



Published in final edited form as:

Immunogenetics. 2005 April ; 57(1-2): 151–157.

Origin and evolution of the Ig-like domains present in mammalian leukocyte receptors: insights from chicken, frog, and fish homologues

Nikolas Nikolaidis, Jan Klein, and Masatoshi Nei

Department of Biology, Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA, 16802, USA e-mail: nxn7@psu.edu Tel.: +1-814-865-1034 Fax: +1-814-863-7336

Abstract

In mammals many natural killer (NK) cell receptors, encoded by the leukocyte receptor complex (LRC), regulate the cytotoxic activity of NK cells and provide protection against virus-infected and tumor cells. To investigate the origin of the Ig-like domains encoded by the LRC genes, a subset of C2-type Ig-like domain sequences was compiled from mammals, birds, amphibians, and fish. Phylogenetic analysis of these sequences generated seven monophyletic groups in mammals (MI, MII, and FcI, FcIIa, FcIIb, FcIII, FcIV), two in chicken (CI, CII), four in frog (FI–FIV), and five in zebrafish (ZI–ZV). The analysis of the major groups supported the following order of divergence: ZI [or a common ancestor of ZI and F (a cluster composed of the FcIII and FcIV groups)], F, CII (or a common ancestor of CII and MII), MII, and MI–CI. The relationships of the remaining groups were unclear, since the phylogenetic positions of these groups were not supported by high bootstrap values. Two main conclusions can be drawn from this analysis. First, the two groups of mammalian LRC sequences must diverged before the separation of the avian and mammalian lineages. Second, the mammalian LRC sequences are most closely related to the Fc receptor sequences and these two groups diverged before the separation of birds and mammals.

Keywords

Leukocyte receptor complex; Chicken Ig-like receptors; Fc receptors; Ig-like domain groups

The mammalian natural killer (NK) cell receptors fall into two categories, one category belonging to the immunoglobulin superfamily (IgSF) and the other to the C-type lectin superfamily. The Ig-like receptors occupy a genomic region called the leukocyte receptor complex (LRC; Trowsdale et al. 2001). In mammals, the LRC contains several gene families including the killer cell Ig-like receptors (KIR), the leukocyte Ig-like receptors (LILR), and the paired Ig-like receptors (PIR), which form species-specific evolutionary clusters (Martin et al. 2002). Singleton genes have been identified in the LRC of humans, artiodactyls, and rodents (Hoelsbrekken et al. 2003; Maruoka et al. 2004; Morton et al. 2004). The presence of the LRC in all mammals so far studied suggests that this region formed before the mammalian radiation. Evolution of the LRC genes has thus far been studied mainly in mammalian species for which gene sequences have been available (Volz et al. 2001; Hughes 2002; Martin et al. 2002). Here we investigate the origin of Ig-like domains encoded by the mammalian LRC genes using recently generated homologous sequences from chicken, frog, and fish.

The annotated human and mouse LRC genomic sequences [from the National Center for Biotechnology Information (NCBI) builds 34 and 32, respectively] and the chicken Ig-like receptors (CHIR)-A and CHIR-B (Dennis et al. 2000) were used in tBLASTn (Basic Local Alignment Search Tool) searches (Altschul et al. 1990). The databases searched were as follows: chicken (*Gallus gallus*), the whole genome shotgun (WGS as of 04/04) database of NCBI and the ENSEMBL database (v22.1.1) (Washington University Genome Sequencing Center); frog (*Xenopus tropicalis*), the preliminary assembly (v1.0) of the genome (Department of Energy Joint Genome Institute); zebrafish (*Danio rerio*), the WGS (05/04) database of NCBI (Sanger Genomic Institute). BLAST searching was conducted by using high expected *E*-values ($E=10$) to ensure that most sequences homologous to the mammalian LRC genes would be retrieved. The resulting BLAST hits were sorted according to the *E*-value and the coverage of the query sequence. Only hits covering at least 85 amino acid residues (the minimum length of the C2-type Ig-like domain; Klein and Hořejší 1997) were used for further analysis. Gene structures were predicted by using GenomeScan (Yeh et al. 2001). In the case of chicken, several genes annotated in the ENSEMBL database, as well as additional sequences identified by us were used. Because of the incomplete status of the chicken and the frog genome projects and the fact that most of the genes identified reside in small non-overlapping contigs, the precise gene structure and the genomic organization of the chicken and frog genes are not known.

Ig-like domains were identified by using the SMART (Letunic et al. 2004) and PFAM databases (Bateman et al. 2004). To determine whether the sequences were of the C2-type, characteristic of the LRC genes (Martin et al. 2002), they were subjected to structure-based alignment analysis using C2-type domains with resolved tertiary structure as profiles. The Ig-like domain sequences were aligned using the profile alignment option of ClustalX v.1.81 (Thompson et al. 1997). Nucleotide alignments were obtained by correspondence to the amino acid alignments. Phylogenetic trees were constructed by using the neighbor-joining (NJ) method with the *p*-distance (proportion of differences; MEGA v2.1; Kumar et al. 2001). The *p*-distance is known to give generally a higher resolution of branching patterns because of the smaller standard errors (Nei and Kumar 2000). We also constructed parsimony trees, but since they were essentially the same as the NJ trees with respect to the major branching patterns, they are not shown.

The extent of divergence of amino acid sequences was found to be higher than that of the nucleotide sequences in most pairwise comparisons (Suppl. Table 1). For example, in the sequences of the CI group the average amino acid identity was 57%, whereas the average nucleotide identity was 75%. One possible explanation of this difference could be positive selection acting on these sequences (as has been proposed for the KIR genes; Hughes 2002; Hao and Nei 2004). To test this hypothesis, we calculated the number of non-synonymous (p_N) and synonymous (p_S) substitutions per site using the original version of the Nei-Gojobori method (as implemented in MEGA2; Kumar et al. 2001). The analysis showed that in most comparisons the p_N/p_S ratio was less than 1 (mean value being 0.8505) and the Z-test suggested that the null hypothesis of neutrality could not be rejected at the 5% significance level (Nei and Kumar 2000). Thus, even if positive selection were restricted to specific sites of these sequences [as has been shown for the peptide binding sites of major histocompatibility complex (Mhc) class I molecules; Hughes and Nei 1988], it would not explain the overall high divergence of the amino acid sequences. Although the phylogenetic trees based on amino acid and nucleotide sequences generated topologies with similar branching patterns with respect to the major clades, the bootstrap values of the former trees were lower than those of the latter because of the high degree of protein sequence divergence. For this reason, only the trees based on nucleotide sequences are presented here.

Nearly 200 C2-type Ig-like domain sequences have been identified from mammals, chicken, frog, and zebrafish. Preliminary phylogenetic analysis of these sequences generated 18

monophyletic groups with two or more sequences (Suppl. Fig. 1). The mammalian sequences (human and mouse) could be divided into seven groups, two of which consisted of LRC sequences (groups MI and MII) and five were composed of domains belonging to the receptors for the Fc region of immunoglobulins (Fc receptors; FcI, FcIIa, FcIIb, FcIII, FcIV). Two groups (CI and CII) consisted only of chicken sequences, four groups were frog-specific (FI–FIV) and five groups were zebrafish-specific (ZI–ZV). Since the branching patterns of the different groups were not supported by high bootstrap values in this analysis, apparently because of the large number of sequences and the small number of nucleotide sites used (Nei and Kumar 2000), in the subsequent examination a different analysis was conducted.

To resolve the relationships among the individual groups, profile alignments and phylogenetic analyses were performed in two steps. In the first step, representative sequences from each group were used. In the second step, the relationships among the most closely related groups were resolved, and then the more divergent groups were progressively added one by one. These analyses suggested that the mammalian MI and MII groups formed a cluster with the chicken CI and CII groups (the CM cluster in Fig. 1), which was supported by high bootstrap values. The frog groups (FI–FIV) were closely related to the mammalian Fc receptors. In particular, groups FcIII and FIII formed a cluster supported by high bootstrap values (cluster F in Fig. 1), which assumed an outgroup position to the avian-mammalian cluster (CM). However, it is not clear whether the entire set of frog (FI–IV) and Fc receptor groups (FcI–IV) form a single monophyletic cluster, since the grouping of FI and FII with F and of FIV with FcIV is supported by low bootstrap values (Fig. 2). Moreover, the largest group in zebrafish (group ZI in Fig. 1), which contains almost 70% of the zebrafish Ig-like sequences (Suppl. Fig. 2), assumes an outgroup position to both the F and the CM clusters. The phylogenetic position of the remaining zebrafish groups (ZII–ZV) is unclear, since none of the outgroup positions of these groups relative to the F and CM clusters is supported by high bootstrap values.

Our results indicate that the Ig-like domains belonging to the mammalian Fc receptors and the amphibian sequences are closely related to the avian CHIR (CI and CII) and the mammalian LRC (MI and MII) sequences (Fig. 1 and Suppl. Fig. 3). Dennis et al. (2000) have originally suggested that Fc and LRC sequences share a common ancestor. These authors have pointed out that, despite the low degree of sequence similarity, the tertiary structures of the Ig-like domains encoded by the *Fc γ RIIB* (D2 of group *FcII* in Fig. 2) and the *KIR2DL1* (D2 of group MII in Fig. 1) genes are similar. An additional link between LRCs and FcRs could be that the *FcaR* (CD89) gene resides in the LRC of all mammals so far studied (Morton et al. 2004). The Ig-like domains of *FcaR* belong to the MI and MII groups of domains.

The mammalian MI and the chicken CI groups are clustered with high bootstrap support (Fig. 3), suggesting that these two groups share a common ancestor, which existed before the separation of birds and mammals. Although the clustering of CII and MII groups is reasonably well supported (Fig. 3), it is not well supported when outgroup sequences are used (Fig. 1). Thus, it is not clear whether the clustering of the CII and MII groups is significant. Three evolutionary scenarios can explain the observed topologies (Fig. 4). In the first scenario, it is assumed that MII and CII are clustered together and it is inferred that two different groups of domains (I and II) were present in the common ancestor of birds and mammals (Fig. 4a). In the second and the third scenarios (Fig. 4b, c) it is assumed that MII and CII are not clustered together. These two scenarios infer that the common ancestor of birds and mammals had at least three different groups of domains [the common ancestor of MI–CI (I), the ancestor of MII, and the ancestor of CII]. All three hypotheses infer that the common ancestor of birds and mammals had at least two different groups of domains. An alternative hypothesis, according to which the common ancestor of birds and mammals had only one group of domains, is not supported by our data (Fig. 4d). Thus, regardless of which of the three scenarios is true we

propose that at least two groups of Ig-like domains existed before the divergence of avian and mammalian lineages.

We suggest a speculative evolutionary scenario for the evolution of the Ig-like domains that have been identified (Suppl Fig. 4). According to this scenario the ZI group shared a common ancestor with the ancestor of the F and CM clusters (F-CM) that existed before the divergence of the fish and tetrapod lineages (Z-F-CM). The divergence of the F and CM clusters from their common ancestor probably occurred after the fish-tetrapod split but before the bird-mammalian split.

Analysis of how the Ig-like domains of single mammalian, avian, amphibian, and fish proteins are distributed among the phylogenetic groups has revealed a number of interesting associations. First, the mammalian group I (MI) contains the first domains (D1 in Fig. 3b) of all the LRC-encoded proteins, except for the KIRs, the third domains of PIRs (D3), and the two domains of the osteoclast-associated receptors (OSCAR). MII contains all three domains of the KIRs, as well as the remaining LRC-encoded domains (Fig. 3 and Suppl. Fig. 5). This result indicates that within each of these groups, there are Ig-like domains belonging to proteins with divergent functions, which have a recent common ancestor (see also Martin et al. 2002). Second, the chicken CI group contains the D1 domains, and the CII group contains the D2 domains of the CHIR proteins (Fig. 3). Third, the mammalian FcI group contains the D1 domains (see Fig. 2b), FcII contains the D2, FcIII the D3, and FcIV the remaining domains (Fig. 2). This result indicates that within each group, there are domains belonging to proteins with distinct functions [the FcG and the IRTA (FcH); Davis et al. 2001; Miller et al. 2002], which share a common ancestor. Fourth, the frog group I (FI) contains mainly the D1 domains, FII the D2, and FIII the remaining domains, while FIV contains two domains encoded by a single gene (Fig. 2). Fifth, the zebrafish group I (ZI) can be divided into two subgroups (ZIA and ZIB in Suppl. Fig. 2). The ZIA subgroup contains the D1 and D2 domains, and the ZIB contains the D3 and D4 domains. The remaining domains form groups ZII–ZV (Suppl. Fig. 2). Although the clustering of the zebrafish groups is supported by low bootstrap values, their origin from a common ancestor is supported by the fact that the genes encoding these domains are located in tandem in a single genomic region (data not shown).

In conclusion, our results show that the Ig-like domains of the mammalian LRC and Fc receptors are present also in other non-mammalian vertebrates. According to our analyses, the two groups of LRC Ig-like domain sequences probably diverged before the separation of the mammalian and avian lineages. In addition, our data suggest that the LRC sequences are most closely related to Fc receptor sequences and that the divergence between these two groups occurred before the separation of birds and mammals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The Washington University Genome Sequencing Center, the DoE Joint Genome Institute and the Sanger Genomic Institute are acknowledged for making the chicken, frog and zebrafish sequences available. We thank the three anonymous reviewers for their comments and suggestions. We also thank Wojciech Makalowski and Dimitra Chalkia for valuable comments and discussions. This work was supported by a grant from the National Institutes of Health (GM20293) to M.N.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410. [PubMed: 2231712]

- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. *Nucleic Acids Res* 2004;32:138–141.
- Davis RS, Wang Y-H, Kubagawa H, Cooper MD. Identification of a family of Fc receptor homologs with preferential B cell expression. *Proc Natl Acad Sci USA* 2001;98:9772–9777. [PubMed: 11493702]
- Dennis G Jr, Kubagawa H, Cooper MD. Paired Ig-like receptor homologs in birds and mammals share a common ancestor with mammalian Fc receptors. *Proc Natl Acad Sci USA* 2000;97:13245–13250. [PubMed: 11078516]
- Hao L, Nei M. Rapid expansion of killer cell immunoglobulin-like receptor genes in primates and their coevolution with MHC class I genes. *Gene*. 2004in press
- Hoelsbrekken SE, Nylenna O, Saether PC, Slettedal IO, Ryan JC, Fossum S, Dissen E. Cutting edge: molecular cloning of a killer cell Ig-like receptor in the mouse and rat. *J Immunol* 2003;170:2259–2263. [PubMed: 12594244]
- Hughes AL. Evolution of the human killer cell inhibitory receptor family. *Mol Phylogenet Evol* 2002;25:330–340. [PubMed: 12414314]
- Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals over-dominant selection. *Nature* 1988;355:167–170. [PubMed: 3412472]
- Klein, J.; Hořejší, V. *Immunology*. 2nd edn.. Oxford Blackwell Science; Oxford: 1997.
- Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 2001;17:1244–1245. [PubMed: 11751241]
- Lanier LM. Natural killer cell receptor signaling. *Curr Opin Immunol* 2003;15:308–314. [PubMed: 12787756]
- Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 2004;32:142–144.
- Martin AM, Kulski JK, Witt C, Pontarotti P, Christiansen FT. Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol* 2002;23:81–88. [PubMed: 11929131]
- Maruoka T, Nagata T, Kasahara M. Identification of the rat IgA Fc receptor encoded in the leukocyte receptor complex. *Immunogenetics* 2004;55:712–716. [PubMed: 14647992]
- Miller I, Hatzivassiliou G, Cattoretti G, Mendelsohn C, Dalla-Favera R. IRTAs: a new family of immunoglobulin-like receptors differentially expressed in B cells. *Blood* 2002;99:2662–2669. [PubMed: 11929751]
- Morton HC, Pleass RJ, Storset AK, Dissen E, Williams JL, Brandtzaeg P, Woof JM. Cloning and characterization of an immunoglobulin A Fc receptor from cattle. *Immunology* 2004;111:201–211.
- Nei, M.; Kumar, S. *Molecular evolution and phylogenetics*. Oxford University Press; New York: 2000.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
- Trowsdale J, Barten R, Haude A, Stewart CA, Beck S, Wilson MJ. The genomic context of natural killer receptor extended gene families. *Immunol Rev* 2001;181:20–38. [PubMed: 11513141]
- Volz A, Wende H, Laun K, Ziegler A. Genesis of the ILT/LIR/MIR clusters within the human leukocyte receptor complex. *Immunol Rev* 2001;181:39–51. [PubMed: 11513150]
- Yeh R-F, Lim LP, Burge CB. Computational inference of homologous gene structures in the human genome. *Genome Res* 2001;11:803–816. [PubMed: 11337476]

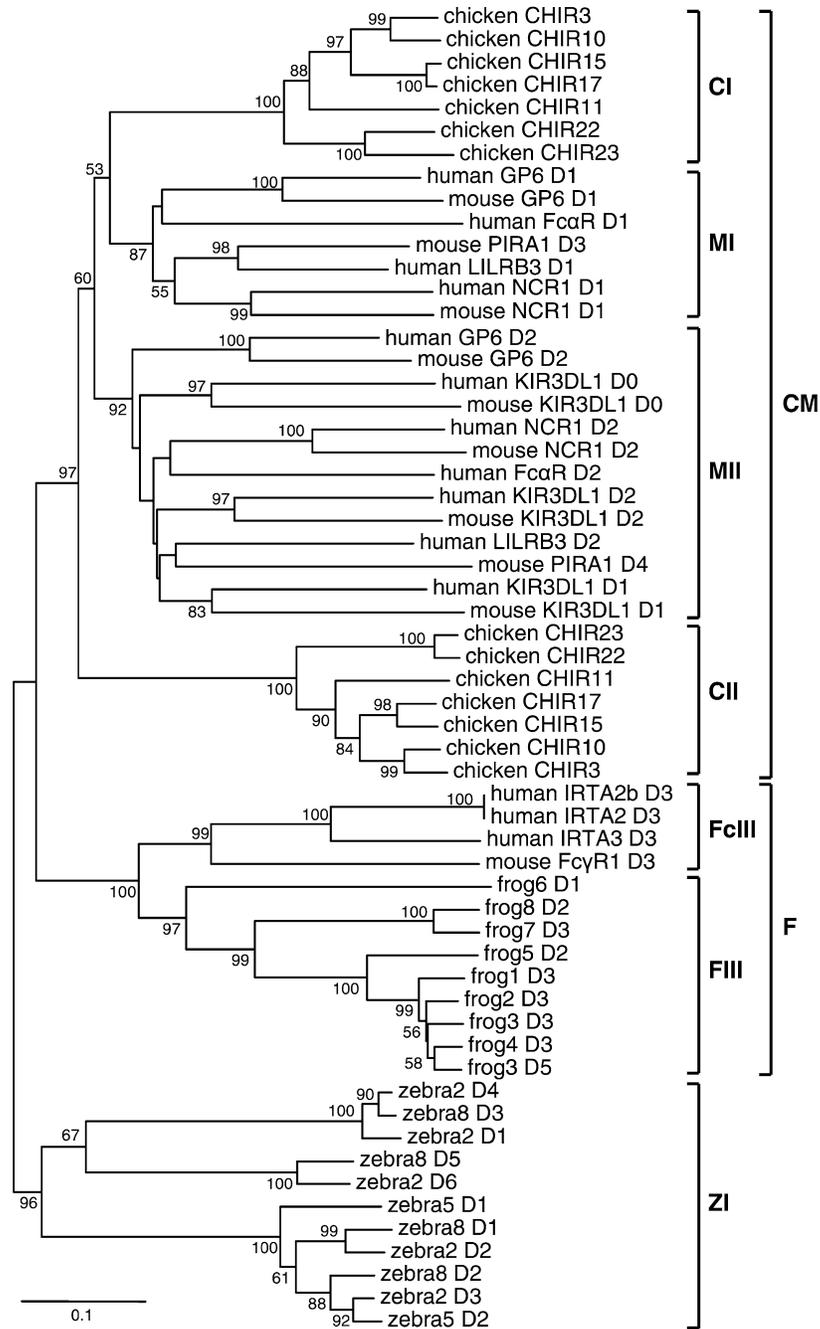


Fig. 1. Phylogenetic tree of C2-type Ig domain sequences from five vertebrate species. Six monophyletic groups (CI, CII, MI, MII, F, and ZI) supported by relatively high bootstrap values have been identified. The tree was constructed by the NJ method with p -distances for 204 nucleotide sites after elimination of alignment gaps (complete deletion option; Nei and Kumar 2000). The numbers for interior branches represent bootstrap values (only values higher than 50 are shown). The accession numbers for the sequences used are given in Figs. 2 and 3 and in the Supplementary Figures

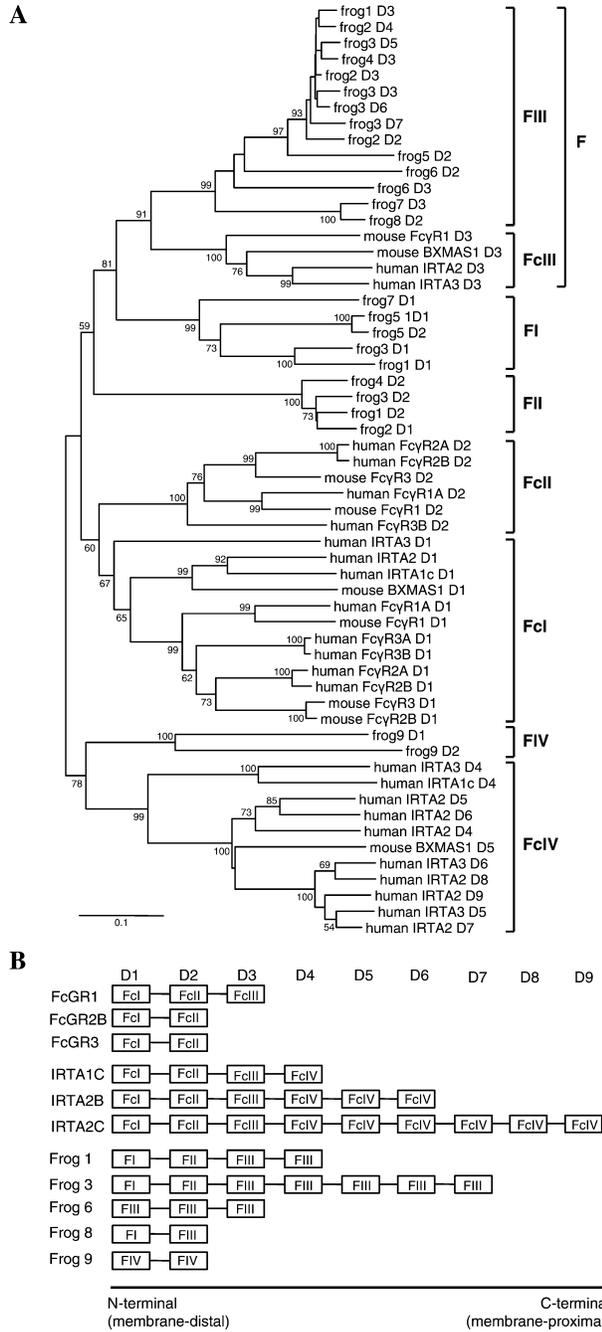


Fig. 2.
a NJ tree showing the Ig-like domain lineages (FcI–IV) of the Fc receptors from human and mouse and the Ig domain lineages of frog (FI–IV). The tree was constructed by using *p*-distances for 146 nucleotide sites after elimination of alignment gaps. The accession numbers for human sequences are: *FcγR1A* NM_000566; *FcγR2A* NM_021642; *FcγR2B* NM_004001; *FcγR3A* NM_000569; *FcγR3B* NM_000570; *IRTA1c* NM_031282; *IRTA2* NM_031282; *IRTA3* AF459027. The accession numbers for mouse sequences are *FcγR1* NM_010186; *FcγR2B* NM_010187; *FcγR3* NM_010188; *BXMAS1* AY158090. For the frog sequences the numbers of the genomic scaffolds on which the gene resides are as follows: frogs 1, scaffold 18812; 2, scaffold 33870; 3, scaffold 7429; 4, scaffold 28895; 5, scaffold 2149; 6, scaffold

15471; 7, scaffold 3806; 8, scaffold 54902; 9, scaffold 64562. **b** Domain (D1–D9) organization of representative Fc receptor and frog molecules. Only the Ig-like domains (*open boxes*) are shown. The phylogenetic group of each domain is given inside the box

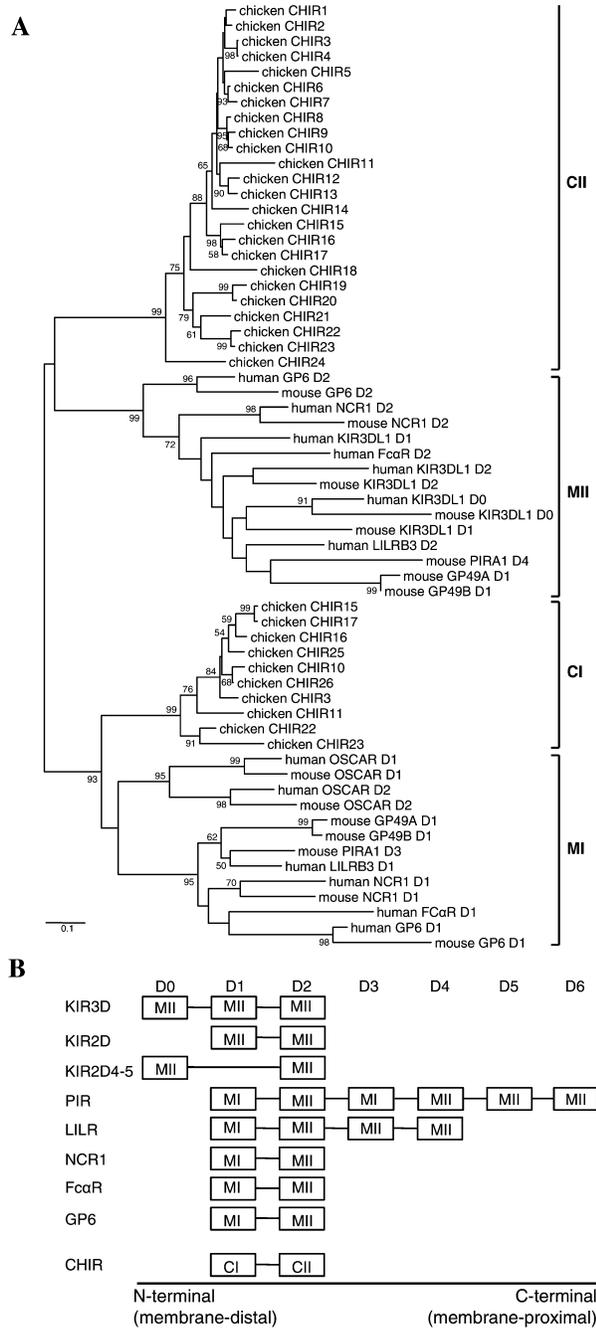


Fig. 3.
a NJ tree for the mammalian MI and MII and the chicken CI and CII Ig-like domain groups. The tree was constructed with *p*-distances for 182 nucleotide sites. The numbers on the interior branches represent bootstrap values (only values higher than 50 are shown). The accession numbers for the LRC sequences are given in Suppl. Fig. 4. Only representative PIR and LILR domain sequences were used, because the six and four domains of PIR and LILR, respectively, belong to MI and MII two groups (see **b** and text for details as well as Hughes 2002; Martin et al. 2002). The nomenclature of Dennis et al. (2000) was used for the chicken genes, but because most genes are incomplete it is not known whether they correspond to activation (–A) or to inhibitory (–B) forms. For this reason the genes were denoted as CHIR followed only by

a number, which does not reflect their genomic location. The accession numbers for the CHIR sequences and the contig number, on which a gene resides, are: CHIR 1, ENSGALT000000325324; 2, contig 8282; 3, contig 5460; 4, ENSGALT000000020530; 5, ENSGALT00000021186; 6, ENSGALT00000005320; 7, contig 6244; 8, contig 25757; 9, ENSGALT00000021683; 10, ENSGALT00000013108; 11, contig 13322; 12, contig 37304; 13, ENSGALT00000004383; 14, contig 38577; 15, ENSGALT00000001991; 16, contig 19145; 17, ENSGALT00000000208; 18, contig 6407; 19, ENSGALT00000021995; 20, ENSGALT00000022411; 21, ENSGALT00000022721; 22, ENSGALT00000010294; 23, ENSGALT00000002835; 24, contig 6240; 25, contig 26710; 26, contig 14147. **b** Domain organization of representative LRC and CHIR sequences. Only the Ig-like domains (*open boxes*) are shown. The phylogenetic group of each domain is given inside the box

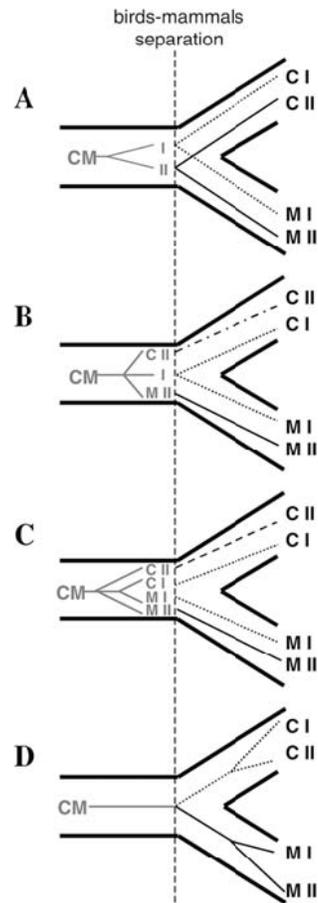


Fig. 4. Four (**a–d**) alternative evolutionary scenarios that could explain the relationships among the Ig-like domain groups of the mammalian LRC and the chicken CHIR genes. Hypothetical ancestral and extant sequences are indicated by *gray* and *black* fonts, respectively