyp3a13 genes.

The overall repeat content of the mouse Pilr locus (106 kb, from the end of Pilra to the end of the Cyp3a13 gene) is not particularly high (41.3% compared with the genome average of 37.5% (24)). However, the content of ERV class II retroviral-like elements withLTRs makes up 9.7% of this region, which is three fold higher than the mouse genome average (24). Large blocks of LTR elements are found at the boundary of the 15-kb block, suggesting these elements may have been involved in the duplication. Transposition mediated by LTR elements is also one possible explanation for how Pilr-ps1, Pilr-ps4 and Pilr-ps3 have come to be in the inverse orientation of Pilra, Pilrb1 and Pilrb2. It has also been proposed that stretches of pyrimidine/purine diners can act as regulatory signals for gene conversion (35). Tandem GTN repeats exist 5’ and AC repeats exist immediately 3’ of the gene conversion-prone 3-kb sequence. No repetitive element insertions were found in the 3-kb region involved in gene conversion events (see Fig. 5 and supporting material). Since disruption of homology by repetitive elements would inhibit gene conversion, it is possible that insertion events have been selected against.

The independent and stochastic expression of highly similar paired receptors on the cell surface of NK cells helps the innate immune system to distinguish between self and foreign antigens. In addition to epigenetic regulation (13), differential transcription factor binding and promoter activity of the killer Ig-like receptors (KIR) has been observed, despite the high similarity of the promoter region sequences (63). Likewise, mouse Pilra and Pilrb1 genes are differentially expressed (52) yet have nearly identical putative promoter regions due to gene conversion. It will be important to see if the recently duplicated (or converted) rat Pilrb genes have evolved regulatory mechanisms similar to the KIR genes. The plasticity of the human PILRB transcript serves as a glimpse into the complex mechanisms that can shape a gene’s structure and expression. While the mosaic structure of PILRB is not predicted to yield a new gene product, it has more than tripled its size and altered its expression pattern. The creation of new gene structures can involve several molecular mechanisms including gene duplication, exon shuffling, mobile element recruitment, de novo origination of a coding region from a noncoding region, and gene fusion (32). PILRB appears to have evolved using all of the above processes. Several mammalian chimeric transcripts have been characterized (for a review see Ref. 32), and to our knowledge PILRB is the one of the most complex transcripts produced by such events. It is unclear what role nonsense-mediated mRNA decay (NMD; reviewed by Ref. 11) would play in the regulation of PILRB transcripts. In mammals, NMD operates by a “50 nucleotide rule” whereby an mRNA is targeted for degradation if a termination codon is found more than about 50 nucleotides upstream of the final exon (38). Furthermore, the regulation of unproductive splicing variants by NMD has been shown to regulate protein translation in several human genes (regulation of unproductive splicing and translation (RUST); reviewed by Ref. 22). Human PILRB transcripts contain sequence homologous to all PVRIG exons (Fig. 4A). Although the PILRB alternative splice variants characterized in this study contain the region homologous to the PVRIG initiation codon, the PVRIG-like sequence is unlikely to be translated into a mature protein as it contains two stop codons within its predicted ORF. It remains to be seen whether or not PILRB is affected by NMD and if the abundant alternative splicing events observed in the PVRIG-like exons can regulate PILRB transcript stability and translation.

Contrary to previous results, the mRNA transcription pattern differs from that of PILRA (37), and extends to cells of the lymphoid lineage. The GNF microarray results support our data and also show relatively high expression of PILRB mRNA in NK cells (55). The difference between PILRA and PILRB mRNA expression suggests that the anti-PILRA 36H2 monoclonal antibody, which does not recognize lymphocytes, is specific to PILRA (18); however, this has not been confirmed. Antibodies raised against PILRB will be important to resolve this question. It is also important to note that in the mouse, Tryobp is necessary for high levels of Pilrb1 signal transduction and expression on the surface of NK cells and BM-DC cells (52). Thus it is possible that 1) human tissues that express PILRB mRNA may not express the protein and 2) even
Fig. 9. A summary and proposed model for the evolution of the mammalian PILR locus. 

1) Insertion of Ig-like genes into ancestral mammalian STAG3/ZCWPW1 locus. The duplication of PILRA to form PILRB would have occurred around the time of the insertion. 2: PILRB mutations truncated the cytoplasmic domain and produced a lysine residue in the transmembrane region. The genomic organization at this stage is characteristic of the dog genome. 3: Tandem duplication of STAG3/PVIRG and separate duplication of PILRB. 4: A duplicon bearing PM2L1/JTV1 inserted in front of PILRB. Both human and chimp share this organization. We cannot say if the duplication of STAG3/PVIRG occurred before, or resulted from the insertion of the duplicon. This duplication resulted in a chimeric transcript. B: A rodent-specific summary. 1: Pilra, Pilrb and a number of Pilrb genes or pseudogenes (not shown) existed downstream of Stag3 before the divergence of the mouse and rat. 2: After the divergence of mouse and rat, an inversion occurred between the Pilr and Cyp3a loci bringing Cyp3a13 into the PILR locus and leaving a Pilrb pseudogene and Pvirg in what was once the Cyp3a locus of the mouse genome. A recent duplication (~5 million years ago) of Pilrb gave rise to Pilrb1 and Pilrb2.
if protein is produced, it may not be significantly expressed or be able to transduce signals without the co-expression of TYROBP.

The rat genome assembly contains more than twice the duplication content of the mouse genome and most of the duplications are tightly clustered intrachromosomally (14, 61). The paired immunoglobulin-like receptor locus is found within the eleventh largest block of segmental duplication detected in the rat genome assembly (61). Many of the recently duplicated segments specific to the rat lineage are related to reproduction, immune system and toxin metabolism proteins (47). Since the rat genome sequence was obtained from two female rats from a highly inbred line (BN/SsNHsd), it is likely that the Pilr genes detected represent one haplotype. Further resolution of the rat PILR locus is needed to determine the ratio of genes to pseudogenes, our combined estimate of 27 Pilr genes is reasonable as the gene segments 1) do not include identical sequences and 2) are not implicated in gene conversion events.

If we were to exclude predictions that have two or fewer nucleotide differences (more than 98.7% identical) we would be left with 22 unique exon 3 sequences. Sixteen of these would maintain the ORF and contain the conserved lysine residue in the transmembrane domain.

The duplication of Pilrb produced new activating receptors in mice and rats (Table 1, top). Our prediction of 27 distinct exon 3 sequences was used to obtain the estimate of 0.76 per gene per MY for the rate of gene expansion in the rat Pilrb gene. This result suggests that the rat Pilrb gene expansion is one of the highest reported for any mammalian gene family and is second only to the Morpheus gene family (39), which was calculated to have a rate of 1.00 per gene per MY (20). In contrast, this rate is much higher than the estimated euchromatic gene duplication rate of 0.001–0.03 gene per MY (33).

Reminiscent of the rat Pilrb gene duplications, a large expansion of Cd99 related genes occurred in the rat, but not the mouse genome. These duplications appear to have involved a divergent paralog of Cd99, Mic211 (Cd99i12). Over 20 copies of Mic211-like genes are predicted in the rat genome. Without functional evidence of their interaction it is premature to speculate about the possible co-evolution of Pilrb/Mic211-like genes. Nevertheless, these extensive gene duplications and their undefined roles in rat physiology make them intriguing subjects for future study.

The rapid evolution observed for the mouse and rat Pilr loci may be indicative of paired receptor genes that are responding to evolutionary pressure provided by pathogens. Based on the biology of the Klra genes, Arase and Lanier (4) proposed that activating receptors evolved from inhibitory receptors under selective pressure imposed by viruses and other pathogens. It has been noted that mouse Cd99, a ligand of Pilrb1 (AB122023), which shares 40% identity to human CD99, is 20% identical to the PE-Pro-Glu polymorphic sequence of Mycobacterium tuberculosis (52). We also note that CD99 shares 37 and 30% identity with two portions of the repetitive region that makes up the cytoplasmic tail of the LMP1 protein of the herpes virus. Details of the ancestral PILR duplication in mammals fit with the proposed mechanism of pathogen-driven evolution; however, the binding of pathogen ligands to PILR gene products has not been substantiated.

This study revealed that the PILR locus is dynamically evolving by means of gene duplication, insertion, mutation and conversion in several mammalian genomes. Due to the high similarity of the extracellular domains of Pilra, Pilrb1 and Pilrb2, it is likely that all three can bind the mouse Cd99 ligand. Unlike more complicated paired receptor loci, the mouse Pilr locus is well suited for a molecular dissection using mouse models. Finally, the intriguing expansions of the Pilr and Cd99 related genes are relatively rare events in any mammalian genome, making them important areas for further evolutionary and functional analyses.

ACKNOWLEDGMENTS

We thank Linda McKinnell, Aura Danby, Peter Hunt, Pamela Schnupf, Marianne Beetz-Sargent, Kris von Schalburg, and Roberto Alberto for assistance with DNA sequencing. We also thank Robert Burke, Sally Goldes, Chris Upton, and Nancy Sherwood for critical readings of the manuscript draft.

Present address of J. Cheung: College of Medicine, The University of Vermont, Burlington, VT 05405.

GRANTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research (B. F. Koop). M. D. Wilson was also supported by a Michael Smith Foundation for Health Research Trainee Award and a Dr. Julius Schleicher Graduate Scholarship.

REFERENCES


