The Genome Resequencing of TCR Loci in *Gallus gallus*
Revealed Their Distinct Evolutionary Features in Avians

Fei Liu, Jing Li, Ivan Y. C. Lin, Xi Yang, Junping Ma, Yan Chen, Na Lv, Yi Shi, George F. Gao and Baoli Zhu

*ImmunoHorizons* 2020, 4 (1) 33-46
doi: https://doi.org/10.4049/immunohorizons.1900095
http://www.immunohorizons.org/content/4/1/33

This information is current as of March 28, 2021.

**Supplementary Material**
http://www.immunohorizons.org/content/suppl/2020/01/27/4.1.33.DCSupplementary

**References**
This article cites 78 articles, 28 of which you can access for free at:
http://www.immunohorizons.org/content/4/1/33.full#ref-list-1

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at:
http://www.immunohorizons.org/alerts
The Genome Resequencing of TCR Loci in *Gallus gallus* Revealed Their Distinct Evolutionary Features in Avians

Fei Liu,*†,‡ Jing Li,*†,‡ Ivan Y. C. Lin,*†,‡ Xi Yang,*† Junping Ma,*† Yan Chen,*† Na Lv,*† Yi Shi,* George F. Gao,* and Baoli Zhu*†,‡,§

*Chinese Academy of Sciences Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; †Beijing Key Laboratory of Microbial Drug Resistance and Resistome, Beijing 100101, China; ‡College of Life Sciences, Beijing Normal University, Beijing 100875, China; §Savai Medical School, University of Chinese Academy of Sciences, Beijing 100049, China; and *Department of Pathogenic Biology, School of Basic Medical Sciences, Southwest Medical University, Luzhou, Sichuan 646000, China

ABSTRACT

The TCR is consisted of four chains: α (TCRα), β (TCRβ), γ (TCRγ), and δ (TCRδ) that are present in all jawed vertebrates. Birds are very important in terms of evolutionary aspects of the adaptive immune system, in which it bridges the evolutionary gap between mammals and other vertebrates. To gain better understanding into the genomic organization and complexity of birds’ TCR loci, we applied cross-reference error-correction sequencing approach by using Illumina and single-molecule real-time sequencing technology to resequence genomic regions of chicken TCR loci based on 10 mapped bacterial artificial chromosome clones. We did de novo classification of V and J genes for all four chains of the TCR loci according to our sequencing results using the Immunogenetics nomenclature. In sum, we identified 85, 8, and 37 TCR V gene segments in the chicken TCRα/TCRδ, TCRβ, and TCRγ loci, respectively. The phylogenetic analysis showed the Vα 7 and Vα family 4 gene sequences shared greater sequence similarity with mammalian species, whereas the other Vα segment sequences are evolutionary closer with sequences from bony fishes. The organization of chicken TCRδ locus is more similar to fish TCRδ locus over mammalian species, as chicken TCRβ locus has a single translocon of its V–D–J–C and exhibits significantly fewer Vβ gene segments. In this study, we present a highly precise genomic map for chicken TCR loci and phylogenetic relationships of TCR variable gene segments against other animal species and verified the relative stability of the receptor structure during evolutional process. *ImmunoHorizons*, 2020, 4: 33–46.

INTRODUCTION

T cells are an essential element for the function of an adaptive immune response by mounting a specific cell-mediated immune response against foreign Ags (1). Each T cell expresses a unique TCR heterodimer that reacts with a specific antigenic peptide. Ligation of the TCR with antigenic peptide initiates a diverse array of developmental and functional immunological events, including T cell selection, differentiation, and cytokine production (2, 3), in which these delicate events are optimally tailored to confer appropriate responses to the dynamic array of infectious agent that the host might come across (4–6). Divergent pathways of T cell
development and maturation are characterized by the expression of either an αβ and a γδ TCR (7). The former lymphocytes express receptors comprised of α and β polypeptide chains, and the latter express receptors comprised of γ and δ polypeptides (8). The frequencies and distributions of αβ and γδ T cells differ among different species. In adult humans and mice, γδ T cells constitute <5% of the peripheral T cells (9). However, γδ T cells make up more than 20% of the peripheral T cells in rabbits and chickens (10, 11). The vast degree of Ag receptor diversity of these T cells are derived from recombination activating genes (RAG)–dependent somatic DNA recombination and by the CDR3 regions made of V(D)J junctions (3, 9).

Although the basic heterodimeric conformations are relatively conserved between αβ and γδ TCRs, there are several key differences in their genetic elements and the nature of recombination, leading to their ultimate diversity and functionality (3, 12). As with Abs, TCRα and TCRγ polypeptides are assembled from gene segments, V, J, and C, whereas a highly variable gene segment, D, is inserted between V and J gene segment in TCRβ and TCRδ, which enhances repertoire diversity (13, 14). Even though the number and organization of V, D, and J genes available for somatic recombination can vary in different animal species, the genes encoding TCRβ chains are embedded within TCRα at a single locus in all mammals, birds, and amphibians investigated to date (15–24). αβ T cells recognize antigenic peptide fragments in a way that is restricted by self-encoded molecules of the MHC, which are capable to present short peptide fragments derived from foreign Ag directly to T cells (3, 25). Unlike αβ T cells, γδ T cells are not restricted to the recognition of peptides presented by MHC complex (26), and a diverse array of ligands have been identified including nonpeptide Ags, metabolites, MHC and non-MHC membrane-bound molecules, soluble proteins and sulfatide, etc. (27–30). Furthermore, mounting evidence demonstrates that γδ T cells exhibit a profound cross-over protection of adaptive into innate immunity (9, 26, 31).

Chicken (Gallus gallus) is not only an economically important farm animal, but it is also an invaluable laboratory model for over 9600 extant avian species. In the evolution tree of life in the animal kingdom, birds shared many characteristics with mammals and other vertebrates like fishes, amphibians, and reptiles (32, 33). In chicken, the TCR gene organization and T cell differentiation and function, in some ways, resemble their mammalian counterparts (34–36), in which four TCR loci (α, β, γ, and δ) are identified in chicken genome, and RAG-dependent somatic DNA recombination is present (37). Although the complete genome of chicken (G. gallus) has been released in 2004 (33) along with some early works dedicated to uncover the chicken TCR loci (22, 36, 38–42), the organization and complexity of chicken TCR loci are still not fully elucidated, and the number and the order of TCR V, J, and C gene segments remain scattered.

In this study, we present a highly precise and accurate genome sequence and gene organization of all four chicken TCR loci. We combined the sequencing strength of second-generation sequencing and single-molecule real-time (SMRT) sequencing instruments to overcome the limitations of each sequencing system (43). We also employed nested PCR to sequence TCR repertoire that had undergone somatic DNA recombination and was expressed as spliced mRNA transcripts for the attempt of identifying the complete genome map representing the chicken TCR loci. Besides, the evolutionary relationship between chicken TCR loci and other animal species was also analyzed.

MATERIALS AND METHODS

Bacterial artificial chromosome clone information
Fifteen bacterial artificial chromosome (BAC) clones covering chicken TCRα, β, δ, and γ loci were purchased from red jungle fowl G. gallus BAC clone library (CHORI-262) from Children’s Hospital Oakland Research Institute (Oakland, CA). Clones 71M20, 73F9, 961F6, 64J27, 89G18, 61E3, 75B4, and 172L21 were mapped to TCRα/δ locus, clones 22I6, 63N4, 166K12, 33O1, and 167M2 covered TCRβ locus, and clones 174P24 and 35M9 covered TCRγ locus (BAC clone information is provided in Supplemental Table I). The clones were subjected to confirmation using both endonuclease digestion and PCR amplification. As there are overlapping parts between clones, we selected clones 71M20, 73F9, 64J27, 89G18, 61E3, 75B4, and 172L21 to cover TCRα/δ locus, clones 63N4 and 167M2 to cover TCRβ locus, and clone 174P24 to cover TCRγ locus.

Genomic DNA cloning and template preparation
The selected clones were purified and prepared for library construction using NEXTFLEX Rapid DNA-Seq Kit (Bioo Scientific, Austin, TX), according to manufacturer’s specification. For Illumina sequencing, the genomic DNA was fragmented by sonification to ~500–700 bp in length. For Pacific Biosciences (PacBio) sequencing, the genomic DNA was fragmented to ~10,000–20,000 bp in length.

DNA sequencing, sequence assembly, and correction
For Illumina sequencing, the DNA sequence analysis was outsourced to Novogene (Beijing China), using Illumina MiSeq PE250. Assembly of the high-throughput sequencing result was performed with ABySS (44) sequence assembly software. The raw data of high-throughput sequencing results are provided in Supplemental Table III. Concurrently, we also used PacBio RS II DNA sequencing system (PacBio, Menlo Park, CA) as the DNA sequencing platform. A 10-kb SMRT bell library was prepared from sheared genomic DNA using a 10-kb template library preparation workflow according to the manufacturer’s recommendation. The SMRT sequencing data reads were assembled with the Hierarchical Genome Assembly Process algorithm in SMRT Portal (version 2.2.0). Then, the assembled sequences were corrected with Illumina sequencing reads by using ICORN2 (45) (the corrected sequences are in Supplemental Table II). The chicken TCR loci were analyzed using the publicly available TCR mRNA repository at ImmunoGeneTics (IMGT) and the National Center for Biotechnology Information (NCBI) with the Nucleotide Basic Local Alignment Search Tool (BLAST) (identity >80%
and coverage >50%). The V and J segments were located by similarity to corresponding segments from previously identified chicken TCR segments and other species by identifying the flanking conserved recombination signal sequences (RSS), whereas D and C segments were located by similarity to corresponding segments from previously published chicken TCR segment sequences at IMGT and GenBank.

**Nomenclature**

Chicken TCR gene segments were annotated according to the IMGT nomenclature established for human and mouse. V segments were annotated according to their location from the 5′ to 3′ end of the locus, as V followed by the locus name (α, β, γ, or δ) and the family number and the gene segment number if there were >1 family. For example, VβL14 is the 14th Vβ gene in family 1. J segments were annotated 3′ to 5′ according to the nomenclature proposed by Koop et al. (21) and followed by IMGT.

**Phylogenetic analysis**

Nucleotide sequences that encode the V regions identified from each one of the chicken TCR loci were compared with TCR V sequences from other species retrieved from GenBank and IMGT. The sequences were aligned using MUSCLE (46) programs. The neighbor-joining trees were constructed using MEGA version 6.0 (47). Confidence values were obtained from bootstrap analyses using 1000 replications. GenBank accession numbers for sequences used in the tree construction are in Supplemental Table IV.

**RESULTS**

From the *Gallus Children’s Hospital Oakland Research Institute* BAC library, clones 71M20, 73F9, 64J27, 89G18, 61E3, 75B4, and 172L21 were sequenced to cover TCRα/δ locus, clones 63N4 and 167M2 were sequenced to cover TCRβ locus, and clone 174P24 was sequenced to cover TCRγ locus. We used the sequence result derived from Illumina sequencer to the correct PacBio sequence result (summary of raw reads from Illumina sequencer is in Supplemental Table III). DNA sequences representing the germline DNA sequence of *G. gallus* TCRα/δ, TCRβ, and TCRγ were obtained. A comparison of the existing chicken TCR loci (genome assembly Galgal4) against our hybrid error-correction sequencing results revealed a large amount of misassembled sequence fragments (shown in Fig. 1). There are two gaps in the draft genome that are not covered by BAC clones; both gaps are located in the TCRα/δ locus.

The chicken TCR α, β, and γ gene segments were classified using all publicly available TCR mRNA sequences at IMGT and NCBI. To precisely classify the chicken TCR gene segments, we also sequenced and analyzed the TCR repertoire derived from the peripheral whole blood of a White Leghorn chicken. The result could provide limited help for mapping those gene segments, as an uneven usage of V gene segment was observed when the repertoire sequence is cross-matched against our annotated TCR loci. Nevertheless, this finding might be explained by uneven preferential usage of V and J gene segments into somatic DNA recombination, as well as the negative selection of T cells during maturation before released into the peripheral blood. In sum, we identified 85, 8, and 37 TCR V gene segments in the chicken TCRα/TCRδ, TCRβ, and TCRγ loci, respectively. Gene segments were annotated according to the IMGT nomenclature established for human and mouse. V segments were annotated according to their location from the 5′ to 3′ end of the locus.

**TCRα/TCRδ locus is mapped to chromosome 27**

The chicken TCRδ locus is embedded within TCRα locus on chromosome 27, and usually, it is named as TCRα/TCRδ locus spanning ~800 kb. To investigate the degree of genomic conservation in this region, we attempted to identify the genes syntenic to chicken TCRα/TCRδ locus and compared these to the available...
genomic data for mouse, human, cow, platypus, and opossum using the whole-genome BLAST available NCBI (https://www.ncbi.nlm.nih.gov/) and the BLAST/BLAST-like alignment tool from ENSEMBL (http://www.ensembl.org). Flanking the chicken TRA/TRD locus were the genes encoded for a member of the wingless-type MMTV integration site (48), defender against cell death gene 1 (DADI) and interspersed with several truncated retroviral-derived elements throughout the locus, including sequences encoded for POL-like, RT-nLTR-like, and RT-ART (49–51). Although the chicken locus has many typical features common to TCRα/TCRδ locus found in other tetrapods, the 5′ end flanking genes, methyl-transferase–like 3, zinc finger protein, and interspersed olfactory receptor, are absent in chicken (15–17, 23, 52). The 3′ end of TCRα/TCRδ locus appears to be the most conserved synteny across tetrapods. The gene that encodes for DADI is also found at the 3′ end of the amphibian, mammal, marsupial, monotreme, and other birds. However, a well-conserved feature observed in mammals, a reverse oriented Vδ segment located immediately downstream of Cδ region, is absent in chicken TCRα/TCRδ locus.

Based on nucleotide identity of available chicken TCRα and TCRδ mRNA sequences, there are a total of 85 TCR V gene segments in the chicken TCRα/TCRδ locus, a number intermediate to that of human and mouse, 54 of which were identified as Vα and 31 that were identified as Vδ. The chicken Vα segments have an average length of ~340 ± 39 bp, whereas the mean length of Vδ is ~370 ± 31 bp (Supplemental Fig. 1). All V gene segments are followed by canonical 23-bp spacer RSS, which are the recognition substrate of the RAG recombinase complex (53).

In other animals, the V gene segments located in TCRα/TCRδ locus are generally designated as either Vα or Vδ, in other words, used in TCRα or TCRδ chains, respectively. In some cases, specific Vα or Vδ have been found expressed interchangeably between TCRα or TCRδ chains and have been designated Vα/δ. To determine which of the 84 V gene segments are expressed interchangeably, TCR repertoire derived from White Leghorn chicken peripheral blood were used to map the TCR expression pattern. Eight Vα segments and two Vδ segments were found expressed interchangeably, resulting in a total of 10 Vα/δ gene segments (shown in Fig. 2, Supplemental Fig. 2).

The number and complexity of D and J gene segments in the chicken TCRα/TCRδ locus is comparable to that of human and mouse. There are at least 67 Ja segments flanked by regions encoded for Cα and Cδ (Fig. 2), and all Ja gene segments are

---

**FIGURE 2.** Genome map of the chicken TCRα/TCRδ locus. TRAV (red), TRDV (yellow), TRAJ (light green), TRDJ (dark green), and TRDD (brown) gene segments are shown numbered by their corresponding location in order across the locus. TRAC (light blue) and TRDC (dark blue) are also indicated. TRAV segments are designated with a subgroup number followed by a period and a designated number according to their genomic location. Transcriptional orientation is indicated by the direction of the arrow on each gene segment. Arrows are not in proportion to the actual gene sizes. Interchangeably used V gene segments are indicated with a φ. Broken lines indicate the sequence gaps in the study, which is supplemented by the currently available Gallus gallus genome assembly (Galgal 4) accordingly. Syntenic and some unrelated genes are presented as gray boxes.
flanked with canonical 12-bp spacer RSS at the 5’ end. The large number of Jα genes is consistent with the postulation that the large number of Jα may be required during secondary Vα to Jα rearrangements in developing αβ T cells if the primary rearrangements are nonproductive or in need for replacement (54). Primary TCRα rearrangements typically begin at the 5’ end of the Jα cluster and can progress downstream toward the 3’ end Jαs during thymocyte maturation (55). Within TCRβ locus, there are at least two D and two J gene segments, all of which are located downstream of Vβ cluster and upstream of a single Cβ. The Dβ gene segments were asymmetrically flanked by an RSS containing a 12-bp spacer on the 5’ end and 23-bp spacer on the 3’ end side, as has been revealed previously for TCR Dδ gene segments in other species (16, 23, 52, 56).

During the course of evolution, V gene segments are either duplicated or deleted, resulting in a certain degree of relatedness among segments (57). These V gene segments are therefore defined as subgroups with the segments belonging to the same subgroup by sharing 75% or greater nucleotide identity (58). By this criterion, the identified 54 Vα segments are subdivided into 10 families. The 31 Vβ segments are placed into four families. The TCRα/TCRβ locus map revealed an interesting gene duplication scenario, as the V gene segments on TCRα/TCRβ locus might be duplicated in a pair. Of all 14 Vα family 4 segments, all of them are accompanied by another Vα segment located immediately upstream throughout the entire TRAV section of the locus. Furthermore, although Vα 7 did not fall into Vα family 4, phylogenetic analysis (Fig. 3) showed close sequence similarity between Vα 7 and Vα family 4 gene segments.

We then performed phylogenetic analysis using the chicken Vα or Vβ sequences to evaluate their evolutionary relatedness. In the analysis, the representative TCRV gene sequences found in amphibian, bird, fish, and mammal (including marsupial and monotreme) were retrieved from freely available public repositories, including IMGT and NCBI. The platypus Vα and Vβ genes were retrieved from the publication written by Parra et al. (23).

The phylogenetic analysis showed eight major groups of V gene sequences (designated I through VIII) from the chosen Vα sequences. The chicken Vα gene sequences clustered together into two groups (group I and VIII). However, there are a small number of gene sequences from other animal groups clustered in close relationships with chicken Vα genes in group I, which included mouse 895, cattle 636, human 261, and rhesus monkey 866 Vα gene sequences. As aforementioned, Vα 7 and Vα family 4 gene segments are closely related and clustered apart from the other Vα gene segments. As shown, Vα 7 and Vα family 4 gene sequences shared greater sequence similarity with mammalian species, whereas the other Vα segment sequences are phylogenetically closer with sequences from bony fishes. Therefore, it is possible that both Vα gene types developed before the divergence of bony fish during evolution.

There were five phylogenetic clusters from the Vβ gene sequences (Fig. 4). The chicken Vβ sequences only clustered within one group and share sequence similarity with zebra finch TRDV gene cluster (group II). An interesting finding was revealed in both TRDV and TRAV phylogenetic analysis, in which the amphibian and cartilaginous fish TCR V gene sequences shared a significant similarity and clustered fairly close to each other, group IV and group III in Vα and Vβ phylogenetic tree, respectively. The chicken Vα and Cβ are encoded by four and three exons, respectively (Supplemental Fig. 3). Two C region genes were present: a Cα that is located at the most 3’ end coding segment in the locus, and a Cβ-orientation 5’ of the Jα genes. Collectively, all content and organization of the chicken TCRα/TCRβ locus appeared fairly generic to other animal species.

**TCRβ locus is mapped to chromosome 1**

The chicken TCRβ locus is physically mapped to chromosome 1 and found to span a region of ~120 kb, making it significantly smaller in size than that of human, mouse, and opossum (15, 59, 60). The chicken TCRβ is flanked by genes encoding for Nedd4 family interacting protein 2 at the 5’ end and interspersed with truncated retroviral-derived elements in the locus. Based on the nucleotide identity of chicken TCRβ mRNA sequences, we identified eight Vβ segments, one Dβ segment, four Jβ segments, and one Cβ segment (Fig. 5).

The chicken TCRβ locus has a translocon organization (with the exception of a reverse-oriented Vb gene segment that is discussed later), with multiple Vβ segments followed by Dβ, Jβ, and Cβ encoding region. The human and mouse TRB loci consist of two such D–J–C cassettes, whereas four of such cassettes are found in opossum (60, 61).

Of the total of eight Vβ gene segments, two different Vβ families were categorized, with each family consisting of four Vβ gene segments. The number of TCRBV genes in chicken is significantly less compared with that of human with up to 67 Vβ (62), 22 Vβ in mouse, and 35 Vβ in the opossum (15, 59). A conserved decanucleotide (decamer) sequence found within the promoter region of mammalian TCR Vβ regions was also found in four of the chicken Vβ gene segments that belong to Vβ1 family (Fig. 6) (63, 64).

It has been reported that Vβ decamer is required for high-level expression of murine Vβ gene segments, which serves as the binding site of a transcription activator (63). Similar to mammalian and some bony vertebrates, each chicken Vβ segment consists of two exons that are spliced posttranscriptionally. The Vβ segments share a conserved leader exon of 45-bp coding sequence, followed by the splice donor sequence, GT, with an intron length of 412 bp for all Vβ1 genes and an average length of 88 ± 4 bp for Vβ2 genes (Supplemental Fig. 1). Exon two of the Vβ segment is preceded by the splice acceptor sequence, AG, and is 296 bp in length for Vβ1 genes and 287 ± 4 bp in length for Vβ2 gene segments. All Vβ segments are followed by canonical 23-bp spacer RSS, typical of Vβ genes. Of the total eight Vβ gene segments, most are oriented 5’ to 3’. The marked exception is a Vβ segment (Vβ2.4) 5.2 kb downstream of Cβ in the reverse orientation relative to the other gene segments.

A single Dβ and four Jβ gene segments are found in the TRB locus. The Dβ segment is flanked by asymmetrical RSS, as described in the TCRα/TCRβ locus. The segment is located between the Vβ and Jβ clusters, downstream of Vβ2.3 and upstream of the 5’-most Jβ gene, Jβ4. The Jβ segments have an average length of 50 ± 2 bp. The Jβ segments have 12-bp spacer RSS at the 5’ end and cartilaginous fish TCR V gene sequences shared a significant similarity and clustered fairly close to each other, group IV and group III in Vα and Vβ phylogenetic tree, respectively. The chicken Vα and Cβ are encoded by four and three exons, respectively (Supplemental Fig. 3). Two C region genes were present: a Cα that is located at the most 3’ end coding segment in the locus, and a Cβ-orientation 5’ of the Jα genes. Collectively, all content and organization of the chicken TCRα/TCRβ locus appeared fairly generic to other animal species.

**TCRβ locus is mapped to chromosome 1**

The chicken TCRβ locus is physically mapped to chromosome 1 and found to span a region of ~120 kb, making it significantly smaller in size than that of human, mouse, and opossum (15, 59, 60). The chicken TCRβ is flanked by genes encoding for Nedd4 family interacting protein 2 at the 5’ end and interspersed with truncated retroviral-derived elements in the locus. Based on the nucleotide identity of chicken TCRβ mRNA sequences, we identified eight Vβ segments, one Dβ segment, four Jβ segments, and one Cβ segment (Fig. 5).

The chicken TCRβ locus has a translocon organization (with the exception of a reverse-oriented Vb gene segment that is discussed later), with multiple Vβ segments followed by Dβ, Jβ, and Cβ encoding region. The human and mouse TRB loci consist of two such D–J–C cassettes, whereas four of such cassettes are found in opossum (60, 61).

Of the total of eight Vβ gene segments, two different Vβ families were categorized, with each family consisting of four Vβ gene segments. The number of TCRBV genes in chicken is significantly less compared with that of human with up to 67 Vβ (62), 22 Vβ in mouse, and 35 Vβ in the opossum (15, 59). A conserved decanucleotide (decamer) sequence found within the promoter region of mammalian TCR Vβ regions was also found in four of the chicken Vβ gene segments that belong to Vβ1 family (Fig. 6) (63, 64).

It has been reported that Vβ decamer is required for high-level expression of murine Vβ gene segments, which serves as the binding site of a transcription activator (63). Similar to mammalian and some bony vertebrates, each chicken Vβ segment consists of two exons that are spliced posttranscriptionally. The Vβ segments share a conserved leader exon of 45-bp coding sequence, followed by the splice donor sequence, GT, with an intron length of 412 bp for all Vβ1 genes and an average length of 88 ± 4 bp for Vβ2 genes (Supplemental Fig. 1). Exon two of the Vβ segment is preceded by the splice acceptor sequence, AG, and is 296 bp in length for Vβ1 genes and 287 ± 4 bp in length for Vβ2 gene segments. All Vβ segments are followed by canonical 23-bp spacer RSS, typical of Vβ genes. Of the total eight Vβ gene segments, most are oriented 5’ to 3’. The marked exception is a Vβ segment (Vβ2.4) 5.2 kb downstream of Cβ in the reverse orientation relative to the other gene segments.

A single Dβ and four Jβ gene segments are found in the TRB locus. The Dβ segment is flanked by asymmetrical RSS, as described in the TCRα/TCRβ locus. The segment is located between the Vβ and Jβ clusters, downstream of Vβ2.3 and upstream of the 5’-most Jβ gene, Jβ4. The Jβ segments have an average length of 50 ± 2 bp. The Jβ segments have 12-bp spacer RSS at the 5’ end.
FIGURE 3. Phylogenetic tree of 54 chicken Vᵦ and 61 Vᵦ gene sequences from 2 cartilaginous fish, 6 bony fish, 2 amphibian, 15 mammalian, and 1 avian species. The eight major phylogenetic clusters are bracketed. Chicken Vᵦ sequences are indicated with the gene number assigned previously. The three-digit number following the Vᵦ animal names are the last three digits of the GenBank accession number referenced in Supplemental Table III.

https://doi.org/10.4049/immunohorizons.1900095
followed by the splice donor sequence, GT, at the 3′ end. All four Jβ segments are closely spaced within 10 kb. There is only a single chicken Cβ region in contrast to two in human and mouse. The chicken Cβ is encoded by four exons.

Phylogenetic analysis revealed that the chicken Vβ gene sequences can be clustered into two groups, in which the chicken Vβ1 genes share closer sequence similarity to fishes, whereas Vβ2 genes have closer sequence similarity to mammalian species (Fig. 7). Such a result might suggest the establishment of both Vβ1- and Vβ2-type TRBV gene lineages before the divergence of reptile and bird because both chicken Vβ clusters have similar phylogenetic distances in relation to amphibian species.

**TCRγ locus is mapped to chromosome 2**

The chicken TCRγ locus has been mapped to chromosome 2. As the 5′-most Vγ gene segment to the 3′-most single Cγ region, the chicken TRG locus spans only ~82 kb (Fig. 8), significantly smaller than that in human (150 kb), mouse (205 kb), and dog (360 kb) (15, 65, 66). DNA-dependent protein kinase catalytic subunit and leucine-rich repeat flightless-interacting protein 2 are found immediately flanking the TRG locus.

Different from human, mouse, and dog, the chicken TCRγ locus consists of a single Vγ, Jγ, and Cγ gene cassette. The human Vγ segments are upstream of two separate sets of Jγ–Cγ gene segments, whereas mouse and dog have four and eight tandemly organized V–J–C cassettes, respectively (65–67).

There are 37 Vγ gene segments present in the chicken TCRγ locus, and these are divided into 11 subgroups based on nucleotide similarity (Fig. 9), but only 15 of the Vγ segments have been found expressed in the peripheral blood (Supplemental Fig. 2). In contrast, other loci described above the TRGV gene segments have a similar mean length of 360 bp but possess a significantly greater SD, 82 bp (Supplemental Fig. 1). Such findings may be direct evidence indicating that ligand targets of the chicken γδ TCR are
not exclusive to MHC-processed peptide complex (68), allowing the recognition of architecturally and chemically more diverse Ags, unlike that seen with αβ TCR recognition of MHC-processed peptide (69, 70). The number of Vγ segments is significantly greater to that in eutherians, in which 7 Vγ gene segments are identified in mouse, 16 Vγ gene segments are identified in dog, and 6 are found in human (3, 59, 66). An apparent expansion of Vγ gene segments in chicken TCR is therefore evident, which might be an indication of extensive usage of γδ TCR–lymphocyte in chicken. Phylogenetic analysis revealed seven major phylogenetic clusters using the chosen V gene sequences; the chicken Vγ segments are clustered within group I, group II, and group IV. Group I and group II are phylogenetically closer to mammalian Vγ sequences, whereas group IV has a greater sequence similarity toward bony fish and amphibian Vγ gene sequences. It was also noteworthy that mallard, another avian species, also presented in group I and group IV, indicate that both Vγ lineages were established before the divergence of bony fish and passed on through the bird and reptile evolutionary branch.

Three Jγ gene segments are identified in the TRG locus. They are closely spaced within 3.4 kb and have an average length of $53 \pm 1$ bp. Consistently, the Jγ gene segments have 12-bp spacer RSS at the 5′ end and are followed by the splice donor sequence. The Jγ cluster is positioned between the Vγ cluster, and the region encodes for Cγ. There is only a single chicken Cγ region in contrast to four in mouse and two in human. The chicken TRGC is encoded by three exons.

DISCUSSION

Chickens are arguably the most extensively studied avian species, in which the draft genome sequence of the red jungle fowl, G. gallus, was published in 2004 (33). It is not only one of the world’s most common sources of dietary protein (71), it also serves as a model species that bridges the evolutionary gap between mammals and other tetrapods, making it indispensable in evolutionary studies (33). Furthermore, chickens are an essential intermediate

![FIGURE 5. Annotated map of the chicken TCRβ locus showing the location of the Vβ (red), Dβ (orange), Jβ (green), and Cβ (blue). Conserved syntenic genes are indicated in gray box.](https://doi.org/10.4049/immunohorizons.1900095)

![FIGURE 6. Genomic organization of the chicken TCRβ Vβ gene segments. The decamer (purple), Vβ coding region (light blue), and RSS (gray) are indicated. Genomic splice donors (GT) and acceptors (AG) are indicated, and intron sizes are indicated. Vβ2.4 is found approximately 5.2 kb downstream of the Cβ on the complementary strand in reverse orientation.](https://doi.org/10.4049/immunohorizons.1900095)
FIGURE 7. Phylogenetic tree of eight chicken Vβ and 54 Vβ gene sequences from 2 cartilaginous fish, 2 bony fish, 2 amphibian, 15 mammalian, and 1 avian species.

The eight major phylogenetic clusters are bracketed. Chicken Vβ sequences are indicated with the gene number assigned previously. The three-digit number following the Vβ animal names are the last three digits of the GenBank number referenced in Supplemental Table III.

https://doi.org/10.4049/immunohorizons.1900095
host and a reservoir of many infectious agents, such as influenza virus (72). Therefore, in-depth characterization of the immune system in the chicken, particularly T cell immunology, is important for better understanding of the immunological interplay between chicken and these diseases. Complete characterization of chicken TCR genomics is a milestone to achieve this final goal.

In this study, we present the foremost accurate genomic content and complexity of the TCR loci for chicken G. gallus as well as the most comprehensive TCR loci description for Galliformes. The current chicken genome assembly, Galgal 4, possesses a considerable amount of sequence gaps due to the limitation of sequencing high-density repeat sequences and the limitation of Illumina sequencers. We combined the strength of high accuracy and long read through the use of second and third generation sequencing technologies, respectively. The sequencing strategy of this hybrid approach provides high-quality assemblies with fewer errors and gaps, which enable more accurate downstream analyses.

The TCRα/TCRδ locus is significantly smaller in size compared with that of human at 1 Mb and mouse at 1.65 Mb (59, 60) but substantially greater than its ortholog in the zebra finch at 150 kb (17) and similar to that of rabbit at 876 kb (73). Despite the fact that the TCRα/TCRδ locus was incomplete because of the presence of inevitable BAC clone gaps, the gaps were complemented using the current Galgal 4 draft genome assembly. The organization and level of complexity are similar to that of their mammalian counterparts. However, DADI was the only syntenic gene present in chicken TCRα/TCRδ locus, and an evidence-based observation suggests that some of the chicken Vα genes were duplicated in pairs (Supplemental Fig. 4). The total number of chicken Vα/δ may be underestimated as it is possible that some Vα/δ may be infrequently expressed or some combinations are negatively selected in the thymus before reaching the periphery. Collectively, the level of complexity for Vα and Vδ genes in chicken still shows considerable similarity to its evolutionarily distant mammalian cousins (15, 59). It should be noted that annotations of V segments as selectively Va or Vd are challenging because it is difficult to know whether they may be infrequently used with the reciprocal C gene segment. Our findings might suggest that bony fishes, an evolutionary intermediate stage between cartilaginous fish and amphibian, have developed their own sets of Vα and Vδ gene sequences to acclimatizing into their own niche after divergence. We also analyzed the phylogenetic relationship by combining both Vα and Vδ gene sequences (data not shown), and the phylogenetic tree shows that the Vα gene sequences from various jawed-vertebrate species forms clusters separate from the Vδ gene sequences. It is possible that the Vα and Vδ gene sequences were evolutionarily diverged before the divergence of fishes and amphibia because fish Vα genes do not appear to be at the basal level of the tree. This is somewhat consistent with the fact that both αβ and γδ T cell lineages are also found in cartilaginous fishes (74, 75), suggesting that TRAV and TRDV lineages were established before the divergence of cartilaginous fishes and bony vertebrates. Because of the limited data, the characterization of the second TCRδ locus was not performed.

The organization of chicken TCRβ locus is more similar to fish TCRβ locus over mammalian species as chicken TCRβ locus has a single translocon organization of its V–D–J–C, unlike the multiple tandem cassettes of D–J–C found in mammalian species. The locus exhibited significant fewer Vβ gene segments, suggesting decreased usage of αβ T lymphocytes can have a direct impact on gene duplication during the course of evolution. Phylogenetic analysis of the Vβ sequences reveals the coexistence of both Vβ families before the divergence of birds and reptiles. The reverse-oriented Vβ gene was also found in other animal species, indicating it is an ancient arrangement. The chicken TCRβ locus organization is highly conserved among eutherians. A clear ortholog of this gene is present in human, mouse, opossum, and zebrafish, making this an ancient arrangement (15, 60, 76, 77). Despite this peculiar organization, such inverted Vβ cassettes observed in human, mouse, and zebrafish have been shown to undergo somatic rearrangement by chromosomal inversion (61, 77, 78). Collectively, the organization and complexity of the chicken TCR genes are indicated in the gray box.

**FIGURE 8.** Annotated map of the chicken TRG locus showing the location of the Vγ (red), Jγ (green), and Cγ (blue).
TRB locus are significantly simpler compared with their mammalian counterparts, especially the number of Vβ genes and the single translocon organization. It is possible that because the αβ TCR<sup>lowγδ</sup> TCR<sup>highγδ</sup> nature of chicken, the chicken TCRβ gene locus has not been as extensively used as in mammalian species (αβ TCR<sup>highγδ</sup> TCR<sup>lowγδ</sup>), leading to limited gene duplication (35, 57).
Chicken TCRγ locus is noticeably smaller than its mammalian counterpart but possesses a more diversified array of Vγ gene segments. The number of chicken Vγ gene segments is more than doubled in some mammalian species, suggesting a positive relationship between extensive usage (μ6 TCRlow/γδ TCRhigh animal) of the locus and the incidence of gene duplication. The relative expansion of chicken Vγ gene segments might be relevant to more γδ T cells of the peripheral T cells in chickens. Noteworthy, the average length of Vγ gene segments has the greatest variability among V gene segments from the other loci, which might be indirect evidence showing that the ligands of γδ T lymphocytes are not restricted to MHC-processed peptides, providing a versatile Ag recognition system for architecturally and chemically more diversified TCR ligands.

Altogether, at genome organization and complexity levels of chicken TCR loci, our data confirm the earlier observations that the chicken TCR are developmentally similar to the mammalian TCR. Therefore, the TCR Ag recognition system appears to have predated the evolutionary divergence of reptiles and mammals, and the pre-existing structure of the receptor has been generally unchanged during the estimated 250–300 million years of divergent evolution.

DISCLOSURES

The authors have no financial conflicts of interest.

REFERENCES


https://doi.org/10.4049/immunohorizons.1900095

Downloaded from http://www.immunohorizons.org/ on March 28, 2021


Supplemental files

**Supplemental Table 1: The red jungle fowl Gallus gallus TCR gene loci’s BAC clone information.**

<table>
<thead>
<tr>
<th>TCR loci</th>
<th>Chromosome</th>
<th>Chromosomal location of the TCR locus (Kb)</th>
<th>BAC clone</th>
<th>Clone length (Kb)</th>
<th>Chromosomal location (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCRα/δ</td>
<td>27</td>
<td>0~1000</td>
<td>71M20</td>
<td>120</td>
<td>23~142</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>73F9</td>
<td>193</td>
<td>179~372</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>96I6</td>
<td>211</td>
<td>191~402</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>64J24</td>
<td>188</td>
<td>347~535</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>89G18</td>
<td>198</td>
<td>433~631</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>61E3</td>
<td>276</td>
<td>648~927</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75B4</td>
<td>116</td>
<td>877~993</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>172L21</td>
<td>181</td>
<td>943~1127</td>
</tr>
<tr>
<td>TCRβ</td>
<td>1</td>
<td>81180~81370</td>
<td>22I9</td>
<td>201</td>
<td>81166~81367</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>63N4</td>
<td>269</td>
<td>81320~81588</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>166K12</td>
<td>182</td>
<td>81276~81457</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>33O1</td>
<td>197</td>
<td>81374~81571</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>167M2</td>
<td>197</td>
<td>81166~81364</td>
</tr>
<tr>
<td>TCRγ</td>
<td>2</td>
<td>49150~49230</td>
<td>174P24</td>
<td>205</td>
<td>49109~49314</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35M9</td>
<td>229</td>
<td>49132~49361</td>
</tr>
</tbody>
</table>
Supplemental Table 2: The complete sequences of BAC clones using cross-reference error-correction sequencing approach.
### Supplemental Table 3: Raw data of the chicken TCR loci BAC library’s high-throughput sequencing result.

<table>
<thead>
<tr>
<th>Clone name</th>
<th>Read length</th>
<th>Number of read pairs</th>
<th>Raw data (Mb)</th>
<th>Read quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Raw</td>
<td>High quality</td>
<td>Beside vector</td>
</tr>
<tr>
<td>71M20</td>
<td>250</td>
<td>686806</td>
<td>390587</td>
<td>372660</td>
</tr>
<tr>
<td>73F9</td>
<td>250</td>
<td>677107</td>
<td>431183</td>
<td>408733</td>
</tr>
<tr>
<td>64J24</td>
<td>250</td>
<td>358339</td>
<td>204122</td>
<td>194527</td>
</tr>
<tr>
<td>89G18</td>
<td>250</td>
<td>435819</td>
<td>268341</td>
<td>255902</td>
</tr>
<tr>
<td>61E3</td>
<td>250</td>
<td>325733</td>
<td>120277</td>
<td>114914</td>
</tr>
<tr>
<td>75B4</td>
<td>250</td>
<td>909424</td>
<td>891360</td>
<td>853733</td>
</tr>
<tr>
<td>172L21</td>
<td>250</td>
<td>201328</td>
<td>211561</td>
<td>203393</td>
</tr>
<tr>
<td>167M2</td>
<td>250</td>
<td>268903</td>
<td>167888</td>
<td>159380</td>
</tr>
<tr>
<td>63N4</td>
<td>250</td>
<td>235069</td>
<td>149425</td>
<td>143946</td>
</tr>
<tr>
<td>174P24</td>
<td>250</td>
<td>473548</td>
<td>315551</td>
<td>302066</td>
</tr>
</tbody>
</table>
Supplemental Table 4: GenBank accession numbers for phylogenetic tree.
Supplemental Figure 1: Statistical illustration of the length of chicken TCR V gene segments.
Supplemental Figure 2: T cell receptor repertoire derived from peripheral whole blood of a white leghorns chicken. The expressed genes segments (dark) are shown by their corresponding location in order across the locus.
Supplemental Figure 3: exon and intron organization of the chicken TCR genes. Indicated in white circles are the donor splice site (GT) and in black the acceptor splice sites (AG).
Supplemental Figure 4: An illustration of pair duplication scenario of Vα gene segments on TCRα/TCRδ locus. A colour key of the gene segments is provided at the bottom.