

Analysis of the reptile *CD1* genes: evolutionary implications

Zhi Yang¹ · Chunyan Wang¹ · Tao Wang¹ · Jianhui Bai¹ · Yu Zhao² · Xuhan Liu¹ · Qingwei Ma^{3,4} · Xiaobing Wu⁵ · Ying Guo¹ · Yaofeng Zhao¹ · Liming Ren¹

Received: 29 December 2014 / Accepted: 15 April 2015 / Published online: 29 April 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract CD1, as the third family of antigen-presenting molecules, is previously only found in mammals and chickens, which suggests that the chicken and mammalian CD1 shared a common ancestral gene emerging at least 310 million years ago. Here, we describe *CD1* genes in the green anole lizard and Crocodylia, demonstrating that CD1 is ubiquitous in mammals, birds, and reptiles. Although the reptilian CD1 protein structures are predicted to be similar to human CD1d and chicken CD1.1, CD1 isotypes are not found to be orthologous between mammals, birds, and reptiles according to phylogenetic analyses, suggesting an independent diversification of CD1 isotypes during the speciation of mammals, birds, and reptiles. In the green anole lizard, although the single CD1 locus and MHC I gene are located on the same chromosome, there is an approximately 10-Mb-long sequence in between, and interestingly, several genes flanking the CD1 locus belong

to the MHC paralogous region on human chromosome 19. The CD1 genes in Crocodylia are located in two loci, respectively linked to the MHC region and MHC paralogous region (corresponding to the MHC paralogous region on chromosome 19). These results provide new insights for studying the origin and evolution of CD1.

Keywords CD1 · Reptile · Isotype · Evolution

Introduction

CD1, which is a family of antigen-presenting molecules, can bind bacterial and autologous lipid, glycolipid, and lipopeptide antigens for presentation to T and NKT cells (Brigl and Brenner 2004; Jayawardena-Wolf and Bendelac 2001; Matsuda and Kronenberg 2001; Moody et al. 2004; Porcelli and Modlin 1999). Although CD1 is related to both MHC class I and class II molecules (Koch et al. 2005; Martin et al. 1986; Porcelli 1995), CD1 is structurally more closely related to MHC class I molecules due to high sequence identity, similar domain organization, and association with β 2-microglobulin (β 2m) (Calabi and Milstein 1986; Martin et al. 1986; McMichael et al. 1979). However, CD1 is functionally more similar to MHC class II, as the tissue distribution of CD1 is highly restricted (Brigl and Brenner 2004; Dougan et al. 2007). CD1 molecules are expressed on the surface of antigen-presenting cells, and most CD1 proteins appear to be localized to endosomal MHC II compartments, in which the MHC II molecules are thought to be loaded with exogenous antigens (Sugita et al. 1996).

The mammalian CD1 family is composed of five nonpolymorphic genes (*CD1A*, *CD1B*, *CD1C*, *CD1D*, and *CD1E*), which are categorized into three groups based on their genomic organization, sequence identity, and cellular

Electronic supplementary material The online version of this article (doi:10.1007/s00251-015-0837-2) contains supplementary material, which is available to authorized users.

✉ Liming Ren
renlm@cau.edu.cn

¹ State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, People's Republic of China

² Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

³ Bioyong (Beijing) Technology Co., Ltd., Beijing, People's Republic of China

⁴ Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, People's Republic of China

⁵ Anhui Province Key Laboratory for Conservation and Exploitation of Biological Resource, College of Life Sciences, Anhui Normal University, Wuhu, People's Republic of China

functions: group 1 (CD1a, CD1b, and CD1c), group 2 (CD1d), and group 3 (CD1e) (Adams and Luoma 2013; Brigl and Brenner 2004). The sequence similarity is substantially higher for the same isotypes from different species than that for different isotypes within the same species (Porcelli 1995), suggesting that each group of CD1 molecules has a different function and present different antigens. The CD1 isotypes are differentially expressed with restricted tissue distribution and can interact with both $\gamma\delta$ and $\alpha\beta$ T cells (Balk et al. 1991; Castano et al. 1995; Jayawardena-Wolf and Bendelac 2001; Moody et al. 2004; Porcelli et al. 1989; Porcelli and Modlin 1999). Despite the functional importance of CD1 in mammals, non-mammalian CD1s have previously been only found in chickens, which have two *CD1* genes that are comprised of two isotypes (Maruoka et al. 2005; Miller et al. 2005; Salomonsen et al. 2005).

The genomic locations of CD1 genes are different between mammals and chickens, which may reflect the origin and evolution of CD1 to some degree. In mammals, the CD1 and MHC gene loci are located on different chromosomes (Albertson et al. 1988; Calabi and Milstein 1986; Dascher and Brenner 2003). For example, in humans, CD1 and MHC genes are located on two paralogous chromosomes, chromosome 1 and chromosome 6, respectively. However, the chicken *CD1* genes are closely linked with MHC genes in the same region (Maruoka et al. 2005; Miller et al. 2005; Salomonsen et al. 2005).

It is now commonly accepted that CD1 genes originated from MHC I in some stage of vertebrate evolution (Dascher 2007; Kasahara 1999; Martin et al. 1986; Maruoka et al. 2005; Miller et al. 2005; Salomonsen et al. 2005). However, it remains still controversial how and when this occurred. Currently, three models have been proposed to explain the associated evolution of CD1 with MHC genes: 1, Class I genes duplicate to give class I and CD1 genes in the primordial MHC gene locus, both of which are then distributed in different MHC paralogous regions during 2R (two rounds of whole genome duplication in vertebrate evolution) (Holland et al. 1994; Ohno 1970), followed by differential silencing of MHC and CD1 genes in different paralogous regions for different lineages (Salomonsen et al. 2005). 2, class I genes in the primordial MHC are distributed in different MHC paralogous regions during 2R, followed by evolution of the class I gene to CD1 in one paralogous region, retention of class I gene in another paralogous region and silencing of the class I genes in the other two paralogous regions (Kasahara 1999). 3, class I genes in the primordial MHC are distributed in different MHC paralogous regions during 2R, followed by retention of class I genes in two paralogous regions and silencing of the class I genes in the other two paralogous regions, and then evolution of class I gene to CD1 in one paralogous region in a close ancestor of birds and mammals (Miller et al. 2005).

In this paper, we describe several reptilian *CD1* genes that are homologous to mammalian and chicken CD1. The results revealed that reptiles express distinct CD1 isotypes that are not orthologous to mammalian or chicken CD1, suggesting that the CD1 isotypes formed independently in the distinct species during speciation. The analysis of the genomic locations of reptilian CD1 showed features that are both identical to and different from mammalian and chicken CD1. These results provide a new opportunity to trace the origin and evolution of CD1.

Materials and methods

Animals, DNA and RNA isolations, and reverse transcription

Approximately 3-year-old green anole lizards (*Anolis carolinensis*) were purchased from a local pet market in Beijing. The Siamese crocodile (*Crocodylus siamensis*) was purchased from a crocodile breeding farm in Tianjin, and the Chinese alligator (*Alligator sinensis*) tissue samples were collected from the Anhui Research Centre for the Reproduction of the Chinese Alligator. The genomic DNA was isolated using a standard phenol-chloroform extraction method. The total RNA from the different tissues was prepared using a TRIzol kit (Tiangen Biotech, Beijing, China). The reverse transcription was conducted using M-MLV reverse transcriptase following the manufacturer's instructions (Invitrogen, Beijing, China).

Amplification of the conserved CD1 cDNA fragments using degenerate primers

Two degenerate primers (CD1F: 5'-CCY RTK GCT GTG GTC TTT GCC C-3'; CD1R: 5'-CTS CKG AKC TGG TAG GTS AGG TCG-3') were designed according to the previously reported chicken CD1 cDNA sequences (Miller et al. 2005) and the green anole lizard CD1 sequence identified in this study. RT-PCR using reptile spleen cDNA was carried out under the following conditions: 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 50 °C for 30 s, and 72 °C for 30 s; and a final extension at 72 °C for 7 min. The polymerase used was the LA-Taq DNA polymerase (Takara, Dalian, China). The resultant PCR product was cloned into the pMD19-T vector (Takara, Dalian, China) and sequenced.

Amplification of the complete cDNA sequences

We used the 3' RACE System for Rapid Amplification of cDNA Ends (Invitrogen, Beijing, China) for the 3'-end amplification. The RACE PCRs were performed according to the manufacturer's instructions. Two primers were derived from

the conserved sequences for the nest-PCR reaction for the Siamese crocodile (CrsiF1: 5'-TTT GCC GGG TCA CTG GCT TC-3'; CrsiF2: 5'-TGC GGG ATG GTG AGG AGG TG-3') and the green anole lizard (AncaF1: 5'-GGA GCC TCC TGC AAC CAC TG-3'; AncaF2: 5'-CCT TGC CCA GCT CAA GGA TC-3'). The PCRs were performed using total spleen cDNA under the following conditions: 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 90 s; and a final extension at 72 °C for 7 min. The polymerase used was the LA-Taq DNA polymerase. The resultant PCR products were cloned into pMD19-T and sequenced. We used the 5' RACE System for Rapid Amplification of cDNA Ends (Invitrogen, Beijing, China) for the 5'-end amplification with three specific primers that were designed using the cDNA sequence obtained from the Siamese crocodile (CrsiCD1.1gsp1: 5'-GCC ACG ATG AGA AGT GTG AC-3'; CrsiCD1.1gsp2: 5'-TGC TGT GTT CCA CAT GAC AG-3'; CrsiCD1.1gsp3: 5'-CAG CTG GTA GGT CAG ATC TG-3'; CrsiCD1.2gsp1: 5'-CTG CAG CCA ACA ATG AAA TG-3'; CrsiCD1.2gsp2: 5'-CGC AGC TGG TAG GTC AGG TC-3'; CrsiCD1.2gsp3: 5'-GCA GCC AGG TCA TGT GAA TG-3') and the green anole lizard (AncaCD1gsp1: 5'-CTG CAG ATA GAA TAA CAC TC-3'; AncaCD1gsp2: 5'-TCC CAC AGG ATG ACA AGA CT-3'; AncaCD1gsp3: 5'-CCT GCA AAC ATA ACT GTG TG-3'). Gsp1 was used to synthesize the first-strand cDNA. The RACE PCRs were performed according to the manufacturer's instructions. The resultant PCR products were cloned into the pMD19-T vector and sequenced. We designed specific primers based on the products of the 3'RACE and 5'RACE to amplify the entire Siamese crocodile CD1 cDNA sequences (CrsiCD1.1F: 5'-AGA AGC CCC TCC AAA GCC TG-3'; CrsiCD1.1R: 5'-AAT GGA AGA AGG AGA GAA TC-3'; CrsiCD1.2F: 5'-TGC GAT GAT GCA GCA GCT TCC-3'; CrsiCD1.2R: 5'-CCT CCG TAA CTG AGA GAG AAC-3') and the green anole lizard CD1 cDNA sequences (AncaCD1F: 5'-TGG CCT GCA GAT ATT TCC TG-3'; AncaCD1R: 5'-CTG CTT TAG ATG AAC TTA AG-3'). Additionally, primers (AlsiCD1.1F: 5'-CCA GAG CAT GCT GCC TCC TCT-3'; AlsiCD1.1R: 5'-ATC CAC TGC TTT ATA ACA CAC-3'; AlsiCD1.2F: 5'-TGC TCG CCT TCC CCA TGT CAT-3'; AlsiCD1.2R: 5'-CCT GGT CTT GCT TAG TTC AAG-3') were designed for the amplification of two transcribed Chinese alligator CD1 genes.

Southern blotting

The $\alpha 1$ domain-encoding sequences of the Siamese crocodile and green anole lizard CD1 were used as probes for Southern blotting. These cDNA fragments were labeled using a PCR DIG Probe Synthesis Kit (Roche, Beijing, China) using the following primers: CrsiCD1.1pF: 5'-TGC GTC TGC TGC AGA CCA TC-3'; CrsiCD1.1pR: 5'-CCC ATA GAG CAC

TGA GTC AC-3'; CrsiCD1.2pF: 5'-TCC CCC TCC CTT GCC TCT TT-3'; CrsiCD1.2pR: 5'-CCA CGG TGA CAA CTG TAT TG-3'. AncaCD1pF: 5'-GTC CAT CCA GCC TTC TTT CA-3'; 5'-CTG CAA CCA TGT TGT TGA TG-3'. The hybridization and detection were performed using the DIG High Prime DNA Labeling and Detection Starter Kit II (Roche, Beijing, China), following the manufacturer's instructions.

Detection of the gene expression in different Siamese crocodile tissues via quantitative real-time PCR

The cDNA samples from seven tissues (heart, liver, spleen, lung, kidney, small intestine, and stomach) were used to detect the CD1 expression in the Siamese crocodile via quantitative real-time PCR. The PCRs were performed using a LightCycler 480 and the LightCycler 480 SYBR Green I Master Mix (Roche, Beijing, China). Each sample was run in triplicate. The Siamese crocodile *EF1a1* gene was chosen as the internal control. The PCRs were performed under the following conditions: 95 °C for 10 min; 35 cycles of 95 °C for 10 s, 60 °C for 20 s, and 72 °C for 15 s; and a final extension at 72 °C for 7 min. The PCR primers were as follows: CrsiCD1.1F2: 5'-CTC AGG CAA GTG GGT AGC TC-3'; CrsiCD1.1R2: 5'-TGT GTC AAT TGT GCC CTT GT-3'; CrsiCD1.2F2: 5'-TGC AGT TCC TGC TCC AGA AC-3'; CrsiCD1.2R2: 5'-TCC TGC CTC TTC AGT GTC TC-3' and EF1a1F: 5'-TGA TGC TCC TGG ACA CAG AG-3'; EF1a1R: 5'-GCC CAT TCT TGG AGA TAC CA-3'.

Sequence alignments, comparisons, three-dimensional structural modeling, and construction of the phylogenetic tree

MegAlign (DNASar/Lasergene) (Hein 1990) was used for the sequence comparisons and identity calculations. The three-dimensional structures of the reptile CD1 were predicted via SWISS-MODEL (<http://swissmodel.expasy.org/>). The PDB files used in analysis are 1ZT4 (human CD1d), 3JVG (chicken CD1.1), and 3DBX (chicken CD1.2). PyMOL was used to display the cartoon representation. The phylogenetic tree was made using MrBayes3.1.2 (Ronquist and Huelsenbeck 2003) and viewed in TreeView (Page 1996). Multiple sequence alignments were performed using Clustal X1.83 (Thompson et al. 1997). The accession numbers of the sequences used for the comparisons and constructions were as follows: human (*Homo sapiens*) huCD1a: NP_001754.2; huCD1b: NP_001755.1; huCD1c: NP_001756.2; huCD1d: NP_001757.1; huCD1e: CAA33100.1; HLA-A: NP_002107.3; HLA-B: NP_005505.2; HLA-C: NP_002108.4; HLA-DRB1: NP_002115.2; HLA-DRB3: NP_072049.2; northern brown bandicoot (*Isodon macrourus*) IsmaCD1: ABI99485.1; chicken (*Gallus gallus*) chCD1.1: AAX49403.

1; chCD1.2: AAX49406.1; BF1: NP_001038148.1; BF2: NP_001026509.1; BLB1: NP_001038159.1; BLB2: NP_001038144.2; *Xenopus* (*Xenopus laevis*) Xela-Ia: AAA16064.1; Xela-IIb: NP_001108243.1; zebrafish (*Danio rerio*) DareDEB: NP_571552.1; Dare-UBA: NP_571546.1; nurse shark (*Ginglymostoma cirratum*) Gici-UAA01: AAF66110.1; Gici-IIb1: AAF82681.1; green anole lizard (*Anolis carolinensis*) AncaCD1: KJ191193; Chinese alligator (*Alligator sinensis*) AlsiCD1.1: KJ191190; AlsiCD1.2: KJ191189; Siamese crocodile (*Crocodylus siamensis*) CrsiCD1.1: KJ191192; CrsiCD1.2: KJ191191; spectacled caiman (*Caiman crocodilus*) CacrIa: AHC72441.1; CacrIIb: AAF99284.1; inshore hagfish (*Eptatretus burgeri*) IgSF3: BAE93396.1.

Results

Identification of the *CD1* genes in certain reptiles

Reptiles and birds belong to the Reptilia and share a common ancestor approximately 220 million years ago (Mya) (Kumar and Hedges 1998). As the *CD1* gene has been identified in chickens (Miller et al. 2005; Salomonsen et al. 2005), likewise the *CD1* gene is highly likely also present in reptiles. Using the chicken *CD1* amino acid sequences, we searched the green anole lizard genomic databases in Ensemble (Assembly AnoCar2.0) and discovered a homologous sequence at position 189410454–189410738 in chromosome 2. Upon sequence alignment, we found that the sequence was more similar to the *CD1* than the *MHC I* gene. 5' RACE and 3' RACE were thus performed to obtain a complete cDNA sequence for the green anole lizard *CD1*. Based on the green anole lizard and chicken *CD1* sequences, a pair of degenerate primers was designed to screen for *CD1* genes in other reptiles. The PCR reactions were performed using spleen-derived cDNA from the red-eared turtle (*Trachemys scripta elegans*), Siamese crocodile (*Crocodylus siamensis*), beauty snake (*Orthriophis taeniurus*), and Burmese python (*Python bivittatus*). Because of the non-specificity of the primers, a homologous sequence was only amplified in the Siamese crocodile, and two similar full-length cDNA sequences were gained via the 5' RACE and 3' RACE.

Using the human, chicken, green anole lizard, and Siamese crocodile *CD1* sequences, we searched the Chinese alligator (*Alligator sinensis*) genomic database and the other available reptilian genomic databases. Several predicted *CD1* sequences and *CD1*-like sequences were found from the genomic data on NCBI, including those in the Chinese alligator (*Alligator sinensis*), American alligator (*Alligator mississippiensis*), Burmese python (*Python bivittatus*), green sea turtle (*Chelonia mydas*), Chinese soft-shell turtle (*Pelodiscus sinensis*), and

western painted turtle (*Chrysemys picta*) (Supplemental table 1). The majority of the *CD1* sequences were not used for further analyses because some predicted sequences were overlong or incomplete. Three Chinese alligator *CD1* sequences were identified in the genomic database. RT-PCR using spleen-derived cDNA was subsequently employed to confirm whether they are transcriptionally functional. The results showed that two of the three sequences were transcribed, and the third one was not detected via the RT-PCR. For the two transcribed sequences, one was functional, while the other one seemed to be a pseudogene due to 13 missing nucleotides in $\alpha 2$ encoding exon, leading to a premature stop codon.

Phylogenetic analysis of the *CD1*, *MHC I*, and *MHC II* genes

In comparison with the full-length chicken and human *CD1* and *MHC I* amino acid sequences, the amino acid sequence identities of the full-length reptilian *CD1* sequences for the green anole lizard, Siamese crocodile, and Chinese alligator are ~25.9–38.3 % with the *CD1* sequences and ~20.1–25.3 % with the *MHC I* sequences, while a comparison with the conserved $\alpha 3$ domain showed that the identities are ~36.6–65.3 % and ~24.2–40.1 %, respectively (data not shown).

To deduce the phylogenetic relationships of the reptilian *CD1* genes with *CD1*, *MHC I*, and *MHC II* genes in other species, we used human, chicken and reptile *CD1*, and the full-length fish, *Xenopus laevis*, reptile, chicken, and human *MHC Ia* and *MHC IIb* amino acid sequences to perform phylogenetic analyses. The phylogenetic analyses were performed independently using three methods including Bayesian, neighbor-joining, and maximum likelihood, and these analyses generated trees with a very same topology. The results strongly supported that the identified sequences in reptiles were *CD1* genes, as they formed a unique clade with *CD1* but not *MHC* genes from other species (Fig. 1, Supplemental Fig. 1).

The phylogenetic analysis also revealed that the crocodylian *CD1* should be divided into two distinct isotypes, but the crocodylian isotypes are distinct from the chicken and human *CD1* isotypes. Meanwhile, it was also revealed that the green anole lizard *CD1* was not orthologous to any of the Crocodylia, chicken, or human *CD1* isotypes. We therefore designated the detected reptilian *CD1* genes as *AncaCD1* (GenBank No: KJ191193) for the green anole lizard, *CrsiCD1.1* (GenBank No: KJ191192) and *CrsiCD1.2* (GenBank No: KJ191191) for the Siamese crocodile, and *AlsiCD1.1* (GenBank No: KJ191190) and *AlsiCD1.2* (GenBank No: KJ191189) for the Chinese alligator, which

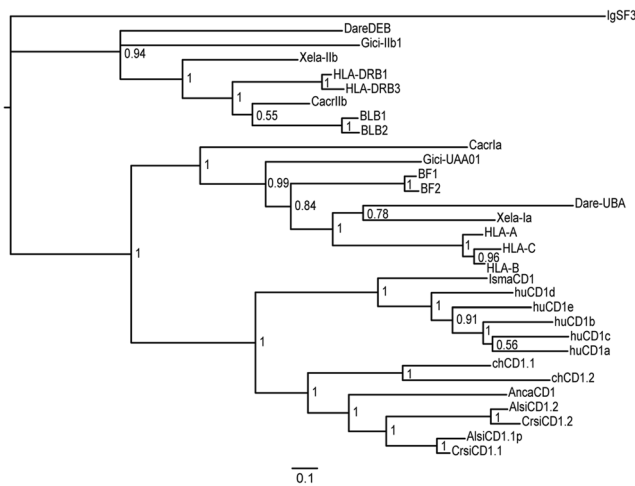


Fig. 1 Phylogenetic tree of the full-length amino acid sequences of fish, amphibian, reptile, bird, and mammalian CD1, MHC I, and MHC II. The phylogenetic tree was constructed using MrBayes3.1.2 and is viewed in TreeView. The credibility value for each node is shown. The inshore hagfish (*Eptatretus burgeri*) immunoglobulin superfamily 3 gene (IgSF3) (AB242223), which has an Ig-like domain that is similar to MHC and CD1, was used as the outgroup in the phylogenetic analysis

was based on the MHC and chicken CD1 nomenclature (Klein et al. 1990; Miller et al. 2005; Salomonsen et al. 2005).

Southern blotting and expressional analyses of the reptilian CD1s

A Southern blotting was performed using the CrsiCD1.1 and CrsiCD1.2 α 1 domain-encoding sequences as probes (Fig. 2a, b). The results showed several bands

for *CrsiCD1.2*. *AlsiCD1.1* and *AlsiCD1.2*, suggesting that there are several copies of these genes in the genome. In contrast, no more than two bands were observed for *CrsiCD1.1* after digestion with each of different restriction enzymes, suggesting that this gene is likely a single copy gene. Another Southern blotting analysis was performed using the AncaCD1 α 1 sequence as a probe (Fig. 2c). The results showed that, likely, green anole lizard CD1 has only one copy of the CD1 gene.

To analyze the crocodilian CD1 expressional pattern, we designed two pairs of qRT-PCR primers according to the full-length *CrsiCD1.1* and *CrsiCD1.2* cDNA sequences. The qRT-PCRs were performed using cDNA from seven tissues from the Siamese crocodile (Fig. 3). The results showed that the highest level of expression of both *CrsiCD1.1* and *CrsiCD1.2* is in the spleen. Both *CrsiCD1.1* and *CrsiCD1.2* displayed low expression levels in the other tissues.

We observed multiple bands in the RT-PCR of *AlsiCD1* from total RNA, and the bands were cloned and sequenced. The results showed that both *AlsiCD1.1* and *AlsiCD1.2* genes expressed multiple transcripts; *AlsiCD1.1* has six different transcripts (X1 to X6), and *AlsiCD1.2* has four (X1 to X4) expressed in the spleen, lung, and small intestine. These distinct transcripts should have arisen from RNA splicing, since a strict RNA splicing rule (i.e., GT-AG splicing site, in rare case GC-AG) was observed when they were aligned with their respective genomic sequences (Supplemental Fig. 2). The schematic splicing patterns of all variants for both *AlsiCD1.1* and *AlsiCD1.2* are shown in Fig. 4. Perhaps because of the

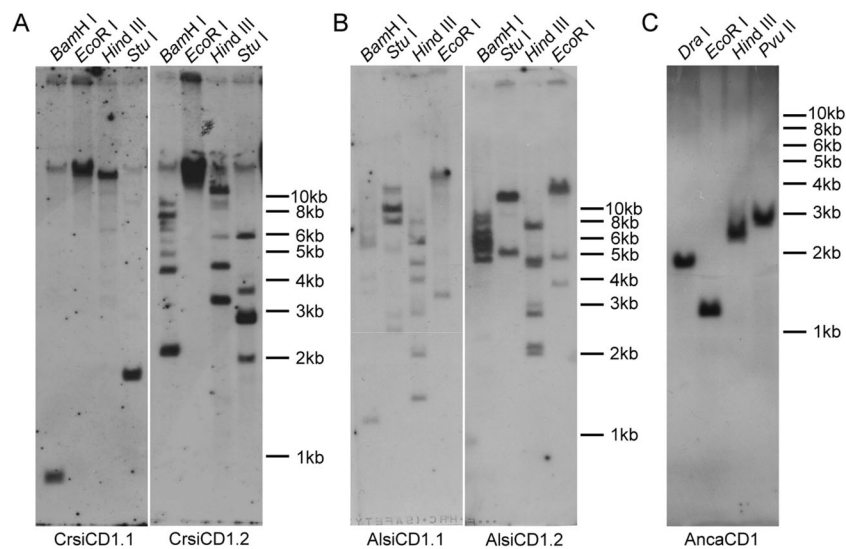


Fig. 2 Southern blotting of the Siamese crocodile, Chinese alligator, and green anole *CD1* genes. The probes were designed based on the CrsiCD1.1 and CrsiCD1.2 α 1 sequences and were used for both the Siamese crocodile and Chinese alligator CD1 due to the high similarity between the α 1 sequences of the two crocodiles. The probe for the green

anole lizard CD1 was designed using its α 1 sequence. Four restriction enzymes were used for each Southern blotting and are indicated at the top. **a** The Southern blotting result of the Siamese crocodile *CD1* genes. **b** The Southern blotting result of Chinese alligator *CD1* genes. **c** The Southern blotting result of green anole lizard *CD1* genes

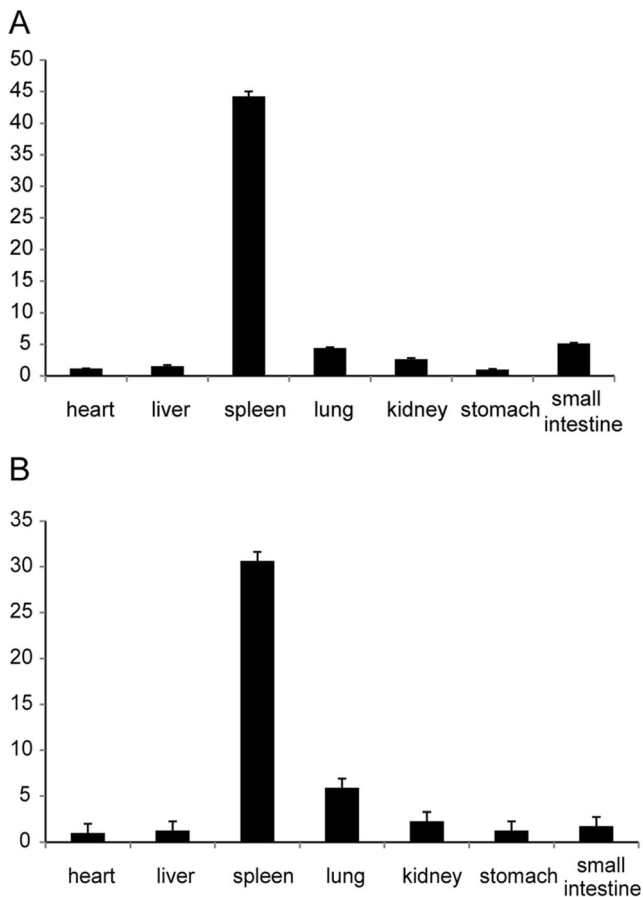


Fig. 3 The tissue expression of *CrsiCD1.1* and *CrsiCD1.2*. The expression levels of *CrsiCD1.1* and *CrsiCD1.2* were examined via qRT-PCR. The Siamese crocodile *eEF1A1* gene was used as an internal control. The seven tissues are listed under the x-axis. The y-axis indicates normalized expression folds. **a** The expression levels of *CrsiCD1.1* in different tissues. **b** The expression levels of *CrsiCD1.2* in different tissues

missing 13 nucleotides in the $\alpha 2$ encoding exon, all the spliced variants for *AlsiciCD1.1* involve only the $\alpha 2$ exon. In

contrast, more exons in *AlsiciCD1.2* are included in splicing events (Fig. 4a, b).

Similarly, in the spleen, *CrsiCD1.2* expresses two alternatively spliced forms, one containing a partial $\alpha 2$ domain and the other lacking the transmembrane region (Fig. 4c, Supplemental Fig. 2). *CrsiCD1.1* does not have any alternative splicing variants.

Comparison of the reptilian *CD1* genes with the human and chicken *CD1* genes

We aligned the full-length reptilian CD1 amino acid sequences with those of two chicken CD1 and five human CD1 for comparison (Fig. 5). The sequence identities are ~50.6–85.7 % among all of the crocodilian CD1 sequences, ~25.9–30.6 % between the reptile and human CD1, and ~33.9–38.3 % between the crocodilian and chicken CD1. The conserved $\alpha 3$ domain show greater similarities than the full-length CD1 between reptiles, chickens, and humans (data not shown). The conserved cysteines that exist in chCD1.1 (C98–C163, C202–C260) and the human CD1s, with the exception of chCD1.2 (Miller et al. 2005; Salomonsen et al. 2005), also exist in most reptilian CD1s, but one of them is missing in *AlsiciCD1.1*, obviously due to a sequence deletion (Fig. 5).

The N-linked N-X(S/T) glycosylation sites in human CD1 and the reptilian CD1 were predicted using NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>). We found nine clusters of glycosylation sites in 12 CD1 sequences, which are marked N1 to N9 (Fig. 5). N1 is highly conserved in the human, chicken, and green anole lizard, but not in crocodiles, whereas N9 is conserved in three reptiles but not in human and chicken. The cytoplasmic tails of human CD1b, -c and -d contain a positively charged membrane anchor followed by the sequence SYQ (huCD1b, SYQNIP; huCD1c, SYQDIL;

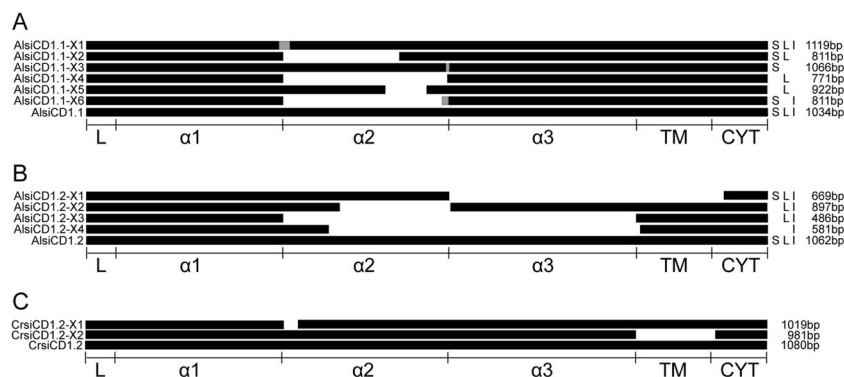


Fig. 4 A schematic diagram showing the alternative splicing variants in Chinese alligator and Siamese crocodile CD1. *AlsiciCD1.1* and *AlsiciCD1.2* cDNA fragments from the spleen, lung, and small intestine, *CrsiCD1.2* cDNA fragments from the spleen were cloned following RT-PCR. Ten to 30 clones from each tissue were sequenced and aligned. The identical residues are indicated in black, and the missing or inserted nucleotides from the sequenced clones are indicated in white or light gray,

respectively. The alternative splicing forms derived from the different tissues are indicated by letters to the right of the schematic diagram; S spleen, L lungs, and I small intestine. **a** Seven PCR products were observed: the full-length *AlsiciCD1.1* is on the bottom. **b** Five PCR products were observed: the full-length *AlsiciCD1.2* is on the bottom. **c** Two PCR products were observed: the full-length *CrsiCD1y2* is on the bottom

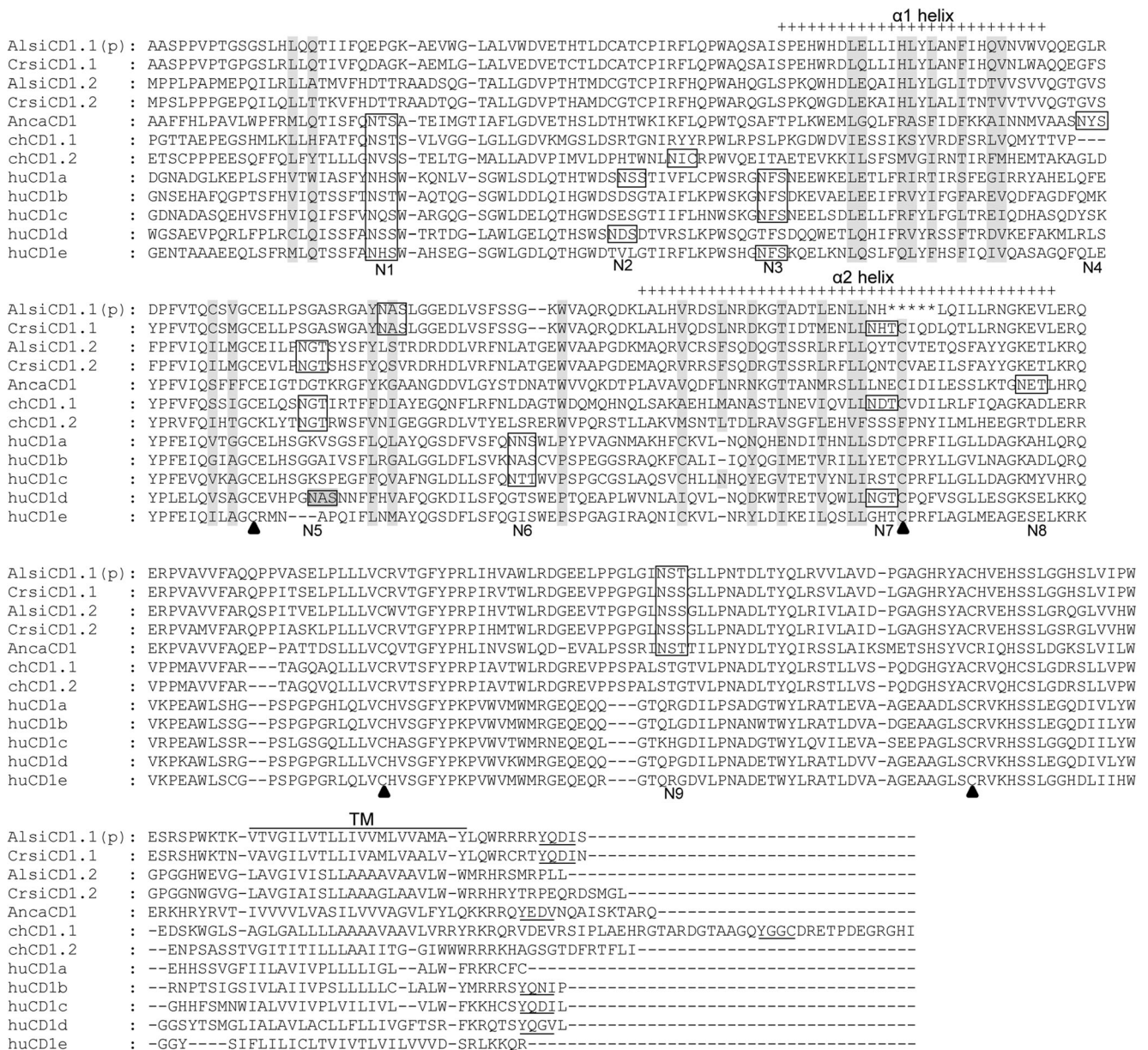


Fig. 5 Alignment of the $\alpha 1$, $\alpha 2$, and $\alpha 3$ domains of the amino acid sequences of all analyzed reptilian CD1s, chicken CD1s, and human CD1s. The $\alpha 1$ and $\alpha 2$ helices are marked with a “+”. The gaps in the alignments and partial sequences are filled with *dashed lines*. The conserved intramolecular disulfide bonds are indicated using *black triangles*. The letter “N” at the bottom of the sequences indicates the positions of the N-linked (NXS/T) glycosylation sites; those that have

been proven using the crystal structure or were predicted by NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc>) are marked with *rectangles*. The potential glycosylation site in human CD1d is marked with a gray-shaded rectangle. N1 to N9 show the nine different N-linked glycosylation sites. The groove-forming residues in chCD1.2 (Zajonc et al. 2008) are marked in *gray*. The transmembrane region is marked *above the line*

huCD1d, SYQGVL). A similar sequence was also found in the cytoplasmic tails of the two murine CD1 proteins (SAYQDIR) (Blumberg et al. 1995). The motif YQXI/V (where X can be any amino acid) in the cytoplasmic tails may also be a signal for internalization and targeting to an endosomal compartment (Sandoval and Bakke 1994). Conversely, the motif YGGC was found in the cytoplasmic tails of chCD1.2 (Miller et al. 2005; Salomonsen et al. 2005). In our study, an identical YQDI motif was found in the cytoplasmic tails of both

crocodilian CD1.1 (Fig. 5). A variant motif, YEDV, was found in the cytoplasmic tails of the green anole lizard CD1.

According to the crystal structure analysis of chCD1.2, 23 groove-forming residues were found to be identical or similar to the human CD1 (Zajonc et al. 2008). In the reptilian CD1, these residues were found to be identical or similar to the human or chicken CD1 (Fig. 5). We performed structural modeling of the reptilian CD1 using SWISS-MODEL. The results

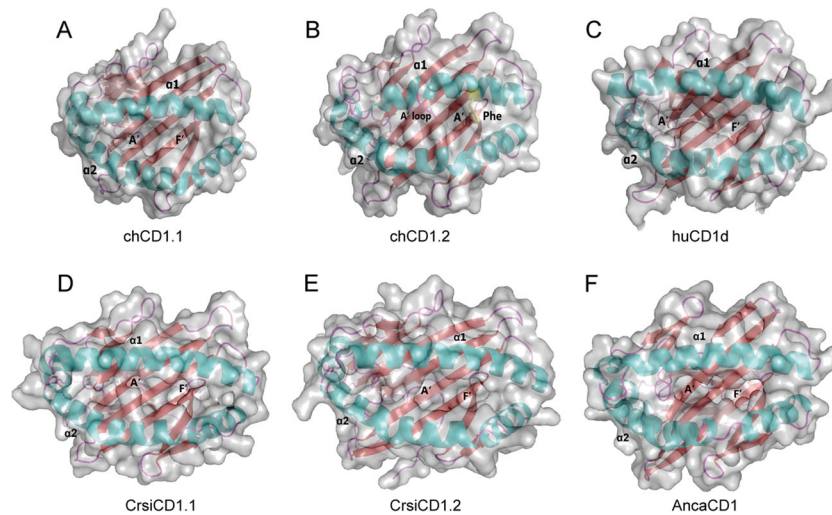


Fig. 6 The structural modeling of the partial reptilian CD1. The PDB files used in analysis are 1ZT4 (human CD1d), 3JVG (chicken CD1.1), and 3DBX (chicken CD1.2). The structural modeling of the reptilian sequences was performed using SWISS-MODEL (<http://swissmodel.expasy.org/>). The cartoon representation was prepared using PyMOL.

The $\alpha 1$ and $\alpha 2$ helices are colored in *cyan*; the β -sheets are colored in *red*. The two pockets are indicated by A' and F'. The A' loop is marked in chCD1.2. Phe, which has a large benzene ring side-chain that can block the entrance of the F' pocket and lead to a missing F' pocket, is shown in chCD1.2 (**b**)

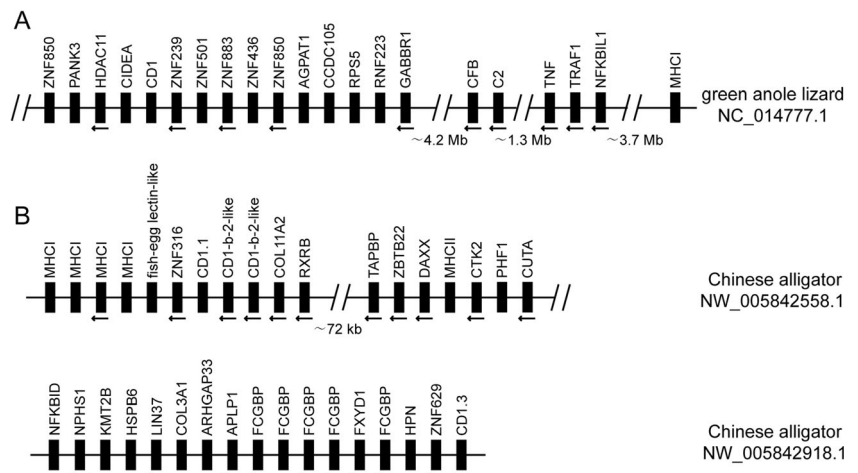
suggest that all of the analyzed reptilian CD1 have a dual-pocket (A' and F'), similar to chCD1.1 and huCD1d (Fig. 6).

The genomic locations of the reptilian *CD1* genes

The genomic locations of chicken *CD1* in the MHC region and human *CD1* in a MHC paralogous region have previously been reported (Calabi and Milstein 1986; Dascher and Brenner 2003; Miller et al. 2005; Salomonsen et al. 2005). The question instead becomes whether the reptilian *CD1* genes are more similar to chickens or humans in terms of their genomic locations. Based on the NCBI genomic database, we found that, similar to the situation in chickens, the green anole lizard *AncaCD1* and Chinese alligator *AlsiCD1.1* genes are located in the MHC locus, whereas the third Chinese

alligator *CD1* gene (the *AlsiCD1.3*, partial sequence) is located in a distinct MHC paralogous region (Fig. 7, supplemental table 2). In detail, the *AlsiCD1.1* is located in scaffold 634_1 (GenBank accession No: NW_005842558.1), in which the MHC locus is also located. There are many CD1-like sequences observed flanking the *AlsiCD1.1*, some of which show more than 80 % amino acid sequence identities with *AlsiCD1.2*, which is consistent with the results of the Southern blotting. *AlsiCD1.2* is found in scaffold 1413_1 (GenBank accession No: NW_005843837.1), and there are no other genes predicted in this scaffold. The third Chinese alligator *CD1* gene (*AlsiCD1.3*, GenBank accession No: XP_006036211) is found in scaffold 113_1 (GenBank accession No: NW_005842918.1), which contains similar genes to those located on human MHC paralogous chromosome 19 (Wan et al. 2013).

Fig. 7 A schematic diagram showing some annotated genes flanking the CD1 genes described in this study. The exact locations of these genes are shown in Supplemental table 2. *Arrow* shows the transcriptional orientation and the regions not included are represented as “//”. **a** Green anole lizard chromosome 2. **b** Chinese alligator scaffold 634_1 and scaffold 113_1



An analysis of the green anole lizard genome database showed that *AncaCD1* is located on chromosome 2 (GenBank accession No: NC_014777.1). Although the lizard MHC locus is located on the same chromosome, the distance between the CD1 and MHC I gene is approximately 10 Mb. Many genes between CD1 and MHC I, such as *GABBR1* and *AGPAT*, belong to the MHC I, MHC II, or MHC III regions, and interestingly, other genes, such as *RNF223*, *RPS5*, *CCDC105*, and *ZNF850* closely flanking the CD1 gene, are same to those located in the MHC paralogous regions on human chromosomes 1 and 19 (3 genes on chromosome 19, 1 on chromosome 1), respectively. This shows that the lizard CD1 gene is more tightly linked to the MHC paralogous region, although it is located on the same chromosome together with MHC genes.

Discussion

In the present study, we have identified CD1 genes in three reptiles including the green anole lizard, Chinese alligator, and Siamese crocodile, demonstrating that CD1 is ubiquitous in reptiles, birds, and mammals. However, CD1 genes differ considerably in gene number, isotypes, and genomic locations in these species, which, on the other hand, may allow us to track the evolutionary process of the CD1 genes.

The types of reptile *CD1* genes can be divided into two isotypes in Crocodylia, whereas the isotypes of *CD1* genes in chickens and mammals are divided into two and five, respectively. Similar to the chicken CD1 isotypes, each of two crocodylian CD1 isotypes is not orthologous to any of five mammalian CD1 isotypes as revealed by phylogenetic analysis. The same thing also applies to the CD1 isotypes between crocodiles and chickens. This strongly suggests an independent diversification of CD1 isotypes during the speciation of mammals, birds, and reptiles.

The finding that CD1 is ubiquitous in reptiles, birds, and mammals suggests that CD1 should have emerged in ancestral species common to reptiles, birds, and mammals. Given that CD1 gene was originated from MHC I gene (Dascher 2007; Kasahara 1999; Martin et al. 1986; Maruoka et al. 2005; Miller et al. 2005; Salomonsen et al. 2005), the major issue remains still unclear: when and how this genetic event occurred. As described previously in the “Introduction”, this is the most intriguing but a quite controversial issue for comparative CD1 studies up to date. There have been three main models proposed to address this puzzling issue, and all three models are based on the finding of four MHC paralogous regions in many vertebrates and 2R hypothesis for vertebrate evolution (Hokamp et al. 2003; Ohno 1970). Briefly, the first model assumes that CD1 was duplicated from MHC I in the primordial MHC, both of which are then distributed in different paralogous regions during 2R and additional genetic events

further deleted or silenced CD1 or MHC I in different paralogous region for different lineages (Salomonsen et al. 2005). The second model hypothesizes that only MHC I was originally distributed to four paralogous MHC region during 2R, followed by evolution of the class I gene to CD1 in one paralogous region, retention, or silencing of the class I genes in the other paralogous regions (Kasahara 1999). The third model differs from the second one by assuming that the evolution of MHC I to CD1 in one paralogous region occurred not in the early stage of vertebrate evolution but late in a close ancestor of birds and mammals (Miller et al. 2005).

In favor of the first model are some evidence derived from this study. The CD1 genes in Crocodylia are located in two loci, respectively linked to the MHC region and a MHC paralogous region (corresponding to the MHC paralogous region on human chromosome 19). In the green anole lizard, the CD1 gene is also more tightly linked to the same MHC paralogous region, although it is located on the same chromosome together with MHC genes. Considering that the chicken CD1 is linked to MHC and human CD1 is located in the MHC paralogous region on chromosome 1 (Calabi and Milstein 1986; Dascher and Brenner 2003; Miller et al. 2005; Salomonsen et al. 2005), CD1 can indeed be found in three of four different MHC paralogous regions albeit in different species. This seems to perfectly implicate that CD1 genes emerged together with the birth of the four MHC paralogous regions, and different species may retain CD1 genes differentially in these regions. However, it is hard to explain with this model why CD1 genes are not found in amphibians and fish.

The pivotal difference between models 2 and 3 is when the CD1 gene evolved from MHC I in vertebrate evolution. The data available now favor model 3 more than model 2, since CD1 has not been identified in amphibians and fish. Taken all the available information together, a slightly modified hypothesis based on model 3 seems more reasonable: class I genes in the primordial MHC are distributed in different MHC paralogous regions during 2R, followed by retention of class I genes in one paralogous regions and silencing of the class I genes in the other three paralogous regions. In the MHC region in a close ancestor of birds, reptiles, and mammals, MHC I was duplicated to generate CD1 by neofunctionalization. This could explain why the CD1 gene is linked to MHC region in several species including chickens and crocodiles as well as the green anole lizard. Then, the chromosome translocation may account for the distribution of CD1 in other MHC paralogous regions in different species. Some clues to putative translocations can be derived the green anole lizard, as in this species, the CD1 is not only located on the same chromosome with MHC genes, but also tightly associated with genes located in MHC paralogous regions on human chromosome 19 and 1. A similar hypothesis was also previously proposed by Dascher C (Dascher 2007), but the author assumes that the MHC I did not emerge in all four MHC paralogous region

originally but appeared in one paralogous region after emergence of jawed vertebrates.

In summary, we have identified CD1 genes in several reptiles and deduced their isotypes and genomic locations. These data are helpful to understand the origin of CD1 genes in the context of MHC I gene evolution. Further analysis of gene components of MHC paralogous regions in more jawed vertebrates such as amphibians, teleost, and cartilaginous fish would be expected to generate more clues to this puzzling issue.

Acknowledgments This study was supported by the National Natural Science Foundation of China (31272433).

References

- Adams EJ, Luoma AM (2013) The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. *Annu Rev Immunol* 31:529–561. doi:10.1146/annurev-immunol-032712-095912
- Albertson DG, Fishpool R, Sherrington P, Nacheva E, Milstein C (1988) Sensitive and high resolution in situ hybridization to human chromosomes using biotin labelled probes: assignment of the human thymocyte CD1 antigen genes to chromosome 1. *EMBO J* 7: 2801–2805
- Balk SP, Ebert EC, Blumenthal RL, McDermott FV, Wucherpfennig KW, Landau SB, Blumberg RS (1991) Oligoclonal expansion and CD1 recognition by human intestinal intraepithelial lymphocytes. *Science* 253:1411–1415
- Blumberg RS, Gerdes D, Chott A, Porcelli SA, Balk SP (1995) Structure and function of the CD1 family of MHC-like cell surface proteins. *Immunol Rev* 147:5–29
- Brigl M, Brenner MB (2004) CD1: antigen presentation and T cell function. *Annu Rev Immunol* 22:817–890. doi:10.1146/annurev-immunol.22.012703.104608
- Calabi F, Milstein C (1986) A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. *Nature* 323:540–543. doi:10.1038/323540a0
- Castano AR (1995) Peptide binding and presentation by mouse CD1. *Science* 269:223–226
- Dascher CC (2007) Evolutionary biology of CD1. *Curr Top Microbiol Immunol* 314:3–26
- Dascher CC, Brenner MB (2003) Evolutionary constraints on CD1 structure: insights from comparative genomic analysis. *Trends Immunol* 24:412–418
- Dougan SK, Kaser A, Blumberg RS (2007) CD1 expression on antigen-presenting cells. *Curr Top Microbiol Immunol* 314:113–141
- Hein J (1990) [39] Unified approach to alignment and phylogenies. In: *Methods in Enzymology*, vol Volume 183. Academic Press, pp 626–645. doi:http://dx.doi.org/10.1016/0076-6879(90)83041-7
- Hokamp K, McLysaght A, Wolfe KH (2003) The 2R hypothesis and the human genome sequence. *J Struct Funct Genom* 3:95–110
- Holland PW, Garcia-Fernandez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. *Development (Cambridge, England) Supplement*:125–133
- Jayawardena-Wolf J, Bendelac A (2001) CD1 and lipid antigens: intracellular pathways for antigen presentation. *Curr Opin Immunol* 13: 109–113
- Kasahara M (1999) The chromosomal duplication model of the major histocompatibility complex. *Immunol Rev* 167:17–32
- Klein J et al (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* 31:217–219
- Koch M et al (2005) The crystal structure of human CD1d with and without alpha-galactosylceramide. *Nat Immunol* 6:819–826
- Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. *Nature* 392:917–920. doi:10.1038/31927
- Martin LH, Calabi F, Milstein C (1986) Isolation of CD1 genes: a family of major histocompatibility complex-related differentiation antigens. *Proc Natl Acad Sci U S A* 83:9154–9158
- Maruoka T, Tanabe H, Chiba M, Kasahara M (2005) Chicken CD1 genes are located in the MHC: CD1 and endothelial protein C receptor genes constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals. *Immunogenetics* 57:590–600. doi:10.1007/s00251-005-0016-y
- Matsuda JL, Kronenberg M (2001) Presentation of self and microbial lipids by CD1 molecules. *Curr Opin Immunol* 13:19–25
- McMichael AJ, Pilch JR, Galfre G, Mason DY, Fabre JW, Milstein C (1979) A human thymocyte antigen defined by a hybrid myeloma monoclonal antibody. *Eur J Immunol* 9:205–210
- Miller MM et al (2005) Characterization of two avian MHC-like genes reveals an ancient origin of the CD1 family. *Proc Natl Acad Sci U S A* 102:8674–8679. doi:10.1073/pnas.0500105102
- Moody DB et al (2004) T cell activation by lipopeptide antigens. *Science* 303:527–531. doi:10.1126/science.1089353
- Ohno S (1970) *Evolution by gene duplication*. George Allen & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag, London
- Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Applic Biosci* : CABIOS 12:357–358
- Porcelli SA (1995) The CD1 family: a third lineage of antigen-presenting molecules. *Adv Immunol* 59:1–98
- Porcelli SA, Modlin RL (1999) The CD1 system: antigen-presenting molecules for T cell recognition of lipids and glycolipids. *Annu Rev Immunol* 17:297–329. doi:10.1146/annurev-immunol.17.1.297
- Porcelli S, Brenner MB, Greenstein JL, Balk SP, Terhorst C, Bleicher PA (1989) Recognition of cluster of differentiation 1 antigens by human CD4-CD8-cytolytic T lymphocytes. *Nature* 341:447–450
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Salomonsen J et al (2005) Two CD1 genes map to the chicken MHC, indicating that CD1 genes are ancient and likely to have been present in the primordial MHC. *Proc Natl Acad Sci U S A* 102:8668–8673. doi:10.1073/pnas.0409213102
- Sandoval IV, Bakke O (1994) Targeting of membrane proteins to endosomes and lysosomes. *Trends Cell Biol* 4:292–297
- Sugita M et al (1996) Cytoplasmic tail-dependent localization of CD1b antigen-presenting molecules to MHCs. *Science* 273:349–352
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Wan QH et al (2013) Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res* 23:1091–1105
- Zajonc DM, Striegl H, Dascher CC, Wilson IA (2008) The crystal structure of avian CD1 reveals a smaller, more primordial antigen-binding pocket compared to mammalian CD1. *Proc Natl Acad Sci U S A* 105:17925–17930