Genomic Organization of the Chicken CD8 Locus Reveals a Novel Family of Immunoreceptor Genes[1]

Hung-Jiun Liaw,* Wei-Ru Chen,* Yu-Chia Huang,* Chung-Wei Tsai,* King-Chan Chang,† and Chia-Lam Kuo[2]*

The genomic organization of the chicken CD8α gene was investigated to determine the basis of its polymorphism. Contiguous to the CD8α gene we identified multiple DNA blocks possessing sequences homologous to CD8α. Gene conversions and recombination over evolutionary time among CD8α and these CD8α homologous genes seem to account for the observed polymorphism. Furthermore, these CD8α-like DNAs encode a distinct multigene family of immunoreceptors that have a charged or polar residue in place of the interspecies-conserved CD8α transmembrane proline residue and a short cytoplasmic tail nonhomologous to CD8α. The identification of this novel multigene family with an organization reminiscent of human killer Ig-like receptors raises compelling questions on their evolutionary relationship among immunoreceptors. ***The Journal of Immunology,*** **2007, 178: 3023–3030.**

The membrane-bound glycoprotein CD8 consists of CD8$\alpha\alpha$ homodimers or CD8$\alpha\beta$ heterodimers and is expressed primarily on thymocytes and CTLs (1). Both the CD8 α- and β-chains are composed of a single extracellular Ig superfamily (IgSF)[3] variable (V) domain, a membrane-proximal hinge (H) region, a transmembrane domain (TM), and a cytoplasmic tail (CY). Biochemical and structural studies have shown that the Ig V domain of CD8α, in conjunction with that of CD8β, interacts with MHC class I molecules, thereby allowing the CD8 molecule to function as a coreceptor of the T cell Ag receptor (2–7).

The α1 and α2 domains of MHC class I molecules are responsible for binding and presenting antigenic peptides to TCR. They are polymorphic and, therefore, determine the nature of the peptide repertoire presented to T cells in the vertebrate immune system. In contrast, the Ig-like α3 domain of MHC class I molecules is relatively nonpolymorphic, especially the CD loop that connects strands C and D and constitutes part of the site contacted by the CDR loops of the CD8α Ig V domain. The interaction between CD8α and MHC class I molecules appeared conserved within a given species and this is reflected by the conserved nature of the CD8α Ig V domain in humans and mice. In marked contrast, the chicken CD8α Ig V domain was reported to be polymorphic in its CDR1 and CDR2 loops (8). This raises questions on the reason for this unusual polymorphism and on the mechanisms that generate it. To determine whether there is a direct relationship between the genomic organization and the polymorphic nature of the chicken CD8α gene, we investigated its genomic structure and in that process discovered a novel multigene family of CD8α-like immunoreceptors.

Materials and Methods

Isolation and assembly of contiguous cosmids

Three semi-nested primer sets, CA1-ACA1 followed by CA2-ACA2 and CA2-ACA3 (CA1, 5′-ATGGCCAGGTCTCCTGCACTG-3′; CA2, 5′-CAGGAGGGACAGCGGCTGGAG-3′; ACA1, 5′-CACACAGGGTTTTGCACCAGGC-3′; ACA2, 5′-CCAGATGAAAATCTCACAGTA-3′; and ACA3, 5′-TCCCGAGCTGAAGTACAGCAT-3′), based on the CD8α cDNA sequence (GenBank accession no. Z22726), were used in PCR using a chicken spleen cDNA library (Clontech Laboratories) as template. The final PCR product primed with CA2-ACA3 were labeled by random priming with [^{32}P]dCTP and used as a probe for the colony hybridization screening of a leghorn liver cosmid library (Clontech Laboratories) and the Red Jungle Fowl (RJF) bacterial artificial chromosome library filter set (provided by J. Dodgson at Michigan State University, East Lansing, MI). Contiguous cosmids were assembled by *Bam*HI, *Eco*RI, *Hin*dIII, *Not*I, *Pst*I, and *Sal*I restriction mappings, by hybridizations with probes made of various restriction fragments, and by nucleotide sequence determination.

Southern blotting and RT-PCR

Genomic DNA isolation and blotting protocols have been described elsewhere (9). Membranes were hybridized overnight in 50% formamide (v/v), 5× SSC (0.15 M NaCl and 0.015 M sodium citrate for 1× SSC (pH 7.0)), 0.02 M sodium phosphate (pH 6.7), 100 μg/ml denatured salmon sperm DNA, 1% SDS, 0.5% nonfat dry milk, and 10% dextran sulfate at 35°C. Following hybridization, membranes were washed twice in 2× SSC with 0.1% SDS at room temperature and twice in 2× SSC with 0.1% SDS at 45°C as a low stringency wash condition. After exposure to x-ray film at −70°C, the wet membranes were then washed twice in 1× SSC with 0.1% SDS followed by twice in 0.2× SSC with 0.1% SDS at 62°C as a high stringency wash condition. Because no difference was observed, results with the high stringency wash conditions were shown. Primer sets for 21384/A21990 and 22958/ACD8α993 (21384, 5′-CCGAGTGCTGAGGGAACAGGA-3′; A21990, 5′-GTGGTAAACTGCAGCCTGCAA-3′; 22958, 5′-AGTGAGCTGAGCAGGAGGATG-3′; ACD8α993, 5′-GGTGTTTGCATCAGGACCGCT-3′) were used by PCR to amplify genomic DNA fragments for probing the genomic structure of the CD8α CY1 and CY2 exons, respectively. Total RNA was extracted from chicken tissues with the TRIzol reagent (Invitrogen Life Technologies) based on the guanidinium thiocyanate method. First-strand cDNA was primed with oligo(dT) and a random hexamer and amplified by PCR using the nested primer sets CA1 and A15845 (5′-AGAGGAGGAGCAGCCAAAT-3′) followed by exCD8AF-3 (5′-AGCACCCCCARGARGGRCAGC-3′, where

*Institute of Biochemical Sciences, †National Taiwan University Medical School, National Taiwan University, Taipei, Taiwan, Republic of China

Received for publication October 19, 2005. Accepted for publication December 21, 2006.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by National Science Council Grant NSC 93-2320-B-002-015 and a grant from the Institute of Biological Chemistry, Academia Sinica, Taipei, Taiwan, Republic of China.

[2] Address correspondence and reprint requests to Dr. Chia-Lam Kuo, Institute of Biochemical Sciences, National Taiwan University, P.O. Box 23-106, Taipei, Taiwan, Republic of China. E-mail address: chialam@ccms.ntu.edu.tw

[3] Abbreviations used in this paper: IgSF, Ig superfamily; CY, cytoplasmic tail; EST, expressed sequence tag; H, hinge; L, leader; NITR, novel immune-type receptor; RHACD8, receptor homologous to α-chain of CD8; RJF, Red Jungle Fowl; TM, transmembrane; UTR, untranslated region; V, variable.

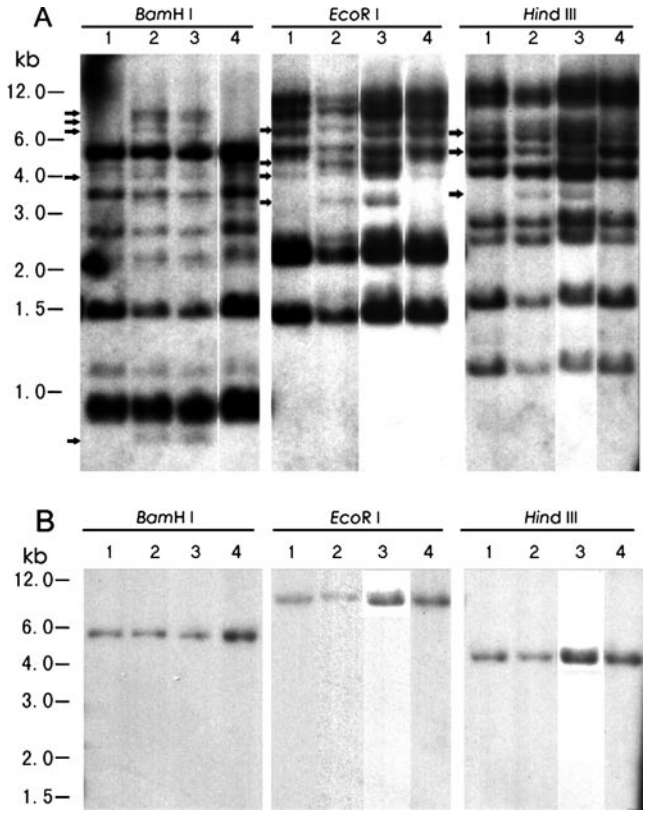

FIGURE 1. Genomic complexity of CD8α V-like sequences in chickens. A set of DNA from RBC of a New Hampshire Brown chicken (*lanes 1*), a locally bred Broil chicken (*lanes 2*), a locally bred Leghorn chicken (*lanes 3*), and a GB2 strain (*lanes 4*) were digested with *Bam*HI, *Eco*RI, and *Hin*dIII, blotted, and hybridized with probes of 268-bp PCR products with a sequence located within the CD8α V region (*A*) or 607-bp genomic PCR products composed of CD8α CY1 and a flanking intron sequence (*B*). Arrows emphasize certain positions showing differences in hybridization intensity between different strains. Size standards in kilobases are indicated on the *left side* of each panel.

R = A or G) and ChX3 (5′-TTGAGGAGGACGCCTGTGT-3′). The primers 5′-TACCACAATGTACCCTGGC-3′ and 5′-CTCGTCTTGTTTTAT GCGC-3′ were used for actin-generated fragments.

Sequence determination and analyses

Clones of cDNA, genomic restriction fragments, and PCR products were sequenced by dideoxy chain termination chemistry using universal and gene-specific infrared primers (LI-COR) in conjunction with the SequiTherm Long-Read cycle sequencing kits (Epicentre Biotechnologies). Sequences were processed via an automated sequencer (LI-COR 4000L). Putative signal peptides and TM and cytoplasmic regions were based on a trout CD8α analysis (10) and the crystal structure of human CD8α (3). Homology searches, alignments, phylogenies, and sequence diversity analyses were conducted using BLAST (www.ncbi.nlm.nih.gov/blast), ClustalW (www.ebi.ac.uk/clustalw), or EKE software (Electronic Knowledge Era).

Results

Cosmids with chicken CD8α gene

By using primers designed according to the published chicken CD8α cDNA sequence (11), we amplified bona fide chicken CD8α cDNA from a chicken spleen cDNA library. After confirming its sequence, we used it as a probe for screening a Leghorn liver DNA cosmid library and for Southern blot analyses of genomic DNA isolated from tissues originating from different chicken strains. Unexpectedly, many bands were observed in each strain when the Southern blot was probed with cDNA within the V region of the CD8α gene, indicating the presence of multiple copies of homologous DNA (Fig. 1*A*). Moreover, the banding pattern varied between different strains. Consistent with genomic DNA blots, nearly 20 cosmid clones were obtained when the cosmid library was screened with the CD8α cDNA. Among them, DNAs from three clones (denoted 7-1, 26-1, and 43-1) displayed strong hybridization signals and were further analyzed. Restriction enzyme mapping indicated that clone 43-1 overlapped with clone 7-1, whereas clone 26-1 appeared to be unlinked to either of them. We then sequenced clone 7-1 because it displayed the strongest hybridizing signal. Its 24-kb sequence contains a CD8α sequence organized into exons coding for the leader (L)[3], V, H, TM, and CY (Fig. 2). This exon-intron organization is similar to the one originally reported for the human CD8α gene. Furthermore, similar to other genes encoding members of IgSF, each of the chicken CD8α introns splits codons between the first and second nucleotide with the exception of the fifth intron, which splits the codon of exons 5 and 6 between the second and the third nucleotide.

Chicken CD8α locus contains multiple duplicated CD8α-like DNA blocks

In addition to the expected chicken CD8α sequence present in cosmid 7-1, two DNA blocks lying upstream of the CD8α gene in tandem that showed sequence homology to CD8α L, V, H, and TM and the corresponding intervening introns were noted. Southern blot results support the presence of multiple copies of DNA homologous to the CD8α L and its 5′ untranslated region (UTR)

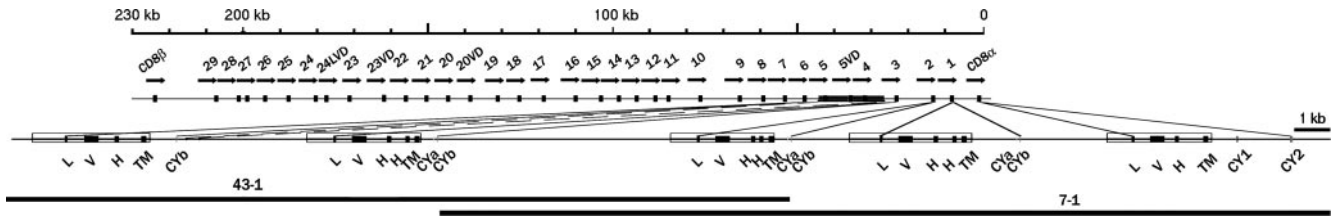

FIGURE 2. Chicken CD8α locus. Chicken CD8α and CD8β bracket a region that contains CD8α-like sequences that are arbitrarily numbered with respect to the proximity to the CD8α gene. Their relative chromosomal locations are indicated on the *top line* according to the RJF genomic sequence (GenBank locus NW_060360) and arrows indicate their orientations with respect to the CD8α gene. In the case of the CD8α-like genes, VD and LVD indicate deletion of the V exon or of both the L and V exons, respectively. Sequences homologous to genes 6, 3, 2, 1, and CD8α of the RJF strain can be found in two overlapping cosmids (7-1 and 43-1) made from the Leghorn strain, whereas homologues to genes 5, 5VD, and 4 (shaded area) of the RJF were absent in the corresponding location of the Leghorn genome. The scale for the cosmid clones is shown to the *right*, above cosmid 7-1. In the case of the 35.5-kb stretch of Leghorn sequences, exons homologous to the L, V, H, and TM of the CD8α gene are specified and regions homologous to that of CD8α gene are marked by the rectangle. The sequences 3′ to the CD8α-like TM exons in the CD8α-like blocks contain one or two exons (denoted as CYa and CYb) that possess in-frame stop codons and have no homology to the corresponding sequence in CD8α gene (exons of CY1 and CY2) containing the p56^{lck} binding motif.

(data not shown). Furthermore, the lack of sequences homologous to the region containing the CY exons led us to determine whether the CD8α CY exons were also duplicated. By using probes specific to the CY exons, Southern blot analysis showed that CD8α CY exons were not duplicated (Fig. 1*B*). Therefore, our sequencing data confirm the presence of multiple duplicated CD8α L, V, and TM-like sequences in the chicken genome (Fig. 1*A* and data not shown).

To further explore the organization of the DNA sequences flanking the CD8α gene, we analyzed cosmid 43 and obtained additional information by sequencing a stretch of 11.5 kb that lies immediately upstream of the 24-kb region sequenced from cosmid 7-1 (Fig. 2). Two similar CD8α L, V, and TM-like DNA blocks were further identified. Thus a 35.5-kb stretch of Leghorn chicken DNA revealed a genomic organization that contains the CD8α gene and four duplicated DNA blocks displaying homology to the CD8α gene (Fig. 2). Moreover, cosmid walking and Southern blotting predicted that the CD8α locus likely comprises additional copies of these CD8α-like DNA blocks.

Considering that in human and mouse the CD8α gene is separated from the CD8β gene by 25 and 36 kb, respectively, we also isolated and characterized two overlapping CD8β-containing cosmids from the same library to determine the organization of the entire chicken CD8 locus. Southern blot analyses indicate that, unlike the situation observed for CD8α, there is no duplication of CD8β-like sequence in the chicken genome. Further mapping and cloning/sequencing analyses indicated that downstream of the CD8β gene there are CD8α-like DNA blocks with sequences and structures similar to those described above in contiguity to the CD8α gene. Fluorescence in situ hybridization results showed colocalized images when using differentially labeled CD8α and CD8β probes, suggesting that the chicken CD8α gene is also linked with the CD8β gene. However, the organization is markedly different from that in humans or mice in that extensively duplicated CD8α-like DNA blocks separate the CD8α and CD8β gene in the chicken CD8 locus (Fig. 2).

Extensively duplicated CD8α V-like sequences in different chicken strains

We investigated next the duplication of the CD8α V-like sequences in other chicken strains. PCR products corresponding to the CD8α V region were cloned from the genomic DNA of a broiler chicken and a GB2 strain. Their sequences indicated that in addition to the expected CD8α sequence there exist multiple copies of CD8α-like sequences in the genome of both chicken strains. We also probed a bacterial artificial chromosome (BAC) library derived from University of California, Davis (UCD) inbred RJF DNA with various DNA fragments isolated from cosmid 7-1 and identified 13 clones possessing sequences homologous to the CD8α V region. Together with genomic DNA blots (Fig. 1*A*), this analysis suggested that extensively duplicated CD8α V-like sequences are present in chickens of different strains. Moreover, variations in the organization and/or copy number and nucleotide sequences exist between different strains.

While we were in the process of finalizing the organization of the chicken CD8 locus, the National Center for Biotechnology Information released on July 2004 the near complete chicken genome sequence of the RJF strain (GenBank locus NW_060360). By comparing it with our own sequencing data obtained from the Leghorn strain, we noted some diversity between the two strains in the nucleotide sequences and in the organization of the CD8α-like DNA blocks (Fig. 2). Importantly the information provided by the RJF genome sequence confirmed the unique features of the CD8 locus originally suggested by our data and allowed us to further refine the organization of the chicken CD8α locus that we arbitrarily define as the region starting from the end of the CD8β gene up to the end of the CD8α gene.

Molecular basis of CD8α gene polymorphism

The CD8α gene is separated from the CD8β gene by extensively duplicated CD8α-like DNA blocks that span at least 220 kb on chromosome 4 of the RJF strain. Due to remaining gaps in the draft sequence, the distance separating the CD8α and CD8β genes cannot yet be precisely determined. Alignments of the 220-kb-long sequence with the CD8α gene allowed us to identify 33 L-, 29 V-, 46 H-, and 31 TM-like gene segments. These segments were organized into 33 CD8α-like regions having a basic "L-intron-V-intron-H-intron-TM-intron" structure arranged in tandem array and in the same orientation as the CD8α gene (Fig. 2). Notably, in each of the CD8α-like DNA blocks the region homologous to the CD8α gene does not extend beyond ~230 bp downstream of the CD8α-like TM exon. This is consistent with a genomic Southern blot showing that sequences corresponding to the CD8α CY were not duplicated within the genome.

Using a BLAST search, we collated cDNA sequences of the chicken CD8α gene from GenBank and the expressed sequence tag (EST) database. Considering that no duplication of the CY exons could be detected by the Southern blot analysis, genes showing sequence identity of ≥96% to the CD8α V exon and containing the CY sequences were assumed to be bona fide *CD8α* alleles, not separate genes. Sequence alignments involving 12 CD8α-encoding sequences revealed 22 nucleotide positions, one in L, 20 in V, and one in H, showing nucleotide substitutions that yielded a total of eight alleles (Fig. 3*A*), confirming that chicken CD8α gene is polymorphic as previously suggested (8). Further alignments of the eight alleles with the genomic CD8α-like sequences indicated that the nucleotide substitutions in 16 of the 22 positions could arise by exchange of blocks of sequences (i.e., gene conversion) present in the CD8α-like DNAs or the CD8α gene (Fig. 3 and data not shown). In the remaining six polymorphic positions one of the two nucleotides in each site could result from an exchange of blocks of sequences from CD8α-like DNAs, suggesting that the other nucleotide sequence could be unique to the CD8α gene and might be present in other *CD8α* alleles that remain to be identified. Alternatively, they might also result from an exchange of blocks of sequences located in the CD8α-like DNAs that either were incompletely sequenced or exhibited strain-specific sequences as described above. Thus, gene conversion is most likely involved in generating a polymorphic CD8α gene, although the existence of a high rate of point mutations cannot be formally dismissed.

CD8α and CD8β bracket a family of novel multigenes belonging to IgSF

We then asked whether the CD8α-like sequences identified in the CD8 locus might constitute bona fide genes. In contrast to chicken Ig genes, where a pool of pseudogenes is maintained for diversification through gene conversion (12, 13), it seems unlikely that the chicken CD8α-like gene cluster is retained simply to act as a reservoir for CD8α gene conversion. It should be noted that in the annotated chicken genome database (GenBank locus NW_060360) and the corresponding article (14), the CD8α-like sequences were described as genes encoding products similar to CD8α precursors or as five additional copies of the CD8α gene, respectively. Our results based on analyzing transcripts from the EST database clearly indicate, however, that they do not correspond to five additional copies of the CD8α gene because their 3′ ends are highly diverged from CD8α, resulting in 29 genes (including pseudogenes) encoding TM and CY sequences that likely have markedly

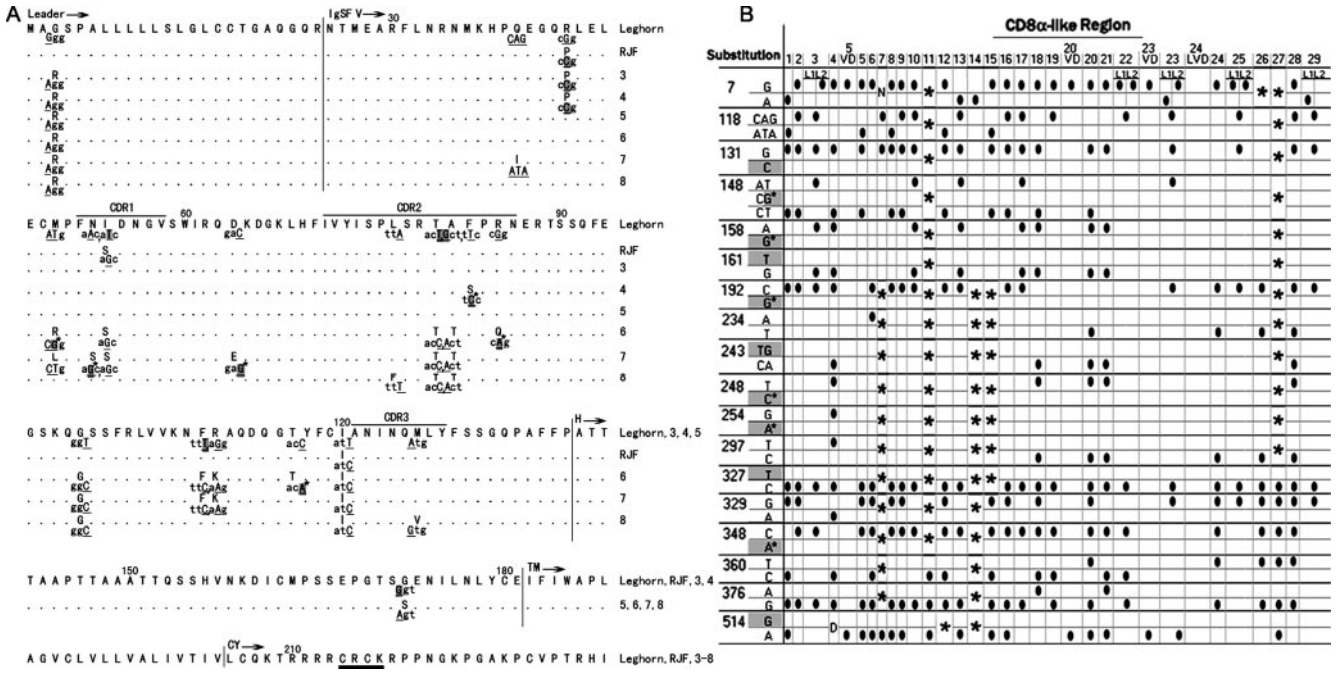

FIGURE 3. Polymorphism of chicken CD8α gene. *A*, Amino acid sequences and nucleotide substitutions among the eight chicken CD8α alleles. Dots indicate identity to the amino acid sequences shown in the *top line*, and codons containing nucleotide substitution(s) (underlined capital letters) are shown under the corresponding amino acid residues. The shaded nucleotides mark the substitutions not found in the blocks of CD8α-like DNAs, and an asterisk is added to the shaded nucleotide if the substitution is found in only one of the eight alleles. The sequences of the alleles 3 to 8 are represented by the sequences of cDNA/EST clones with the following GenBank accession numbers: allele 3 (AY528647, cDNA from thymocytes of the Langshan strain bred in China; and AY528649, cDNA from thymocytes of the Beijing fatty strain bred in China); allele 4 (AY528651, cDNA from thymocytes of the Camellia strain bred in China); allele 5 (Y11472, from strain H.B15, substrain of H12, see Ref. 8; and AY528650, cDNA from thymocytes of the Big Bone strain bred in China); allele 6 (Z22726, cDNA from the spleen of RPL line 7, see Ref. 11); allele 7 (AY519197, cDNA from thymocytes of the Leghorn strain bred in China); and allele 8 (Y11473, from strain H.B15, substrain of H12, see Ref. 8; AY528648, cDNA from thymocytes of the Xianju strain bred in China; and CD740752, cDNA from intestinal lymphocyte of the white Leghorn SC). Regions corresponding to the putative L, IgSF V domain, H, TM, and CY regions of mammalian CD8α are shown above the sequences. The putative CDR1-3 (based on Refs 3, 5, and 8) involved in MHC class I contact are specified by the overlined sequences. The cytoplasmic $p56^{lck}$ binding motif is underlined. Data from the Leghorn are from this report and data from RJF are from GenBank (locus NW_060360). Numbering of the amino acids starts from the first amino acid. *B*, Potential donors in the pool of CD8α-like DNAs in the RJF genome for the nucleotide substitutions within *CD8*α alleles. A filled oval marks the potential donor arbitrarily defined by having a block of sequences at least 6-bp long (maximum, 104 bp between CD8α and RHACD8-10 in the sequence covering the L exon and flanking introns) shared with the substitution-containing allele and by having this block embraced in a stretch of 100-bp-long DNA possessing $\geq$80% sequence identity to the substitution-containing allele. An asterisk marks the CD8α-like DNAs containing undetermined sequences where the potential donor may reside. Numbering of the CD8α-like DNAs is as in Fig. 2, and the shaded or asterisk-marked shaded nucleotide substitution is as in *A*. Numbering of the nucleotide substitutions starts from the first nucleotide of the CD8α gene. L1 and L2 are used to distinguish two L exons within a CD8α-like region, with L2 being proximal to the V-like exon. N denotes the absence of sequences homologous to the CD8α L exon, and D stands for deletion of the CD8α-like exon where the potential donor resides.

different signal transducing properties from that of the CD8α gene (Fig. 4, *A* and *B*) and are followed by a 3′ UTR including a polyadenylation signal, which is absent in the CD8α gene. This stretch of TM-, and CY-containing sequence is conserved among these CD8α-like DNAs, strongly suggesting that these CD8α-like sequences identified to date constitute a novel chicken multigene family with members possessing a basic "L-V-H-TM-CY" structure.

This novel CD8α-like multigene family is bracketed by the CD8α and CD8β genes, spans ~220 kb, and arises by extensive duplication events. These CD8α-like genes have average intron sizes of 535 bp (as shown in Fig. 4*A*) and average intergenic distances of 3 kb. They are not organized in an especially compact region as found in chicken MHC locus, where 11 genes are found in ~44 kb, with intergenic distances (excluding promoters) as small as 30 bp (15). The average GC content for the CD8α locus is ~49.36%, which is higher than the average 38.68% GC for chromosome 4q where the locus resides but lower than the MHC locus (roughly 60% GC) (14, 15). Furthermore, these CD8α-like receptor genes possess the following characteristics. First, they are most conserved in the L sequence, which is identical or nearly identical with the L sequence of the chicken CD8α gene, suggesting that these novel genes code for type I transmembrane molecules as the CD8α molecule. Second, although members of this family are diversified in evolution, the amino acids specifying the folding of their Ig V domains are remarkably conserved among the diversified V-like gene segments, indicating that this family belongs to the IgSF. Third, members of this family have variable numbers of H-like exon ranging from one to three. However, as noted in the H of the CD8α gene, nearly all of the H regions of the CD8α-like genes contain residues for *O*-glycosylation. Fourth, it is notable that the proline residue that constitutes one of three consecutive and evolutionarily conserved amino acids (tryptophan, alanine/valine, and proline) in the TM domain of the CD8α molecule is uniformly replaced in these CD8α-like genes by either a charged amino acid (arginine) or a polar residue (asparagine). Molecules with a charged residue in the TM domain generally interact with a molecule possessing a TM domain with an oppositely charged residue, and a TM segment with a polar residue, i.e., asparagine or glutamine, will drive the formation of homodimer or homotrimer of such segments in the plasma membranous environment (16, 17).

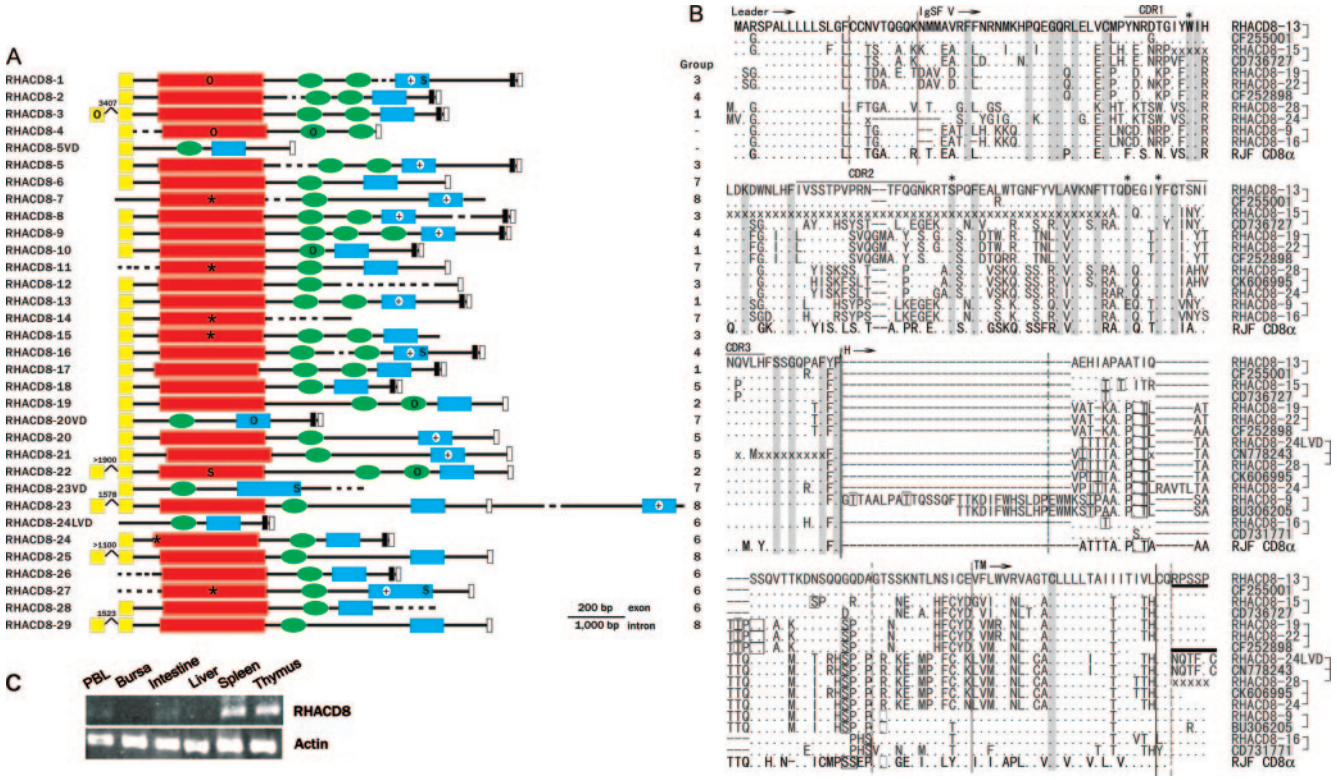

FIGURE 4. The family of RHACD8s. *A*, Exon-intron organization of RHACD8 genes on the RJF genome. Most genes are composed of a L (a yellow square), a V (a red rectangle), a H (a green oval), and a TM (a blue rectangle) exon homologous to that of CD8α and of two distinct and alternatively used exons denoted CYa and CYb (filled and open vertical bars, respectively) encoding a CY. Each gene is numbered as in Fig. 2. The presence of sequence gap is indicated by the symbol > (in the case of two contiguous L exons), by an asterisk (in the case of exons), or by broken lines (in the case of introns). O marks frame-shifted exons caused by nucleotide additions/deletions; S marks exons with stop codons and the plus sign (+) shows the presence of a positively charged residue in the TM region. Based on sequence homology analysis, members are segregated into eight groups, shown on the *right side*, except for RHACD8-4 and RHACD8-5VD, which do not fit in the defined groups. The bar corresponds to 200 and 1,000 bp in the case of exon and intron, respectively; with the exceptions that were specified in base pairs between two contiguous L exons. *B*, ClustalW alignment of amino acids deduced from EST clones, RHACD8 genes most similar to the EST clones, and CD8α of the RJF. Dots indicate identity to the top line of sequence, dashes are used to maximize the alignment, and x denotes gaps corresponding to the nucleotide sequence gaps in the RJF genome sequence or in the EST sequences. Regions corresponding to the putative CD8α-like L, V, H, and TM are shown *above* the sequences. Region borders are denoted with solid vertical lines and corresponding exon borders are shown by dashed vertical lines. The CYa-encoded sequence is overlined with a black horizontal bar, whereas the CYb-encoded sequence is underlined. Shaded residues are residues identical among all RHACD8s, available EST clones, and CD8α, except for the asterisk-marked W (all but RHACD8-22 replaced by a stop codon, and RHACD8-6 by R), S (all but RHACD8-12 by P), D (all but RHACD8-9 by conserved E), and Y (all but pseudogene RHACD8-4 by S). Residues potentially subjected to *O*-linked glycosylation (XPXX, glycosylated if X is S or T) (36) are boxed. Regions corresponding to the CD8α CDRs are designated by overlining the top line of sequence. Tissues from which EST clones (named by their GenBank accession numbers) were isolated are spleen (CF255001 and CF252898), intestine (CD736727 and CD731771), blood macrophage (CK606995), small intestine (BU306205), and a mix of thymus, bursa, spleen, PBL, and bone marrow (CN778243). *C*, Expression of RHACD8s. Primers located in V-like exon and 3′ UTR were used for PCR amplification of first-strand cDNAs from various tissues. Transcripts of the appropriate size were detected within the intestine, PBL, thymus, and spleen.

Moreover, as already mentioned, the very short CY (7 or 8 residues containing no identifiable motif (Fig. 4*B*)) of these novel gene products markedly differs from those of the CD8α and the CD8β, which further suggests that these products diverge from CD8α that contains the motif for the coupling of CD8α to p56^{lck} (18). Based on the attribute that CD8α is the only gene identified by BLAST homology

Table I. *Average nucleotide differences (%) between V gene segments of RHACD8s of different groups*[a]

Group	1	2	3	4	5	6*	7†	8‡	CD8α
1	**4.07**	21.14	22.42	23.84	24.35	26.17	30.16	32.13	21.58
2	21.14	**1.87**	24.70	23.80	26.88	31.11	26.65	31.35	26.16
3	22.42	24.70	**4.53**	13.49	22.95	27.77	29.82	25.43	22.39
4	23.84	23.80	13.49	**3.32**	24.93	27.11	28.21	25.63	22.74
5	24.35	26.88	22.95	24.93	**5.42**	14.53	22.84	28.21	15.45
6*	26.17	31.11	27.77	27.11	14.53	**8.01**	30.15	28.95	19.61
7†	30.16	26.65	29.82	28.21	22.84	30.15	**6.87**	32.40	21.30
8‡	32.13	31.35	25.43	25.63	28.21	28.95	32.40	**2.80**	29.81

[a] RHACD8-27 (*), RHACD8-11 (†), and RHACD8-7 (‡) are excluded in the analysis due to the presence of a large unresolved sequence gap in their respective V regions. Average values of differences between V regions of the same group are in boldfaced type.

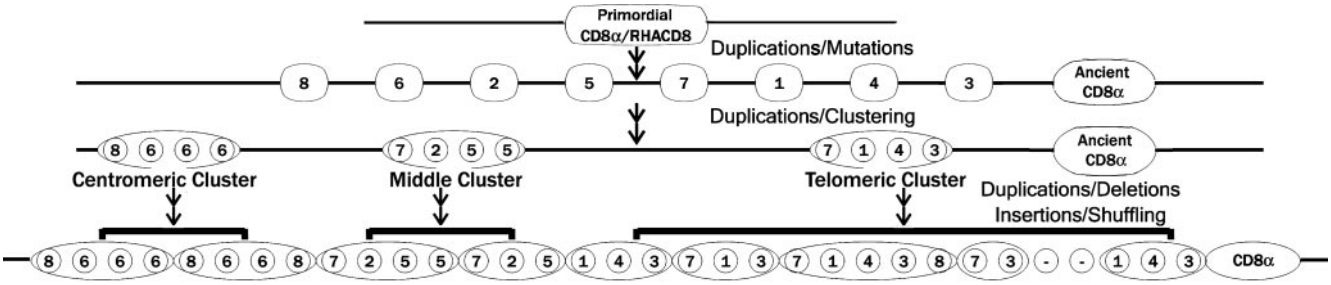

FIGURE 5. A hypothetical evolutionary pathway that gives rise to the polymorphic CD8α gene and the RHACD8 family in chickens. A region of the primordial CD8α/RHACD8 unit with a structure of CYa-space-CYb-space-L-space-V-space-H-space-TM-space that excludes the modern CD8α CY exons is hypothesized as the object of successive duplications. The first homologous unit found at the 5′ end of the cluster, ~6.7 kb downstream of CD8β gene, is structurally the same. The ancestral CD8α gene might arise by combining a sequence containing a $p56^{lck}$ binding motif 3′ to this unit, thereby constituting the cytoplasmic region 3′ to the exons of L, V, H, and TM of the CD8α/RHACD8 unit. Successive duplications of the primordial unit resulted in the individualization of a constellation of eight representative RHACD8 genes in tandem (denoted 1–8, 5′ to the ancestral CD8α gene), each thus structured with the potential to produce transcripts of L-V-H-TM-CYa and/or CYb via transcribing from L through TM in one unit into the CYa/CYb in the unit 3′ next to it (see Figs. 2 and 4*A*). Moreover, three intermediate clusters were structured on the genome by the events of duplications and unequal crossover of the eight representative RHACD8 genes. Finally, tandem duplications and unequal crossover of the intermediate clusters together with shuffling, insertions, and/or deletions of DNA blocks gave rise to the three large clusters that constitute the present polymorphic chicken RHACD8-CD8 locus.

search as homologous to members of this multigene family, we will denote this family as the family of receptors homologous to the α-chain of CD8 in chickens (RHACD8) (Fig. 4, *A* and *B*). In addition to the information deduced from EST clones, the results of the RT-PCR using nested primers that allow the amplification of ~20 RHACD8 gene-containing sequences demonstrated that these genes are expressed in the small intestine, PBL, thymus, and spleen, but not the bursa (Fig. 4*C*). Because the reverse primers were specific to the 3′ UTR conserved among these members and the forward primers were specific to the Ig V-like region, several introns were crossed and genomic DNA contamination could be readily ruled out.

The 29 V region-containing members of RHACD8s could be categorized into eight groups based on the nucleotide sequence differences calculated between their V gene segments with gaps ignored (Fig. 4*A* and Table I). The average percentage of nucleotide difference was <10% for members of the same group and 20–30% for members of different groups, with the exception of members between groups 3 and 4 that showed ~13.5% difference. Nearly identical grouping patterns were obtained for the TM exons except for RHACD8-7, -13, -18, and -27, where the V and the TM did not cosegregate. Because the H exons of RHACD8-7 and -18 cosegregated with their V regions and the H exons of RHACD8-13 and -27 with their TMs, it thus suggested that the shuffling of TM exons of RHACD8-7 and -18 and of the V domains of RHACD8-13 and -27 was involved in generating members of this family. Moreover, based on the grouping profile within the locus, we propose that at some time point during evolution the eight basic group-representing members were structured into three clusters: centromeric, middle, and telomeric. Further duplication and unequal crossover events on each cluster, including shuffling, insertions, and/or deletions of DNA blocks, then gave rise to the large centromeric, middle and telomeric clusters constituting the RJF CD8α locus (Fig. 5). The classification of groups and clusters should provide a useful base for haplotyping this evolutionarily diversified multigene family showing variations in the organization and/or gene contents between different chicken strains.

Discussion

The finding of potential donor sequences within the RHACD8 genes of the chicken RJF strain supported the view that homologous recombination events involving gene conversion give rise to the polymorphic chicken *CD8α* alleles. Further support comes from the finding that a 1,260 bp-long stretch of sequences extending from ~460 bp upstream of the L exon to ~140 bp downstream of the CD8α V exon is highly conserved in the corresponding region of RHACD8 genes (85% identity in average), which would therefore facilitate the process of gene conversion within this region. This view is also validated by the observed diversity displayed by the RHACD8 genes.

What might be the effects on chickens for being polymorphic in CD8α contributed by the expansion and diversification of this gene family in evolution? It has been documented that in humans and mice the α3 domain of the MHC class I molecule interacts with the CDRs of the CD8α V domain. Among the eight chicken CD8α alleles there are 11 nonconserved residue changes, including two in CDR1, four in CDR2, and one near CDR1 (Fig. 3*A*). Interestingly, the chicken class I α3 domain is polymorphic, because among 37 chicken cDNA sequences there are eight nonconserved residue changes, including one in the AB loop and three in the CD loop (data not shown), contrasting with the nonpolymorphic nature of the α3 domain of mice and humans. These findings therefore suggest that the interaction between CD8α and MHC I could be evolutionarily conserved and important enough to serve as a strong force to give rise to polymorphic MHC I α3 domain in chickens, a process considered as a relatively uneasy one because the chicken B locus contains only two class I genes and no pseudogene (15). Based on the involvement of CD8α in the process of T cell development in the thymus and T cell activation in the periphery, further investigations on CD8α structure and the interaction of CD8$\alpha\beta$ or CD8$\alpha\alpha$ with its ligand in individual chicken strains may help us to understand whether the unusual polymorphic nature, as compared with that of humans and mice, contributes to unique immunity features such as the tight association between the haplotype and viral susceptibility (reviewed in Ref. 19) observed in chickens.

In addition to indirectly affecting the chicken T cell by producing polymorphic CD8α as described above, the RHACD8 molecules may be directly involved in regulating CD8α function. For example, the expression of RHACD8 genes in the thymus raises an intriguing question of whether they pair with CD8α or substitute for CD8α in CD8$\alpha\beta$, which is possible considering the existence of NKG2C (lysine residue in TM)/CD94/DAP12 (aspartic acid residue in TM) complex in NK cells (reviewed in Ref. 20) and thus play a role similar to that of CD8α' in mice. The expression of alternatively spliced CD8α', lacking the Lck-binding site, in nearly equal levels to CD8α in the mouse thymus (21) is known to

permit the expression of CD8$\alpha'\beta$ heterodimers with normal ligand-binding properties but "tuned down" Lck activation properties. The combination of RHACD8 and CD8β would affect the ligand-recognition and/or divert the Lck signal pathway of CD8$\alpha\beta$. Therefore, further studies of RHACD8-interacting molecules should provide important clues as to whether there is an evolutionarily conserved mechanism involved in regulating the T cell development, and the use of CD8α' represents an alternative means in the mouse thymus. Furthermore, CD8$\alpha\alpha$ has been shown to function differentially from CD8$\alpha\beta$ in mice. For example, CD8$\alpha\alpha$ has a unique and strong preference to interact with the nonclassical MHC class I molecule TL (22), can be expressed on different T cell subsets and dendritic cells (23, 24), and has been suggested to endorse a specific function to CD8$\alpha\alpha$-expressing cells (25). It is known that both CD8$\alpha\alpha$ and CD8$\alpha\beta$ exist in chickens and that the pattern of expression of these two molecules in different lymphoid tissues is generally conserved in chickens and mice (11). Therefore, it would be interesting to determine whether these RHACD8 molecules could interact with CD8α and how such interaction might affect chicken immunity.

To understand further the way these new genes have evolved, National Center for Biotechnology Information databases were extensively analyzed. No significant positive hits were found when using the nucleotide sequences of the Ig V domain of RHACD8 genes, except for the expected CD8α genes from ducks and mammals and the related EST clones from chickens. It could be the effect of gene conversion occurring over evolutionary time between these clustered genes that thus maintains the CD8α-like sequence conserved among them. However, interesting results were obtained by PHI-BLAST searches that take a motif pattern and protein sequence as input to look for proteins containing conserved regions similar to the motif entered. The sequences with an expect (E) value better than the default threshold from species occupying phylogenetic positions below avian stood out. They are rainbow trout novel immune-type receptors (NITR)-2 (27% identity and 50% similarity to RHACD8-18, 28% identity and 48% similarity to RHACD8-20), zebra fish NITR-4a (28% identity and 47% similarity to RHACD8-2 and -9), and nurse shark IgW (28% identity and 44% similarity to RHACD8-28). For CD8α there are NITR-1 (32% identity and 51% similarity) and IgH (27% identity and 39% similarity) from channel catfish, Igκ (31% identity and 44% similarity) from nurse shark, and IgL from Siberian sturgeon (28% identity and 47% similarity). NITR has been identified in bony fish during the search for phylogenetic origins of the Ag-binding receptors (26). Members of this IgSF have V ectodomain(s) exhibiting extensive germline variation and often containing J-like sequences. With only a few exceptions, fish NITRs contain either a CY with an ITIM or a TM segment with a positively charged amino acid. Human DAP12, a transmembrane polypeptide involved in immunoreceptor signaling and possessing both a negatively charged TM segment and an ITAM in its CY was able to interact with the positively charged NITR TM segments and to function as an activating subunit for the NITR (26–32). Recently, it was shown that sequence variations of the V domains of zebra fish NITRs are distributed in a manner similar to CDR diversities in Igs and TCRs and that the existence of NITR seemed to be limited to bony fish (30).

The RHACD8 multigene family shows a high level of complexity. Each chicken strain may express a unique RHACD8 haplotype distinct from that of another strain in its gene content and organization. Importantly, each of the RHACD8 IgSF V domains is associated with a charged or polar TM domain. These genetic characteristics and the detected similarity by PHI-BLAST described above led us to propose that chicken RHACD8s may be evolutionarily related to fish NITRs. In addition, the detection of RHACD8 transcripts in spleen, small intestine, and macrophages implies that the members of this novel gene family likely encode receptors involved in chicken innate immunity. Furthermore, the genomic organization, level of complexity, and IgSF V domain-containing property of the RHACD8 genes resemble very much that of the human killer Ig-like receptor (KIR) gene family. The later are polymorphic, clustered in tandem in the genome, derived by gene duplication, and evolve rapidly, and members with short CY have a Lys residue located within the TM that is required for association with the DAP12 protein (reviewed in Ref. 20). In contrast with RHACD8 genes, KIR genes contain more than one V ectodomain and also include inhibiting members with ITIMs in the CY region (31–34). Therefore, future functional studies to determine whether the products of RHACD8 genes, siblings of the CD8α, could recognize classical/nonclassical MHC class I, class I-related molecules, or even perhaps the decoy ligands encoded by viruses as those described for NK receptors in humans and mice (reviewed in Refs. 20 and 33–35) are expected to further our understanding of both the evolutionary relationship among these immunoreceptors important to innate immunity and the immune system of chickens, a species of great economic value to humans.

Acknowledgments

We thank Bernard Malissen and Fonald Yee for critical reading of the manuscript and Jerry Dodgson at Michigan State University for providing the RJF BAC library filter set.

Disclosures

The authors have no financial conflict of interest.

References

1. Janeway, C. A., Jr. 1992. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. *Annu. Rev. Immunol.* 10: 645–674.
2. Salter, R. D., R. J. Benjamin, P. K. Wesley, S. E. Buxton, T. P. Garrett, C. Clayberger, A. M. Krensky, A. M. Norment, D. R. Littman, and P. Parham. 1990. A binding site for the T-cell co-receptor CD8 on the α 3 domain of HLA-A2. *Nature* 345: 41–46.
3. Gao, G. F., J. Tormo, U. C. Gerth, J. R. Wyer, A. J. McMichael, D. I. Stuart, J. I. Bell, E. Y. Jones, and B. K. Jakobsen. 1997. Crystal structure of the complex between human CD8$\alpha(\alpha)$ and HLA-A2. *Nature* 387: 630–634.
4. Gao, G. F., Z. Rao, and J. I. Bell. 2002. Molecular coordination of $\alpha\beta$ T-cell receptors and coreceptors CD8 and CD4 in their recognition of peptide-MHC ligands. *Trends Immunol.* 23: 408–413.
5. Sanders, S. K., R. O. Fox, and P. Kavathas. 1991. Mutations in CD8 that affect interactions with HLA class I and monoclonal anti-CD8 antibodies. *J. Exp. Med.* 174: 371–379.
6. Sun, J., D. J. Leahy, and P. B. Kavathas. 1995. Interaction between CD8 and major histocompatibility complex (MHC) class I mediated by multiple contact surfaces that include the α 2 and α 3 domains of MHC class I. *J. Exp. Med.* 182: 1275–1280.
7. Kern, P. S., M. K. Teng, A. Smolyar, J. H. Liu, J. Liu, R. E. Hussey, R. Spoerl, H. C. Chang, E. L. Reinherz, and J. H. Wang. 1998. Structural basis of CD8 coreceptor function revealed by crystallographic analysis of a murine CD8$\alpha\alpha$ ectodomain fragment in complex with H-2Kb. *Immunity* 9: 519–530.
8. Luhtala, M., C. A. Tregaskes, J. R. Young, and O. Vainio. 1997. Polymorphism of chicken CD8-α, but not CD8-β. *Immunogenetics* 46: 396–401.
9. Wang, K., L. Gan, C. L. Kuo, and L. Hood. 1997. A highly conserved apoptotic suppressor gene is located near the chicken T-cell receptor α chain constant region. *Immunogenetics* 46: 376–382.
10. Hansen, J. D., and P. Strassburger. 2000. Description of an ectothermic TCR coreceptor, CD8 α, in rainbow trout. *J. Immunol.* 164: 3132–3139.
11. Tregaskes, C. A., F. K. Kong, E. Paramithiotis, C. L. Chen, M. J. Ratcliffe, T. F. Davison, and J. R. Young. 1995. Identification and analysis of the expression of CD8 $\alpha\beta$ and CD8 $\alpha\alpha$ isoforms in chickens reveals a major TCR-$\gamma\delta$ CD8 $\alpha\beta$ subset of intestinal intraepithelial lymphocytes. *J. Immunol.* 154: 4485–4494.
12. Reynaud, C. A., A. Dahan, V. Anquez, and J. C. Weill. 1989. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* 59: 171–183.
13. Carlson, L. M., W. T. McCormack, C. E. Postema, E. H. Humphries, and C. B. Thompson. 1990. Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion. *Genes Dev.* 4: 536–547.
14. Hillier, L. W., W. Miller, E. Birney, W. Warren, R. C. Hardison, C. P. Ponting, P. Bork, D. W. Burt, M. A. Groenen, M. E. Delany, et al. 2004. Sequence and

comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.
15. Kaufman, J., S. Milne, T. W. Gobel, B. A. Walker, J. P. Jacob, C. Auffray, R. Zoorob, and S. Beck. 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923–925.
16. Call, M. E., J. Pyrdol, M. Wiedmann, and K. W. Wucherpfennig. 2002. The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell* 111: 967–979.
17. Call, M. E., and K. W. Wucherpfennig. 2005. The T cell receptor: critical role of the membrane environment in receptor assembly and function. *Annu. Rev. Immunol.* 23: 101–125.
18. Turner, J. M., M. H. Brodsky, B. A. Irving, S. D. Levin, R. M. Perlmutter, and D. R. Littman. 1990. Interaction of the unique N-terminal region of tyrosine kinase p56lck with cytoplasmic domains of CD4 and CD8 is mediated by cysteine motifs. *Cell* 60: 755–765.
19. Kaufman, J. 2000. The simple chicken major histocompatibility complex: life and death in the face of pathogens and vaccines. *Philos. Trans. R. Soc. London B* 355: 1077–1084.
20. Lanier, L. L. 2005. NK cell recognition. *Annu. Rev. Immunol.* 23: 225–274.
21. Tagawa, M., H. Nakauch, L. A. Herzenberg, and G. P. Nolan. 1986. Formal proof that different-size Lyt-2 polypeptides arise from differential splicing and post-transcriptional regulation. *Proc. Natl. Acad. Sci. USA* 83: 3422–3425.
22. Leishman, A. J., O. V. Naidenko, A. Attinger, F. Koning, C. J. Lena, Y. Xiong, H. C. Chang, E. Reinherz, M. Kronenberg, and H. Cheroutre. 2001. T cell responses modulated through interaction between CD8$\alpha\alpha$ and the nonclassical MHC class I molecule, TL. *Science* 294: 1936–1939.
23. Traver, D., K. Akashi, M. Manz, M. Merad, T. Miyamoto, E. G. Engleman, and I. L. Weissman. 2000. Development of CD8α-positive dendritic cells from a common myeloid progenitor. *Science* 290: 2152–2154.
24. Hayday, A., E. Theodoridis, E. Ramsburg, and J. Shires. 2001. Intraepithelial lymphocytes: exploring the third way in immunology. *Nat. Immunol.* 2: 997–1003.
25. Gangadharan, D., and H. Cheroutre. 2004. The CD8 isoform CD8$\alpha\alpha$ is not a functional homologue of the TCR co-receptor CD8$\alpha\beta$. *Curr. Opin. Immunol.* 16: 264–270.
26. Strong, S. J., M. G. Mueller, R. T. Litman, N. A. Hawke, R. N. Haire, A. L. Miracle, J. P. Rast, C. T. Amemiya, and G. W. Litman. 1999. A novel multigene family encodes diversified variable regions. *Proc. Natl. Acad. Sci. USA* 96: 15080–15085.
27. Litman, G. W., M. K. Anderson, and J. P. Rast. 1999. Evolution of antigen binding receptors. *Annu. Rev. Immunol.* 17: 109–147.
28. Yoder, J. A., M. G. Mueller, S. Wei, B. C. Corliss, D. M. Prather, T. Willis, R. T. Litman, J. Y. Djeu, and G. W. Litman. 2001. Immune-type receptor genes in zebrafish share genetic and functional properties with genes encoded by the mammalian leukocyte receptor cluster. *Proc. Natl. Acad. Sci. USA* 98: 6771–6776.
29. Hawke, N. A., J. A. Yoder, R. N. Haire, M. G. Mueller, R. T. Litman, A. L. Miracle, T. Stuge, L. Shen, N. Miller, and G. W. Litman. 2001. Extraordinary variation in a diversified family of immune-type receptor genes. *Proc. Natl. Acad. Sci. USA* 98: 13832–13837.
30. Yoder, J. A., R. T. Litman, M. G. Mueller, S. Desai, K. P. Dobrinski, J. S. Montgomery, M. P. Buzzeo, T. Ota, C. T. Amemiya, N. S. Trede, et al. 2004. Resolution of the novel immune-type receptor gene cluster in zebrafish. *Proc. Natl. Acad. Sci. USA* 101: 15706–15711.
31. Shilling, H. G., L. A. Guethlein, N. W. Cheng, C. M. Gardiner, R. Rodriguez, D. Tyan, and P. Parham. 2002. Allelic polymorphism synergizes with variable gene content to individualize human KIR genotype. *J. Immunol.* 168: 2307–2315.
32. Rajalingam, R., P. Parham, and L. Abi-Rached. 2004. Domain shuffling has been the main mechanism forming new hominoid killer cell Ig-like receptors. *J. Immunol.* 172: 356–369.
33. Moretta, L., and A. Moretta. 2004. Killer immunoglobulin-like receptors. *Curr. Opin. Immunol.* 16: 626–633.
34. Uhrberg, M. 2005. The KIR gene family: life in the fast lane of evolution. *Eur. J. Immunol.* 35: 10–15.
35. Lodoen, M. B., and L. L. Lanier. 2006. Natural killer cells as an initial defense against pathogens. *Curr. Opin. Immunol.* 18: 391–398.
36. Classon, B. J., M. H. Brown, D. Garnett, C. Somoza, A. N. Barclay, A. C. Willis, and A. F. Williams. 1992. The hinge region of the CD8 α chain: structure, antigenicity, and utility in expression of immunoglobulin superfamily domains. *Int. Immunol.* 4: 215–225.