Analysis of the Chinese Alligator TCRα/δ Loci Reveals the Evolutionary Pattern of Atypical TCRδ/TCRμ in Tetrapods

Xifeng Wang, Jinwei Huang, Peng Wang, Renping Wang, Chaolín Wang, Di Yu, Cuncun Ke, Tian Huang, Yu Song, Jianhui Bai, Kongpan Li, Liming Ren, Robert D. Miller, Haitang Han, Xin Zhou and Yaofeng Zhao

*J Immunol* published online 26 June 2020
http://www.jimmunol.org/content/early/2020/06/25/jimmunol.2000257

Supplementary Material  http://www.jimmunol.org/content/suppl/2020/06/25/jimmunol.2000257.DCSupplemental

Why *The JI*? Submit online.

- **Rapid Reviews! 30 days** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

*average

Subscription  Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions  Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts  Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

*The Journal of Immunology* is published twice each month by The American Association of Immunologists, Inc., 1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2020 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Analysis of the Chinese Alligator TCRα/δ Loci Reveals the Evolutionary Pattern of Atypical TCRδ/TCRμ in Tetrapods

Xifeng Wang, Jinwei Huang, Peng Wang, Renping Wang, Chaolin Wang, Di Yu, Cuncun Ke, Tian Huang, Yu Song, Jianhui Bai, Kongpan Li, Liming Ren, Robert D. Miller, Haitang Han, Xin Zhou, and Yaofeng Zhao

Atypical TCRδ found in sharks, amphibians, birds, and monotremes and TCRμ found in monotremes and marsupials are TCR chains that use Ig or BCR-like variable domains (VHδ/Vμ) rather than conventional TCR V domains. These unconventional TCR are consistent with a scenario in which TCR and BCR, although having diverged from each other more than 400 million years ago, continue to exchange variable gene segments in generating diversity for Ag recognition. However, the process underlying this exchange and leading to the evolution of these atypical TCR receptor genes remains elusive. In this study, we identified two TCRα/δ gene loci in the Chinese alligator (Alligator sinensis). In total, there were 144 V, 154 J, nine Dδ, eight Dμ, two Ce, and five Cδ gene segments in the TCRα/δ loci of the Chinese alligator, representing the most complicated TCRα/δ gene system in both genomic structure and gene content in any tetrapod examined so far. A pool of 32 VHδ genes divided into 18 subfamilies was found to be scattered over the two loci. Phylogenetic analyses revealed that these VHδ genes could be related to bird VHδ genes, VHδ/Vμ γ genes in platypus or opossum, or alligator VH genes. Based on these findings, a model explaining the evolutionary pattern of atypical TCRδ/TCRμ genes in tetrapods is proposed. This study sheds new light on the evolution of TCR and BCR genes, two of the most essential components of adaptive immunity. The Journal of Immunology, 2020, 205: 000–000.

Cell receptors play a significant role in the specific recognition of pathogen-associated epitopes (1). Four TCR chains, called α, β, γ, and δ, are expressed as heterodimers either as α with β on αβ T cells or γ with δ on γδ T cells.

Published June 26, 2020, doi:10.4049/jimmunol.2000257

Abbreviations used in this article: BAC, bacterial artificial chromosome; RSS, recombination signal sequence.

OrCID: 0000-0001-8298-0177 (D.Y.); 0000-0002-8479-8819 (R.D.M.); 0000-0003-4499-8610 (Y.Z.).

Received for publication March 11, 2020. Accepted for publication May 22, 2020.

The sequences presented in this article have been submitted to the National Center of Biotechnology Information GenBank (https://www.ncbi.nlm.nih.gov/genbank/) under accession numbers MT081963 and MT081964.

Address correspondence and reprint requests to Dr. Haifang Han and Prof. Yaofeng Zhao or Prof. Xin Zhou, State Key Laboratory of A globotomy Technology, College of Biological Sciences, China Agricultural University, 2 Yuanmingyuan Western Road, Beijing 100193, People's Republic of China (X.Z.). E-mail addresses: hanhaitang@cau.edu.cn (H.H.) and yaofengzhao@cau.edu.cn (Y.Z.).

The online version of this article contains supplemental material.

Copyright © 2020 by The American Association of Immunologists, Inc. 0022-1767/20$37.50

www.jimmunol.org/cgi/doi/10.4049/jimmunol.2000257
similar to TCR_β has been identified, called NAR-TCR. NAR-TCR also employs two V domains, each encoded by separate, somatically rearranging VD segments (11, 18). The distal V domain of NAR-TCR is more closely related to the V domains of IgNAR than to the V domains of conventional TCR. IgNAR is an H chain homodimer that is not associated with light chains and is found only in cartilaginous fish (19, 20). It can be assumed that the structures of TCR_μ and NAR-TCR may represent an ancient receptor system. However, phylogenetic analysis and structure are consistent with NAR-TCR and TCR_μ being more likely the results of convergent evolution rather than being related by direct descent (10, 17).

The origins of TCR_μ remained unclear until the discovery of VH6 genes in the TCRα/β locus of platypus, which provided a model for the TCR_μ origin (10). Such atypical V genes that appear to be indistinguishable from VH genes but are located within the conventional TCRα/β locus have also been detected in amphibians, the platypus, passeriform birds, coelacanths, and nurse sharks (10, 14–16, 21). The VH6 gene, which is used with C_δ as part of TCRδ repertoire, was first identified as a cluster of gene segments in a reverse transcriptional orientation at the 5’ end of the Xenopus tropicalis TCRα/β locus (14). Later, one additional VHδ gene was identified in the TCRα/β locus of platypus (10). The phylogenetic relationship of VHδ and V_μ genes in platypus provided a critical link in reconstructing the evolutionary history of TCR_μ. Possibly, V gene segments have recurrently translocated between the IgH and TCRδ loci. The TCRδ genes are hypothesized to have translocated out of the TCRα/β locus early in evolution of mammals, creating the TCR_μ locus (10). Supporting this, a second TCRδ locus that uses Ig-like V domains was found in Galliformes, which is unlinked to the conventional TCRδ/β locus. Additionally, flanking genes of the second TCRδ locus cannot be found in the conventional TCRα/β locus (15). A Ig-like V segment, nestled within the TCRδ locus and expressed with TCRδ (but not IgH) was also found in the nurse shark and termed as TAILVs (16).

VH6 genes exist in TCRα/β locus perhaps because of direct transfer of VH genes from an IgH locus to the TCRα/β locus. But it should be pointed out the TAILV/VHδ/μ genes found in the nurse sharks, coelacanths, frogs, birds, and monotremes are not closely related. Phylogenetic evidence did not support TAILV as the progenitor of VHδ (16). Indeed, VHδ and V_μ genes in tetrapods appear to have derived from different, ancient VH clans. Avian VHδ are derived from VH clan I, amphibian VHδ from VH clan II, and platypus VHδ from VH clan III, for example (14, 15, 17). The translocation of VH into the TCRα/β locus occurred to have occurred independently in different lineages. Strangely, birds are thought to only have VH clan III genes in their IgH locus, yet all avian VHδ groups belong to VH clan I (10, 22). This raises interesting questions regarding the origins of the VHδ and V_μ genes in tetrapods.

In this study, we generated a detailed genomic organization for the TCRα/β locus in the Chinese alligator, an ancient reptile that represents a key evolutionary position linking amphibians, birds, and mammals (23). Notably, a large number of VHδ genes were found to be dispersed throughout the TCRα/β locus in the Chinese alligator, and phylogenetical analysis suggested some of them grouped with frog VHδ, some with bird VHδ, and some with VHδ and VHμ genes in the platypus and the opossum. Based on this study, we developed and present a model to explain the evolutionary pattern of atypical VHδ genes in tetrapod.

Materials and Methods
Sample collection, DNA and RNA extraction, and reverse transcription
Blood samples of Chinese alligators were collected from the Beijing Zoo. Genomic DNA was extracted from blood cells according to a routine protocol. Various tissues of Chinese alligator were provided by the Administration Bureau of Chinese Alligator National Nature Reserve Protection of Anhui Province. Total RNA was extracted from the spleen and thymus using a TRIzol kit (TIANGEN Biotech, Beijing, China) according to the manufacturer’s instructions. Reverse transcription was conducted using Moloney Murine Leukemia Virus Reverse Transcriptase (Promega, Madison, WI) with an oligo(dT) adapter primer. Our studies were approved by the Animal Care and Use Committee of China Agricultural University.

Bacterial artificial chromosome screening
Based on the sequences of TCRα and TCRβ derived from chicken (Co, ABU79779.1; C61, AAD51740.1 [both from GenBank]), we performed tblast against National Center for Biotechnology Information GenBank. Two scaffolds (scaffold1585_1 and scaffold955_1) were identified. Based on the sequences of Co and C61 from scaffold1585_1 and scaffold955_1, positive bacterial artificial chromosome (BAC) clones covering the Chinese alligator TCRα/β loci were isolated from the BAC library via PCR-based screening. The first positive BAC clone was sequenced from both ends, and the end sequences were used to design primers for the next round of screening to determine the BAC clone overlap. Finally, 17 positive BAC clones were sequenced by shotgun sequencing (the sequence had >7× coverage of each BAC) and assembled with a next-generation sequencing platform by Beijing Genomics Institute (Beijing, China).

Genome walking
We mapped the gap between Y251L7 and Y104L3 of Chinese alligator TCRα/β locus A by a genome walking kit (Clontech Laboratories) according to the manufacturer’s instructions. All amplifications were conducted using LA Taq DNA Polymerase (Takara, Dalian, China) with proofreading activity. The PCR products were sequenced.

Gene annotation and characterization
IgBLAST (http://www.ncbi.nlm.nih.gov/igblast/) was used to predict the V gene segments by assessing similarity to homologs from human and mouse. Germline Vα and Vβ genes were grouped into families using the ImMunoGeneTics numbering system (24). If there was more than one member in the subgroup, we named the V gene from 3’ to 5’, with the subgroup number followed by the gene segment number. The recombination signal sequences (RSSs) for the V, D, and J gene segments were analyzed by using the online programmer FUZZYNUC (http://embossgui.sourceforge.net/demofuzzynuc.html). Non-TCR genes located in or flanking the TCRα/β locus were identified using GENSCAN (http://genes.mit.edu/GENSCAN.html), and then BLAST was used on all predicted exon domains against the GenBank database using blastp to identify the genes. DNA and protein sequence editing, alignments, and comparisons were performed with MegAlign software (DNASTAR) (25). Dot blot analyses of the V regions of the TCRα/β locus were conducted with the dotter programmer (26).

Southern blotting
Restriction endonucleases were used to digest the genomic DNA. Digested genomic DNA was electrophoresed in a 0.9% agarose gel for 6 h and then transferred to a positively charged nylon membrane (Roche, Mannheim, Germany) for hybridization with individual C_λ1, C_κ2, C_β1, C_β2, C_δ5, and C_δ6 probes, which were labeled by a PCR digoxigenin probe synthesis kit (Roche). Hybridization and detection were performed according to the manufacturer’s instructions.

Construction of phylogenetic trees
The phylogenetic trees were constructed in MEGA version 6.0 (27) using the neighbor-joining method with 1000 bootstrap replicates. Multiple amino acid/nucleotide alignments for tree construction were performed using ClustalW. Each V amino acid/nucleotide alignments for tree construction were performed using the neighbor-joining method with 1000 bootstrap replicates. Multiple
enzyme, Pyrobest DNA polymerase (Takara). The PCR products were cloned into the pMD19-T vector (Takara) and sequenced.

Results
Identification of two TCRα/δ loci in the Chinese alligator

To analyze the TCRα/δ locus in the Chinese alligator, we employed a BAC genomic library that was generated by using peripheral blood leukocytes isolated from a Chinese alligator (28). Using a PCR-based approach and shotgun sequencing, 17 overlapping genomic TCRα/δ gene–positive BAC clones (Y12D21, Y135N2, Y144M13, Y251L17, Y104L3, Y257J18, Y179I18, Y107L2, Y13B10, Y183J22, Y132K5, Y75A3, Y360O6, Y173I18, Y236E17, Y96L4, and Y110J22) were identified. Upon filling a gap between Y251L7 and Y104L3 by genome walking, we obtained two contiguous distinct genomic sequences containing the alligator TCRα/δ genes. These spanned ∼547 kb and 1013 kb of DNA sequences, respectively, and we labeled them as TCRα/δ locus A and locus B.

Subsequently, we annotated and characterized the genes of the two TCRα/δ loci. As in all tetrapods analyzed so far, the genes encoding the alligator TCRα and TCRδ chains are tightly linked.

FIGURE 1. Physical map of the two Chinese alligator TCRα/δ loci. Va (VHδ or Vδ), potentially functional variable gene segments; pVa (VHδ or Vδ), pseudovariable gene segments; D, diversity gene segments; J, joining gene segments; and C, C region gene. V, D, and J gene segments and C regions were numbered based on their position on the locus (5'–3') and were color coded as follows: Vα/δ, red; D, green; J, orange; Cα, blue; Cδ, dark green; and VHδ, pink. Non-TCR genes located in or flanking the TCRα/δ locus were shaded by light gray. Transcriptional orientation is indicated by the arrow under each segment except for the D and J genes.
in both loci, with TCRδ genes nested among the TCRα genes (Fig. 1). Flanking genes METTL3 and SALL2 were detected at the S’ end of locus A, which was similar to that in human, mouse, opossum, X. tropicalis, and coelacanth (Fig. 1). We also identified DAD1 and ABHD4 at the 3’ end of locus A, as in mammals, chicken, lizard, turtle, and X. tropicalis (Fig. 1). Additionally, there was also a single Vδ gene in reverse transcriptional orientation downstream of the Cδ1 gene like in mammalian TCRαδ loci (Fig. 1). However, genes (KHNYN, CYLD, NFATC4, ADCY4, LTB4R, IGSF1, and FCRL5) located at the 5’ end of locus B were not found in TCRαδ loci of any other vertebrates. There were, however, several olfactory receptors interspersed among the Vα genes (Figs. 1, 2).

Genomic structures of the two TCRαδ loci

Based on the existence of RSSs, we have annotated a pool of 46 V genes positioned at the S’ of Dδ–Jδ–Cδ clusters, which were followed by a large number of Jα (n = 63) segments and a Cα region in locus A (Fig. 1). The striking feature of the locus A is the presence of four tandemly repeated Dδ–Jδ–Cδ clusters. There are three Dδ segments, four Jδ segments, and one Cδ in the first Dδ–Jδ–Cδ cluster. Only one Jδ segment and one Cδ was found in the second cluster. The Cδ2 appeared to be a pseudogene because of a frameshift mutation at its first exon. Both the third and fourth Dδ–Jδ–Cδ clusters had a single Dδ segment, a single Jδ segment, and one Cδ (Fig. 1). The genomic structure of locus B was simpler than the locus A. There are 98 V genes and one Dδ–Jδ–Cδ cluster containing three Dδ, three Jδ, and one Cδ. A Jα–Cα cluster containing 91 Jα segments and a single Cα is positioned downstream of the Dδ–Jδ–Cδ cluster (Fig. 1). All Dδ and Jδ/Jα gene segments are flanked by canonical RSS (Supplemental Table II).

To determine whether the TCRαδ loci in another alligator species is also organized as the similar structure to that in Chinese alligator, we identified the V, D, J, and C gene segments on two species is also organized as the similar structure to that in Chinese alligator (Alligator mississippiensis). All genes of the TCRαδ locus in American alligator are syntenic to those in Chinese alligator.

Characterization of the Cα and Cδ genes in Chinese alligator

Two Cα were detected, one from locus A and one from locus B; they share an amino acid identity of 64.8% (Figs. 1, 2). Southern blot analyses with the first exons of Cα1 and Cα2 as probes were conducted to verify the copy numbers of the Cα genes, and the two probes each detected a single hybridizing fragment (Supplemental Fig. 1A). An amino acid sequence alignment of the two Cα with those of other vertebrates revealed conserved cysteine distribution (Supplemental Fig. 2A). The Cα of Chinese alligator shared <35% sequence identity with Cα from other species (Supplemental Fig. 2A). Moreover, a phylogenetic analysis indicated that the divergence of Cα1/Cα2 genes in alligators (including the Chinese alligator and American alligator) occurred in a common ancestor (Fig. 3).

We also identified five Cδ genes in the Chinese alligator (Fig. 1). Pairwise comparisons revealed no more than 40% aa sequence identity between Chinese alligator Cδ and that of other vertebrates (Supplemental Fig. 2B). The identical cysteine distribution in the Cδ is conserved between Chinese alligator and Cδ from other vertebrate species. Southern blot analysis corroborated the presence of five Cδ genes in the genome (Supplemental Fig. 1B). We observed only a single band each when using the first exons of Cδ1, Cδ4, and Cδ5 as probes (Supplemental Fig. 1B, 1C). When the first exon of Cδ2 was used, two bands were detected in Xba I–, BamH I–, and Pst I–digested genomic DNA (Supplemental Fig. 1B), most likely because the nucleotide sequence identity between exon 1 in Cδ2 and Cδ3 is to 90.2%.

Based on a phylogenetic analysis, Cδ genes from the reptilian lineage, including birds and nonavian reptiles, clustered together (Fig. 4). They formed three subclades: Cδ1 subclade comprised all avian Cδ genes, alligator Cδ1 and Cδ5 and turtle Cδ1; Cδ II

**FIGURE 2.** Simplified representations of the TCRαδ and TCRμ loci distinct vertebrate lineages. Vαδ genes are marked in red, VHδ and Vμ genes are marked in pink, Cδ and Cμ genes are represented in dark green, and Cα genes are shown in blue. Syntenic genes shown in dark gray are those conserved with other tetrapod TCRαδ locus: methyl-transferase like 3 (METTL3), slapt-like transcription factor 2 (SALL2), olfactory receptors (OR), defender against cell death gene 1 (DAD1), and abhydrolase domain containing 4 (ABHD4). Other genes are marked by light gray.
subclade consisted of turtle Cd2 and alligator Cd2 and Cd3; and Cd III subclade was composed of turtle Cd3 and alligator Cd4 (Fig. 4).

A large number of unconventional VHd gene segments dispersed in the two TCRα/δ loci

As noted above, a total of 144 V gene segments combined were detected in two TCRα/δ loci (Fig. 1). One hundred and twelve conventional Vα/δ genes could be classified into thirty-four families using the rule that sequences sharing 75% nucleotide identity are in the same family. Based on the ImMunoGeneTics rules described previously (29), 97 Vα/δ (~86%) genes were predicted to be functional, and 15 appeared to be pseudogenes because of the presence of unusual RSS or in-frame stop codons, frameshifts, and/or mutations in the sequences (Supplemental Table II). The derived amino acid sequences of these Vα/δ families contained the canonical Vα/δ structure, which included the conservative sites Cys23, Trp41, and Cys104 (Supplemental Fig. 2C).

Thirty-two unconventional V genes were found to share greater identity with VH genes than conventional Vδ genes and were designated as VHδ genes, following the nomenclature established for X. tropicalis (14). Indeed, compared with Ig and TCR V genes from other vertebrate species, these VHδ genes were found to be most similar to VH. Phylogenetic analysis supported that the VHδ genes of Chinese alligator grouped within the same clade as the VH genes rather than the Vα/δ genes, suggesting that the VHδ genes were more closely related to the VH genes (Fig. 5). These VHδ genes were segregated into 18 families based on 75% or greater nucleotide identity (Fig. 1, Table I, Supplemental Fig. 2D).

Eighteen VHδ gene segments located on two genomic scaffolds from the American alligator were also identified and could be classified into 12 families. The position of the VHδ genes in Chinese and American alligators were highly conserved (data not shown).

VHδ genes identified in the Chinese alligator fell into three different ancient VH clans

Ig VH genes from mammals and other tetrapods have been shown to cluster into three ancient clans (22, 30). Previous studies indicated that the X. tropicalis VH all clustered with VH clan II genes (14), whereas all VHδ genes identified in coelacanth clustered into VH

FIGURE 3. Phylogeny of the TCRα constant regions in jawed vertebrates. The phylogenetic tree was constructed using TCR Ca domains in MEGA 6.0 (27) with the neighbor-joining method (42). Chinese alligator sequences are shown in bold. Bootstrap percentage values are shown at the major interior branch nodes. The horizontal length is proportional to the distance score generated by the computer program. Similar topology was generated using the maximum likelihood method.

FIGURE 4. Phylogenetic analyses of Chinese alligator TCRδ constant regions. The tree was constructed using the Ig-C domains and analyzed using the neighbor-joining method. Bootstrap values are indicated per 1000 replicates. Chinese alligator sequences are shown in bold. A distance bar is shown below the tree. In the figure, all bird Cd genes, alligator Cd1 and Cd5 and turtle Cd1 as Cd I in red; turtle Cd2, alligator Cd2 and Cd3 as Cd II in blue; and turtle Cd3 and alligator Cd4 as Cd III in green. Similar topology was generated using the maximum likelihood method.
Expression of VH\(\delta\) genes in the IgH locus. Avian VH\(\delta\) genes, in contrast, are closely related to VH clan I (15). Finally, the VH-related V\(\alpha\) and VH\(\delta\) in platypus both clustered within VH clan III (10). These observations are consistent with VH\(\delta\) having evolved independently from VH multiple times.

To analyze phylogeny of the VH\(\delta\) genes in Chinese alligator, a phylogenetic tree using VH genes from different tetrapods, V\(\alpha\) of opossum and platypus, and VH\(\delta\) of platypus, birds, alligators, soft-shell turtles, and X. tropicalis was constructed (Fig. 6, Table II). The VH\(\delta\) genes from alligators fell into three different ancient clans of VH genes: eight VH\(\delta\) subgroups (VH\(\delta\)6–9, VH\(\delta\)12–13, VH\(\delta\)15, and VH\(\delta\)17) fell into VH clan I, among which VH\(\delta\)12 clustered with platypus VH\(\delta\) in the same clade, suggesting that they share a common origin (Fig. 6). In contrast, nine subgroups, including VH\(\delta\)1–5, VH\(\delta\)10–11, VH\(\delta\)14, and VH\(\delta\)16, were located within VH clan I together with VH\(\delta\) genes of birds (Fig. 6), whereas the VH\(\delta\)18 fell into the VH clan II (Fig. 6).

In addition, Chinese alligator pVH\(\delta\)1.1, V\(\alpha\)2, VH\(\delta\)9.3, and VH\(\delta\)15 were clustered with their own VH\(\delta\), VH\(\delta\), VH\(\delta\), and VH\(\delta\), respectively (Fig. 6). Indeed, these VH\(\delta\) genes in TCR\(\alpha\)/\(\delta\) share a high degree of identity (63.4–91.1%) with their corresponding VH\(\delta\) genes in the IgH locus.

Expression of VH\(\delta\) with C\(\delta\)1, C\(\delta\)3, C\(\delta\)4, and C\(\delta\)5

To identify that VH\(\delta\) genes are found used in Chinese alligator TCR\(\delta\) transcripts, S' RACE PCR was performed. We sequenced and analyzed 296 cDNA clones (including 74 TCR\(\delta\), 20 TCR\(\delta\)3, 25 TCR\(\delta\)4, and 177 TCR\(\delta\)) with unique VJ junctions. Four different C\(\delta\) had different preferences for usage of V gene segments. V\(\delta\)3 was the most frequently used in TCR\(\delta\)1 chains, which accounted for roughly half of the expressed TCR\(\delta\) repertoire (45.95%) and only one TCR\(\delta\)1 clone (clone D2-2-6) contained VH\(\delta\)3 gene. In TCR\(\delta\)5 chains, V\(\alpha\)14/\(\delta\)7 and V\(\alpha\)21/\(\delta\)11 were preferentially used, accounting for 37.85 and 33.90%, respectively, and there were 18 (10.17%) TCR\(\delta\)5 cDNA clones containing VH\(\delta\)6 genes. But in all TCR\(\delta\)3 and TCR\(\delta\)4 transcripts obtained, the C\(\delta\)6 were only found coexpressed with VH\(\delta\)6 genes.

The four functional C\(\delta\)6 genes coexpressed with VH\(\delta\) genes detected in Chinese alligator encoded single V and C domains, similar to the structure of atypical TCR\(\alpha\) chain in X. tropicalis and birds (14, 31). To analyze the sequence diversity of the TCR\(\alpha\)/\(\delta\) transcripts with VH\(\delta\), we compared VDJ junctions in sequenced transcripts. A biased usage of VH\(\delta\) genes segments was observed in these clones (Fig. 7). Only a single TCR\(\alpha\)1 chain containing VH\(\delta\)3 gene segment was found. The 18 TCR\(\delta\)5 clones using VH\(\delta\)6 genes showed a preferential VH\(\delta\)6 usage with most clones using VH\(\delta\)6 and VH\(\delta\)6 (six clones and eight clones, respectively), with VH\(\delta\)10 (three clones) being the third most frequently used (Fig. 7). In the TCR\(\delta\)3 chains, which are only coexpressed with VH\(\delta\)6 genes segments VH\(\delta\)6.2 and VH\(\delta\)8 were found to be nearly equally used, whereas most TCR\(\delta\)4 clones, also only coexpressed with VH\(\delta\), preferentially used VH\(\delta\).3 (92%). Notably, all the VH\(\delta\)6 genes segments tended to be used by the nearest C\(\delta\).

From the sequence analyses, it appears that the C\(\delta\) transcripts use the D\(\delta\) gene segments and J\(\delta\) gene segments located within the same cluster as the C\(\delta\) (Figs. 7, 8). Most of the TCR\(\alpha\)1 and TCR\(\delta\)5 appear to use more than one D gene (VDDJ) because of the presence of the asymmetrical RSS leading to incorporation of
multiple D genes in recombination (32, 33). Therefore, the average length of CDR3 in unconventional TCR\(\alpha\) and TCR\(\beta\) was longer than that in TCR\(\beta\)3 and TCR\(\beta\)4 (13 ± 2.5 versus 7 ± 2.1 aa).

**Rare usage of VH\(\delta\) genes in TCR\(\alpha\) chains**

The use of V\(\alpha\) gene segments was also analyzed. A total of 313 cDNA fragments, including 123 TCR\(\alpha\)1 and 188 TCR\(\alpha\)2, all with unique V–J junctions, were obtained. We analyzed the usage of V and J gene segments and observed a biased usage of V\(\alpha\) and J\(\alpha\) in these clones. C\(\alpha\)1 preferred to use the V\(\alpha\) genes of locus A (75/123, 60.97%), and C\(\alpha\)2 had a biased usage pattern of V\(\alpha\) in locus B (147/188, 78.19%). Additionally, V\(\delta\) and VH\(\delta\) genes, albeit rare, were also found to be used in TCR\(\alpha\) chains.

**Discussion**

In this study, we systematically characterized the genomic organization, phylogeny, and expression of the TCR\(\alpha\)/\(\delta\) locus in Chinese alligator. The results provide a better and deeper understanding of the origin and evolution of VH\(\delta\) genes, as well as TCR\(\alpha\)/\(\delta\) locus as a whole in tetrapods.

*The Chinese alligator TCR\(\alpha\)/\(\delta\) locus exhibits a complicated genomic organization*

The genomic region containing TCR\(\alpha\)/\(\delta\) has been relatively conserved over the 400 million years of evolution of vertebrates because the genes flanking the TCR\(\alpha\)/\(\delta\) gene locus in coelacanth, which is considered to be phylogenetically mostly close to the ancestor of tetrapod, are the same as in *X. tropicalis*, alligators, birds, and mammals (Fig. 1) (5, 14, 21, 31). Chinese alligator TCR\(\alpha\)/\(\delta\) locus A was also flanked by the METTL3, SALL2, DAD1, and the ABHD4 genes, as seen in other tetrapods (6, 14, 34). However, flanking the locus B were KHNYN, CYLD, NFATC4, ADCY4, LTB4R, IGSF1, and FCRL5 genes, none of which could be found in the conventional TCR\(\alpha\)/\(\delta\) locus in amphibians, birds,
and mammals (Figs. 1, 2). These results are consistent with locus A being the ancient and locus B being derived from locus A.

The genomic structure of the two Chinese alligator TCRαδ/δ loci revealed that the genes encoding TCRδ were nested within the TCRα locus, like in other tetrapods (14, 31, 34). What was atypical was four tandemly repeated Dδ–J6–Cδ clusters were found to be present in locus A (Fig. 1). More surprising was the discovery of a large number of VH-related genes dispersed within both Chinese alligator TCRαδ/δ loci, which has not been found previously in any other tetrapod (Figs. 1, 5, 6). In total, there were 144 V (including 32 VHδ), 154 Jα, 9 Jδ, eight Dδ, 2 Cα, and 5 Cδ gene segments in the TCRαδ/δ loci of Chinese alligator (Fig. 1). The results are consistent with Chinese alligator having a few more Jδ segments and Dδ segments than other tetrapod species and that twice the number of Jα gene segments than in other vertebrates because of the presence of the additional locus (10, 14, 15, 31, 34–40). Furthermore, a large number of V gene segments were also found in the TCRαδ/δ loci of Chinese alligator, second only to that found in cattle (Table II) (32, 33).

In brief, although the organization of the Chinese alligator TCRαδ/δ locus has many features typical of tetrapods (6, 10, 14, 34, 35), it appears that the overall gene content and organization of the Chinese alligator TCRαδ/δ locus is the most complex among all tetrapod examined thus far (Fig. 2).

Transfer of VH genes into the TCRαδ/δ locus occurred frequently in vertebrates after the divergence of the IgH and TCR genes

The presence of TCR chains that use Ig-like V domains, such as atypical TCRδ using VHδ (termed as TAILV in the nurse shark), NAR-TCR, or TCRμ, have now been widely found in vertebrates, with the exception of the bony fish and eutherian mammals (10–12, 14–16, 21). NAR-TCR uses an N-terminal V domain related to IgNAR, or TCRμ, may have recombined with Vδ6 genes in the nurse shark evolution (16). The nurse shark TCRαδ/δ locus houses a complex of V segments from multiple lineages, including an IgH-like V segment coexpressed with TCRδ rearrangements. Phylogenetic analyses suggest the proximal IgH clusters could be the ancient precursors to TAILV that illustrate that IgH and TCRδ have freely exchanged gene segments over the course of the nurse shark evolution (16).

X. tropicalis possesses a great number of VHδ genes, all of which cluster with clan II VH gene segments (14). The VHδ5 genes in the TCRαδ/δ locus share a surprisingly high degree of nucleotide identity (91–96%) with VH5 genes in the Xenopus IgH locus, consistent with the VHδ5 genes being paralogous to VH5 genes in X. tropicalis (14). In Chinese alligator, we also identified 32 VHδ gene segments that can be segregated into 18 subgroups (Fig. 1, Table I). Based on nucleotide identity, four of them were found to be more similar to VHδ genes of Chinese alligator itself as supported by the phylogenetic analysis (Fig. 5). It is highly likely that these VHδ genes were translocated from the IgH locus into the TCRαδ/δ locus during the alligator speciation.

Therefore, a transfer of VHδ genes from the IgH locus to TCRαδ/δ locus has been occurring repeatedly after the divergence of the IgH and TCRδ genes in vertebrates (10, 14, 15). This transfer appears to be unidirectional, as thus far, no TCRVδ genes have been reported in the Ig loci (14).

Analysis of the Chinese alligator TCRαδ/δ locus reveals a possible evolutionary route of VHδ genes in tetrapod

Previous studies suggested that the translocation of VHδ genes from IgH to TCR locus appeared to have occurred independently in different species of tetrapods, as the VHδ genes in frogs, birds, and monotremes belonged to different VH clans (10, 14, 15). However, in this study, numerous VHδ genes were identified in Chinese alligators that were found to be related to three distinct VHδ gene clans (Figs. 1, 5). Most interestingly, as supported by the phylogenetic tree, all the avian VHδ and the platypus VHδ and Vμδ identified previously, are associated with distinct VH clans and could be perfectly matched to their phylogenetic counterparts of VHδ genes in the alligator (Fig. 6). This strongly suggested that all VHδ and Vμδ genes identified in birds and mammals had a common origin with the VHδ genes found in alligators. Based on this study and previous efforts by other teams, we propose a model of IgNAR may have recombined with V14–16, 21). NAR-TCR uses an N-terminal V domain related to

Table II. Numbers of V, D, J, and C gene segments in TCRαδ/δ loci of selected tetrapods

<table>
<thead>
<tr>
<th>Species</th>
<th>Vα/Vβ4</th>
<th>VHδ</th>
<th>Jα</th>
<th>Jδ</th>
<th>Dδ</th>
<th>Cα</th>
<th>Cδ</th>
<th>Totalb</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>57 (44)</td>
<td>0</td>
<td>61</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>127</td>
<td>(35)</td>
</tr>
<tr>
<td>Mouse</td>
<td>104 (28)</td>
<td>0</td>
<td>60</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>170</td>
<td>(35, 36)</td>
</tr>
<tr>
<td>Cattle</td>
<td>&gt;400 (&gt;44)</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>&gt;410</td>
<td>(37, 38)</td>
</tr>
<tr>
<td>Opossum</td>
<td>74 (47)</td>
<td>1</td>
<td>53</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>138</td>
<td>(34)</td>
</tr>
<tr>
<td>Platypus</td>
<td>99 (19)</td>
<td>2</td>
<td>32</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>144</td>
<td>(10)</td>
</tr>
<tr>
<td>Chicken</td>
<td>96 (3)</td>
<td>1</td>
<td>48</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>152</td>
<td>(31, 39)</td>
</tr>
<tr>
<td>Zebra finch</td>
<td>&gt;14 (7)</td>
<td>1</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>&gt;37</td>
<td>(15)</td>
</tr>
<tr>
<td>Duck</td>
<td>~79 (8)</td>
<td>1</td>
<td>68</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>155</td>
<td>(40)</td>
</tr>
<tr>
<td>X. tropicalis</td>
<td>57 (31)</td>
<td>14</td>
<td>5</td>
<td>77</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>159</td>
<td>(14, 15)</td>
</tr>
<tr>
<td>Chinese alligator</td>
<td>112 (34)</td>
<td>32 (18)</td>
<td>154</td>
<td>9</td>
<td>8</td>
<td>2</td>
<td>5</td>
<td>322</td>
<td>This study</td>
</tr>
</tbody>
</table>

All genes of TCRαδ/δ in Chinese alligator appear in bold.

4The Vα/Vβ segments include the Vα in either TCRα and/or TCRδ locus.

bNumbers of all elements in different species.

Numbers preceding parentheses are numbers of V gene segments, and numbers in parentheses are numbers of V subgroups.

Number not confirmed.

FIGURE 7. Usage frequency of VHδ and Jο genes in Chinese alligator. In horizontal axis, TCRδ1, TCRδ3, TCRδ4, and TCRδ5 indicate transcripts amplified from different Cδ, the number behind the VHδ means the number of families in Chinese alligator, and the number behind the Jο represents the number of allele. The vertical scale represents the number of the clone.
FIGURE 8. A model of the origin and evolution of VHδ gene in TCRα/δ loci of different tetrapod. Genes are color coded as indicated at the right: Vaδ, red; VHδ and Va, pink; Cδ and Cμ, dark green; and Ca, blue. In X. tropicalis, VH clan II gene or genes were translocated from the IgH locus and inserted into the TCRα/δ locus, then VHδ II genes expanded (14). In cartilaginous fish, V gene segments of the proximal IgH locus were translocated into the TCRδ locus as TAILVs (16). In the evolution of TCRα/δ locus in an ancestral teleost, IgHV-like segments lost because of an incomplete recombination of the V array to the other side of the D–J–C cluster (41). In coelacanth, the TCRα/δ locus contained VHδ genes (VH clan III) derived from IgH gene locus, most likely during the speciation of coelacanth (21). In an amniotic ancestor, VH clan I and VH clan II gene translocated and inserted into TCRα/δ locus. In the ancestor of archosaurs, duplication of Cδ I led to the presence of Cδ I and Cδ II. Then, VHδ III–Cδ III II duplicated to a new VHδ III–Cδ III cluster. In alligator, duplication of whole locus A generated locus B. Then V gene expanded and lost Cδ in TCRα/δ locus A of alligator, whereas in TCRα/δ locus B, additional tandem duplication of the VHδ III–Cδ cluster occurred. In birds, VHδ III genes and Cδ II and Cδ III genes were lost, and another Cδ I gene was duplicated in the ancestor of avian. But in the Passeriformes lineage, one Cδ I gene was lost, and the remaining Cδ I was duplicated after species divergence. In contrast, in the Galliformes lineage, the VHδ–Dδ–Jδ–Cδ cluster was translocated out of the TCRα/δ locus and currently resides on another chromosome (15). In platypus, VHδ I gene was lost, and the VHδ-Dδ-Jδ-Cδ cluster was translocated out of the TCRα/δ locus, giving rise to TCRμ; mammals retained the VHδ III genes (10, 12). The Metatheria lineage lost VHδ from the TCRα/δ locus, remaining a TCRμ locus only. In eutherian mammals, the TCRμ locus was also lost (10, 34).
to explain the evolution of TCRα/β genes in tetrapods, as described below.

In cartilaginous fish, the TCRδ locus is also nested within the TCRα locus and contains the TAILVs segments (IgHV-like) that are the result of a proximal IgH translocation into the TCRδ locus (16). Phylogenetic evidence supports that IgHV segments in proximal IgH clusters of TCRδ locus could be the ancient precursors to TAILVs (16).

In the evolution of TCRα/β locus in an ancestral teleost fish, the IgHV-like segments may have been lost because of an incomplete recombination of the V array to the opposite side of the D–J–C cluster (41). This rearrangement could have produced the unusual locus organization with Vαδ/β genes in an inverted transcriptional orientation 3′ of Cox (41). In coelacanth, which is closely related to the ancestor of terrestrial vertebrates, the TCRα/δ locus contained VHδ genes (VH clan III) derived from IgH gene locus most likely during the speciation of this lineage (21) (Fig. 8).

In the Xenopus lineage, a duplication of the D6–J6–C6 cluster occurred, resulting in the presence of two C6, each with its own set of Dδ and J6 segments (10, 14). Then, VH clan II gene or genes from the IgH locus were translocated into the TCRα/δ locus (Fig. 8) because Xenopus VδH is all clustered with its own VH II gene segments in the phylogenetic tree (14). Subsequently, the region of the VHδ II–Dδ–J6–C6 cluster was inverted and VHδ II genes expanded, leading to the number of VHδ genes present in frog (10, 14). In frogs, IgH and TCRα/δ loci are also closely linked, which may have facilitated translocation of VH into the TCRα/δ locus (14).

In a common ancestor of the amniotes, the VHδ I and VHδ III genes were already present in the TCRα/δ locus as a result of VH clan I and VH clan III gene translocation (Figs. 1, 8). Then, in the ancestor of birds, testudines, and crocodilians, duplication of C6 I led to the presence of C6 I and C6 II with its own set of Dδ and Jδ segments (Figs. 1, 4, 8). Subsequently, VHδ III–C6 II cluster also duplicated in the ancestor of birds, testudines, and crocodilians, generating a new VHδ III–C6 III cluster (Figs. 1, 4, 8).

After alligators diverged from archosaurs, locus B was duplicated from locus A in alligator, an event supported by phylogenetic analyses of the Cα genes and Cδ genes (Figs. 3, 4, 8). However, the two loci have undergone different evolutionary paths following duplication. In locus A, the VHδ III–C6 II cluster duplicated again, leading to presence of the fourth Cαδ gene (Figs. 1, 8). In locus B, a V gene expansion and loss of selected Cδ genes have shaped the present genomic structure, which contains 98 V gene segments and only one Cδ (Fig. 1). Additionally, VH clan II translocated into locus B, allowing a single VHδ gene located at the 3′ end of the locus B (Figs. 1, 8).

During the evolution of bird TCRα/δ locus, the VHδ III genes and Cδ II and Cδ III were lost, but the Cδ I gene was duplicated in the common ancestor of birds (Fig. 8). In the Galliformes lineage (such as chicken and turkey), the VHδ–Dδ–Jδ–Cδ cluster was translocated out of the TCRα/δ locus, creating an additional locus on another chromosome (15). This can explain why all avian VHδ grouped with VH clan I even though no VH clan I genes are found in the avian IgH locus today (15). While in the Passeriformes lineage (such as zebra finch), one ancestral Cδ I gene was lost, and the other remaining Cδ I was later duplicated again after zebra finch separated (Figs. 4, 8).

In contrast to birds, the common ancestor of mammals presumably lost the VHδ I genes, and then the VHδ–Dδ–Jδ–Cδ cluster was translocated out of the TCRα/δ locus, giving rise to TCRμ (10). Therefore, they retained the VHδ III genes (10, 12, 17). However, in the Metatheria lineage, including opossum, VHδ III was lost in the TCRα/δ locus, whereas the TCRμ locus is still maintained in their genomes (12). Later in evolution of eutherian mammals, the TCRμ locus was also lost (Fig. 8).

In summary, this study identified a distinct genomic organization of the TCRα/δ locus in Chinese alligator, which appears to be the most complicated TCRα/δ locus among all tetrapods examined so far. Our study further provides significant insights into the evolution of atypical TCRδ or TCRα/δ genes in tetrapods and the dynamic history of the TCRα/δ locus.

Acknowledgments

We thank Drs. Li Ma and Yonghe Ma for technical help in the BAC screening and genome walking.

Disclosures

The authors have no financial conflicts of interest.

References


