Platypus TCRμ Provides Insight into the Origins and Evolution of a Uniquely Mammalian TCR Locus

Xinxin Wang, Zuly E. Parra and Robert D. Miller

*J Immunol* 2011; 187:5246-5254; Prepublished online 5 October 2011;
doi: 10.4049/jimmunol.1101113

http://www.jimmunol.org/content/187/10/5246

Supplementary Material

http://www.jimmunol.org/content/suppl/2011/10/05/jimmunol.1101113.DC1

References

This article cites 32 articles, 9 of which you can access for free at:
http://www.jimmunol.org/content/187/10/5246.full#ref-list-1

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
Platypus TCR\(\mu\) Provides Insight into the Origins and Evolution of a Uniquely Mammalian TCR Locus

Xinxin Wang, Zuly E. Parra, and Robert D. Miller

TCR\(\mu\) is an unconventional TCR that was first discovered in marsupials and appears to be absent from placental mammals and nonmammals. In this study, we show that TCR\(\mu\) is also present in the duckbill platypus, an egg-laying monotreme, consistent with TCR\(\mu\) being ancient and present in the last common ancestor of all extant mammals. As in marsupials, platypus TCR\(\mu\) is expressed in a form containing double V domains. These V domains more closely resemble Ab V than that of conventional TCR. Platypus TCR\(\mu\) differs from its marsupial homolog by requiring two rounds of somatic DNA recombination to assemble both V exons and has a genomic organization resembling the likely ancestral form of the receptor genes. These results demonstrate that the ancestors of placental mammals would have had TCR\(\mu\) but it has been lost from this lineage. *The Journal of Immunology*, 2011, 187: 5246–5254.

Conventional T cells exist in two distinct lineages based on the composition of their TCR heteroduplex: \(\alpha\beta\) T cells use a TCR composed of \(\alpha\) and \(\beta\)-chains, whereas \(\gamma\delta\) T cells use \(\gamma\) and \(\delta\)-chains. Like Ig, the Ag-binding V domains of the TCR chains are encoded by exons that are assembled from gene segments by somatic DNA recombination. All jawed vertebrates have both \(\alpha\beta\) and \(\gamma\delta\) T cells, and the genes encoding these four TCR chains are highly conserved both in sequence and organization (1–3). Recently, a fifth locus encoding TCR chains, named TCR\(\mu\), was found in marsupial mammals (4). TCR\(\mu\) contains C regions related to TR\(\delta\) but is transcribed in a form that would include double V domains that are more related to Ig H chain V region (VH) than to TCR V genes (2, 4, 5). TCR\(\mu\) does not substitute for TCR\(\delta\) in marsupials because the genes encoding conventional TCR\(\delta\)-chains are highly conserved and expressed (2, 6).

TCR\(\mu\) genes are distinct and unlinked to those that encode conventional TCR chains and have atypical gene organization. The N-terminal V of TCR\(\mu\) (V\(\mu\)) is encoded by somatically recombined genes (V, D, and J), with the recombination taking place in thymocytes, resulting in clonal diversity (4). The second, C-proximal V domain (V\(\mu\)C) is encoded by an exon in which the V, D, and J genes are already prejoined in the germline DNA and are relatively invariant (4). This is the only known example of germline-joined V genes being used in a TCR. The TCR\(\mu\) locus is also organized in tandem clusters, which is also atypical of TCR genes (2, 4).

Searching the available placental mammal, avian, and amphibian genomes failed to uncover TCR\(\mu\) orthologs (2). However, in this study, we show that TCR\(\mu\) is present in a monotreme, the duckbill platypus *Ornithorhynchus anatinus*. The monotremes are oviparous mammals that last shared a common ancestor with marsupials and placentals at least 165 million years ago (MYA) (7). The genomic organization of the platypus TCR\(\mu\) locus reveals insight into the evolution of this uniquely mammalian TCR locus and supports its ancient presence in mammals.

Materials and Methods

Whole genome analysis and annotation

Analyses were performed using the platypus genome assembly version 5.0.1 available at GenBank (http://www.ncbi.nlm.nih.gov/genome/guide/platypus/). Marsupial C\(\mu\) sequences were used to search based on homology using the BLAST algorithm (4, 5, 8). Scaffolds containing C\(\mu\) sequences were retrieved, and exon boundaries were determined by the presence of canonical mRNA splice sites. Platypus cDNA sequences were used to search against the *O. anatinus* genome project to identify the genomic V, D, and J gene segments. The beginning and end of each coding exon of V, D, and J gene segments were identified by the presence of mRNA splice sites or flanking recombination signal sequences (RSS). Supplemental Fig. 1 shows the location of each TCR\(\mu\) V, D, J, and C segments on the scaffolds. Platypus TCR\(\delta\)-chain C region sequence (GenBank accession number XM_001516959) was used to identify the single-copy platypus C\(\delta\) on scaffold 588, which is separate from any of the scaffolds containing the putative platypus TCR\(\mu\) sequences.

PCR and cDNA analyses

A spleen cDNA library constructed from tissue from a Tasmanian platypus was screened by PCR (9). All PCR primer sequences used in this study are presented in Table I. PCR amplification was performed using Advantage-2 HF 2 PCR (BD Biosciences, Clontech Laboratories, Palo Alto, CA) with the following conditions: denaturation at 94°C for 1 min for 1 cycle, followed by 34 cycles of 94°C for 30 s, annealing/extension at 62°C for 4 min, and a final extension period of 68°C for 5 min. Forward and reverse primers complementary to sequence internal to the platypus C\(\mu\) exon were paired with primers in the Agt10 vector used to construct the library to amplify clones containing the 5′ and 3′ untranslated regions (UTR) (10). This approach generated the partial cDNA sequences analyzed. Full-length platypus TCR\(\mu\) cDNA sequences were isolated by PCR using primers complementary to 5′ and 3′ UTR. PCR products were cloned using TOPO TA cloning Kit (Invitrogen, Carlsbad, CA) and sequenced using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). The GenBank accession numbers of the cDNA sequences described in this study are: clone 21, GU458338; clone 26, GU458339; clone 2.22, GU458341; clone 3815, GU475137; clone 1951, GU475138; clone 1953,

Abbreviations used in this article: CP, connecting peptide; CT, cytoplasmic; FR, framework region; L, leader; MYA, million years ago; NAR, new Ag receptor; RSS, recombination signal sequences; TM, transmembrane; UTR, untranslated regions; VH, Ig H chain V region.

Department of Biology, Center for Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, NM 87131

Received for publication April 18, 2011. Accepted for publication September 2, 2011.

This work was supported in part by National Institutes of Health Grant IP20RR18754 from the Institutional Development Award program of the National Center for Research Resources and National Science Foundation Award IOS-0641382 (to R.D.M.).

Address correspondence to Prof. Robert D. Miller, Department of Biology, Center for Evolutionary and Theoretical Immunology, University of New Mexico, MSC03 2020, Albuquerque, NM 87131-1091. E-mail address: rdmiller@unm.edu

The online version of this article contains supplemental material.

Abbreviations used in this article: CP, connecting peptide; CT, cytoplasmic; FR, framework region; L, leader; MYA, million years ago; NAR, new Ag receptor; RSS, recombination signal sequences; TM, transmembrane; UTR, untranslated regions; VH, Ig H chain V region.

Copyright © 2011 by The American Association of Immunologists, Inc. 0022-1767/11/$16.00

www.jimmunol.org/cgi/doi/10.4049/jimmunol.1101113

The Journal of Immunology
Platytrium \( \text{TCR}\mu \) is transcribed in a double V form

To investigate the structure of expressed platypus \( \text{TCR}\mu \), full-length transcripts were isolated from a spleen cDNA library. Transcripts averaged 1300 bp in length, which is longer than a conventional TCR transcript and more similar to the double V encoding opossum \( \text{TCR}\mu \) (Fig. 2, Table I). Each encoded a leader (L) peptide followed by two complete V domains, designated V1 and V2 for the 5′ (N-terminal) and 3′ (C-proximal) domains, respectively. They also contained one C domain along with sequences corresponding to the connecting peptide (CP), transmembrane (TM), and cytoplasmic (CT) regions typical of transmembrane TCR chains (Fig. 2). The clones encoded conserved residues found in conventional TCR including cysteines forming intrachain disulfide bonds in the V and C domains as well as interchain disulfide bond in the CP (Fig. 2). The framework region (FR) 4 of V1 and V2 contain the sequence YGXF and FXXG, respectively, similar to the conserved FGXG motif in conventional TCR and marsupial \( \text{TCR}\mu \) (4, 15, 16) (Fig. 2). Also present are two positively charged amino acids (arginine and lysine) in the TM region that, in conventional TCR chains, participate in association with the CD3 signaling complex (17). Comparison to the genic region revealed that the CP is unusual in platypus \( \text{TCR}\mu \) in that it is encoded on two exons, designated CP1 and CP2 with the conserved cysteine in CP2 (Fig. 2). This is unlike the opossum \( \text{TCR}\mu \) and most conventional TCR in which the CP is encoded by a single exon (4).

Both V1 and V2 are encoded by somatically recombined genes

The germline genes encoding the V1 and V2 domains were identified by comparing 18 unique V1 and 16 V2 sequences from both partial and full-length platypus splenic cDNA clones to the genome assembly. V1 and V2 domains share <65% nucleotide identity to each other and, by convention, are encoded by different V gene subgroups designated V\( \mu \)1 and V\( \mu \)2, respectively. Nine V\( \mu \)1 and six V\( \mu \)2 genes were identified in the germline sequence (Supplemental Fig. I). All nine of the V\( \mu \)1 genes contained upstream exons encoding a conserved L sequence; however, none of the V\( \mu \)2 germline genes had an L exon (not shown). The sequences corresponding to FR4 in V1 and V2 were also used to identify 8 J\( \mu \)1 and 12 J\( \mu \)2 genes, respectively. J\( \mu \)1 and J\( \mu \)2 are easily distinguished by length and sequence, with J\( \mu \)1 being shorter and sharing <50% nucleotide identity with J\( \mu \)2 genes (Fig. 3). All V\( \mu \) and J\( \mu \) genes were flanked by conserved RSS, the recognition substrates for the RAG product (18). The RSS flanking the V\( \mu \) and J\( \mu \) genes contained 23- and 12-bp spacers, respectively, typical of TCR genes (Fig. 3). In all cdDNA sequences analyzed, V\( \mu \)1 were recombined to J\( \mu \)1 and V\( \mu \)2 to J\( \mu \)2. These results support that both the V1 and V2 domains in platypus

<table>
<thead>
<tr>
<th>Sequence (5’ - to - 3’)</th>
<th>Orientation</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCGTGGCATGCGGGGACATGGCTG</td>
<td>R</td>
<td>C( \mu )</td>
</tr>
<tr>
<td>GGATGGATATCTTCACAGCCGAACG</td>
<td>R</td>
<td>C( \mu )</td>
</tr>
<tr>
<td>AGGAGCCTGGCTCAGGTAAAG</td>
<td>F/R</td>
<td>( \lambda ) gt10 vector</td>
</tr>
<tr>
<td>ATTAGGATATCTTCACCAGGTA</td>
<td>F/R</td>
<td>( \lambda ) gt10 vector</td>
</tr>
<tr>
<td>CCGAACCGATGCTGCTGCTCATG</td>
<td>F</td>
<td>C( \mu )</td>
</tr>
<tr>
<td>AACACTGCTGTCAGGTC</td>
<td>F</td>
<td>5′ UTR</td>
</tr>
<tr>
<td>CAGAGGGGAAATGATTAGAC</td>
<td>R</td>
<td>3′ UTR</td>
</tr>
<tr>
<td>CGGACAACAAAAAGAACGACCA</td>
<td>R</td>
<td>3′ UTR</td>
</tr>
<tr>
<td>CGTGAATTACTCGGAGGAT</td>
<td>F</td>
<td>V( \mu )1</td>
</tr>
<tr>
<td>AGGCTTCTGAATGATCTCCS</td>
<td>F</td>
<td>V( \mu )2</td>
</tr>
</tbody>
</table>

F. forward; R. reverse.
TCR\(\mu\) are encoded by exons that are fragmented in the germline DNA and undergo RAG-mediated V(D)J recombination. The sequences corresponding to CDR3 differed both in length and diversity between the V1 and V2 domains (Fig. 2). The V1 CDR3 are longer and up to 22 codons in length, whereas none of the V2 CDR3 exceeded 12 codons. Using the V1 CDR3 sequences identified 35 putative D\(\mu\) genes in the platypus genome assembly, all of which were asymmetrically flanked by RSS containing a 12-bp spacer on the 5' side and 23-bp spacer on the 3' side, as is typical of TCR D genes (Supplemental Fig. 3). Based on length and nucleotide identity, the D genes fell into two groups designated D\(\mu\)1 and -2. D\(\mu\)1 \((n = 20)\) contained coding regions 10–13 nucleotides in length, whereas D\(\mu\)2 \((n = 15)\) were 18 to 19 nucleotides (Supplemental Fig. 3). There was 75% nucleotide identity within each group but 40% nucleotide identity between D\(\mu\)1 and D\(\mu\)2 genes. Although D\(\mu\) genes could be distinguished in the genomic sequence, individual contributions to the V1 junctions were difficult to establish due to their similarity and short length. Nonetheless, it was possible to determine that the V\(\mu\)1–J\(\mu\)1 junctions contained two, three, or four D\(\mu\) genes, in an \(\sim\)1:2:1 ratio, similar to the multiple D genes found in opossum TCR\(\mu\) rearrangements (Fig. 4, Supplemental Table I). Typical of D gene segments, the D\(\mu\) present in V1 junctions were used in multiple reading frames (Supplemental Fig. 3). The gene segments encoding the V1 domains demonstrated extensive trimming and no evidence of P nucleotide additions, although N nucleotide additions were common (Fig. 4).

In contrast to V1, the CDR3 of 14 of the 16 V2 cDNA sequences could be accounted for entirely by recombination between germ-line V\(\mu\)2 and J\(\mu\)2 genes, with evidence for P and N nucleotide additions.

<table>
<thead>
<tr>
<th></th>
<th>Platypus C(\mu) ((n = 6))</th>
<th>Platypus C(\beta) ((n = 1))</th>
<th>Opossum C(\mu) ((n = 8))</th>
<th>C(\beta)(\alpha) ((n = 5))</th>
<th>C(\beta)(\beta) ((n = 5))</th>
<th>C(\beta)(\gamma) ((n = 5))</th>
<th>C(\beta)(\delta) ((n = 4))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platypus C(\mu) ((n = 6))</td>
<td>(80–98) (84)</td>
<td>(43–47) (44)</td>
<td>(50–56) (52)</td>
<td>(41–54) (50)</td>
<td>(21–26) (24)</td>
<td>(25–32) (29)</td>
<td>(27–33) (31)</td>
</tr>
<tr>
<td>Platypus C(\beta) ((n = 1))</td>
<td>(43–47) (44)</td>
<td>100</td>
<td>(43–47) (45)</td>
<td>(46–53) (50)</td>
<td>(26–30) (28)</td>
<td>(29–32) (30)</td>
<td>(29–33) (32)</td>
</tr>
<tr>
<td>Opossum C(\mu) ((n = 8))</td>
<td>(50–36) (52)</td>
<td>(43–47) (45)</td>
<td>(75–96) (83)</td>
<td>(41–54) (48)</td>
<td>(21–30) (25)</td>
<td>(26–34) (30)</td>
<td>(26–33) (29)</td>
</tr>
<tr>
<td>C(\beta) ((n = 5))</td>
<td>(41–54) (50)</td>
<td>(46–53) (50)</td>
<td>(41–54) (48)</td>
<td>(55–83) (67)</td>
<td>(21–30) (25)</td>
<td>(24–31) (27)</td>
<td>(28–34) (31)</td>
</tr>
<tr>
<td>C(\alpha) ((n = 5))</td>
<td>(21–26) (24)</td>
<td>(26–30) (28)</td>
<td>(21–30) (25)</td>
<td>(45–47) (54)</td>
<td>(22–33) (27)</td>
<td>(24–33) (27)</td>
<td>(28–36) (31)</td>
</tr>
<tr>
<td>C(\gamma) ((n = 4))</td>
<td>(25–32) (29)</td>
<td>(29–32) (30)</td>
<td>(26–34) (30)</td>
<td>(24–31) (27)</td>
<td>(63–93) (72)</td>
<td>(28–36) (31)</td>
<td>(28–36) (31)</td>
</tr>
</tbody>
</table>

Values are range of percent nucleotide identity (mean percent nucleotide identity).
\(\alpha\) C\(\beta\) sequences of human, mouse, opossum, bandicoot, and wallaby.
\(\beta\) C\(\beta\) sequences of human, mouse, opossum, echidna, and platypus.
\(\gamma\) C\(\gamma\) sequences of human, mouse, opossum, echidna, and platypus.
\(\delta\) C\(\delta\) sequences of human, mouse, opossum, and platypus.
Number of sequences included in the comparison.
Three scaffolds contain multiple DmTCR

Combining the scaffold analyses with the cDNA sequences reveals a stream of Vm with the identification of several scaffolds ranging in length from 1 kb to 64.8 kb (Supplemental Table I). These results imply there may be multiple Vm genes (4, 28416) containing single Vm1 and Vm2 genes that correspond to those used in expressed recombinations (Fig. 6A, Supplemental Fig. 1). One scaffold (28416) contains single Vm2 and Jm2 genes that correspond to those used in expressed recombinations (Fig. 6A, Supplemental Fig. 1, Supplemental Table I). However, no Dm genes were found on this scaffold consistent with the lack of D segments in the majority of Vm2–Jm2 junctions (Figs. 4, 6, Supplemental Fig. 1, Supplemental Table I).

Full-length cDNA clones containing similar or identical Vm1 sequences also had similar or identical Jm1, Vm2, Jm2, and Cm (Supplemental Table I). The most parsimonious explanation for these observations is a cluster organization of platypus TCRm genes, similar to that found in marsupials (4). In other words, the V, D, and J genes encoding V1 domains are upstream of the V and J gene segments encoding V2, followed by Cm (Fig. 6B). Consistent with this prediction, three scaffolds (19044, 26255, and 33931) contain Jm1 genes upstream of Vm2 genes, and many of the scaffolds containing Cm genes also contained an upstream Jm2 (Fig. 6A, Supplemental Fig. 1). A conservative model for the organization of the platypus TCRm genes is presented in Fig. 6B. The model may be overly conservative because two cDNA clones appeared to use different Vm1 but the same Jm1, whereas two others appeared to use the same Vm1 recombined to two different Jm1 (compare clones 2.34, 10, and 17 in Supplemental Table I). These results imply there may be multiple Vm1 and Jm1 in some clusters or alternatively may be due to trans-cluster recombination, as has been found for both opossum TCRm and shark TCR8 genes (4, 20).

To estimate the possible number of TCRm clusters, the number of unique Cm sequences that could be isolated from an individual platypus was determined. PCR was performed on genomic DNA from a single platypus using primers designed to amplify all 15 Cm identified in the genome assembly. Twenty individual clones were sequenced and yielded nine distinct C sequences consistent with the evidence from cDNA sequences having multiple Dm in the junctions between Vm1 and Jm1 genes (Fig. 6A, Supplemental Fig. 1). One scaffold (28416) contains single Vm2 and Jm2 genes that correspond to those used in expressed recombinations (Fig. 6A, Supplemental Fig. 1, Supplemental Table I). However, no Dm genes were found on this scaffold consistent with the lack of D segments in the majority of Vm2–Jm2 junctions (Figs. 4, 6, Supplemental Fig. 1, Supplemental Table I).

FIGURE 2. Predicted amino acid alignment of full-length platypus TCRm cDNA clones. Dashes indicate identity, and gaps introduced to the alignment are indicated with an *. Conserved residues YGXG and FXXG in FR4 of the V1 and V2 domains, respectively, are noted. The borders of CDR and FR are indicated above the sequences.

*Jr1*
with at least five Cμ exons per haploid platypus genome (not shown). This number is slightly lower but not significantly different from what would be predicted from the platypus whole genome sequence in which 15 different Cμ were identified or a minimum of eight per haploid genome. Whether this is an artifact of the assembly or normal platypus variation remains to be determined.

Discussion

The discovery of a platypus TCRμ homolog confirms that this unconventional TCR locus is not unique to marsupials but rather it is ancient in the mammalian lineage and appeared prior to the divergence of the prototherian (monotremes) and therian (marsupial and placental) mammals 165 MYA (7). TCRμ was clearly retained in the marsupial lineage and, therefore, would have been
present in the last common ancestor of marsupials and placental mammals. However, no TCRμ homolog has been identified in placental mammals, consistent with gene loss in this lineage (2). Furthermore, a TCRμ homolog has yet to be found in the available avian, reptilian, and amphibian genomes, consistent with its appearance in the synapsids (mammals and their extinct relatives) after their divergence from the diapsids (birds and reptiles) 310 MYA (2, 21). This conclusion is also consistent with phylogenetic analyses of TCRμ C region genes published previously, in which marsupial Cμ appears to diverge from Cμ after the split between mammals and birds (4).

The most distinctive feature common to both marsupial and platypus TCRμ is their transcription in a form predicted to encode three extracellular Ig domains (V-V-C) instead of the conventional two domains (V-C). TCR with this characteristic have only been described in one other vertebrate lineage, the cartilaginous fish. Both the elasmobranchs (sharks, rays, and skates) and the holocephalins (ratfish) use an isoform of TCRd, called new Ag receptor (NAR)-TCR, that also has a double V expressed with a conventional Cd (22).

There are a number of common characteristics shared between mammalian TCRμ and shark NAR-TCR, as well as distinctive differences (Table III). In both platypus TCRμ and NAR-TCR, the exons encoding both V domains require somatic DNA recombination to be assembled (22). The supporting or V2 domains in NAR-TCR are encoded by a dedicated subset of Vδ gene segments that, like the platypus Vμ2, lack L sequences and would be unable to encode the N terminus of an extracellular protein (22). This is different, however, in marsupials in which the exon encoding the V2 domain, called Vμj, is preassembled as a germline-joined gene and contains an L sequence that is contiguous with the exon encoding the extracellular V domain (Fig. 6C) (4). In the case of marsupial TCRμ, this L sequence is left out of the Vμj exon in the mature mRNA due to a canonical RNA splice site at the junctions between the L and V sequences (2, 4). This arrangement makes it possible to transcribe a two-domain form of marsupial TCRμ that contains only the Vμj and C regions. Indeed, such transcripts are found in the opossum thymus; however, they are rare in peripheral lymphoid tissues, leading to the current working hypothesis that it is the double-V form that is the mature, functional chain (4). Furthermore, in the opossum, Monodelphis domestica, there are eight tandem clusters of TCRμ genes, and in six of these, the Vμj L sequences contain mutations rendering them nonfunctional (2, 4). Therefore, whereas the shark and platypus have fully deleted the L sequence of the supporting V, the L sequences in marsupials are apparently degenerating due to lack of use.

Both TCRμ and NAR-TCR use V domains more similar to Ab V genes than conventional TCR V genes. The N-terminal V domains in NAR-TCR are related to V used in IgNAR, which are L chainless Abs unique to cartilaginous fishes (22, 23). As already described, the second V in NAR-TCR is a Vδ gene, making the NAR-TCR appear to be a hybrid between IgNAR and TCRδ (22). In contrast, the genes used to encode both V1 and V2 domains in platypus TCRμ are indistinguishable from mammalian clan III Ig VH genes and unrelated to NAR V genes. Marsupial Vμ and Vμj, in contrast, are somewhat intermediary. Vμj are more similar to Ig VH, but do not fall within the three traditional mammalian VH clans, and Vμj appear to be more related to NAR V genes, although this latter relationship is only weakly supported in phylogenetic analyses (Fig. 5).
The current model for the structure of NAR-TCR is an unpaired N-terminal domain, much like the V-NAR domain in IgNAR, binding Ags as a single domain (22, 23). This Ag binding is similar to that which has been described for single V domain IgNAR Abs in sharks and L chainless IgG in camels (24, 25). It seems likely that TCRμ is structured similarly to NAR-TCR, with a single, unpaired N-terminal V domain capable of binding Ag directly. Based on conserved residues, including cysteines, TCRμ is predicted to form a heterodimer with another TCR chain (4). However, because no other TCR-related genes encoding a three-domain chain have been found in the marsupial genome, it is predicted that the partner is a conventional two-domain TCR chain, likely TCRγ, leaving the N-terminal domain unpaired (2).

The common characteristics found in mammalian TCRμ and shark NAR-TCR raise the question of whether these features are due to homology by descent or convergent evolution. An argument could be made that the evolutionary distance between sharks and mammals is sufficiently vast, and the differences between TCRμ and NAR-TCR extensive enough that each evolved independently and appear analogous due to convergence on a common structure and function. This could imply a common evolutionary pressure shared between cartilaginous fish and early mammals to have T cells capable of binding Ag directly using single domain binding sites.

Phylogenetic analyses of platypus and marsupial TCRμ C region support that they are orthologous genes that would have been found in a last common ancestor of the three living mammalian lineages. However, following the divergence of the viviparous monotremes from the viviparous marsupials and placental mammals, TCRμ appears to have followed different evolutionary paths. In the placental mammals, it was lost altogether (2). As discussed earlier, in the marsupials, the genes encoding the V2 domain appear to have been replaced in the germline by a pre-joined V gene, most likely via retrotransposition (4). This novel marsupial adaptation is consistent with the V2 domains serving strictly supporting roles rather than being Ag binding and, therefore, requiring little or no clonal variation. In the platypus, the TCRμ V2 domain is encoded by somatically recombined genes, but variation remains restricted through limited junctional diversity, with no D segments and few N or P additions in the V-J junctions. Comparisons of the length of the CDR3 region in the platypus and marsupial V2 domains, where they are both relatively short, suggests that D segments, if they were ever present, were deleted early in the evolution of TCRμ prior to the divergence of prototherians and therians (4). The mean codon length of the platypus V2 CDR3 is the same (n = 11) as that found in the germline-joined marsupial Vμj genes (Table III). In contrast, the V1 domains of both platypus and opossum TCRμ have comparatively longer and more diverse CDR3 due to the incorporation of multiple D segments during V(D)J recombination in both species (4, 26).

The lack of an intron separating the L from the V in the Vμj exon is evidence of retrotransposition in the evolution of TCRμ in marsupials (4). In other words, Vμj is a functional, partially processed gene. The insertion of joined V genes into the germline by retrotransposition would require coexisting retroelements in the genome, and one noteworthy distinction between the opossum and the platypus genomes is the abundance of retroelements. The opossum has among the highest percentage of retroelements of any vertebrate genome sequenced (27). In contrast, monotremes are relatively devoid of retroelements (14, 28). Whether this extreme difference contributed to the evolution of opossum and platypus TCRμ is not known. Furthermore, this explanation is not fully satisfying because processed pseudogenes have been found.
in the platypus and echidna genomes, consistent with retrotransposition having occurred sometime in the past for some monotreme genes (10).

Phylogenetic analyses support TCR\(\alpha\) being related to and likely derived from a TCR\(\delta\) ancestor (4, 5). As stated earlier, if TCR\(\alpha\) evolved from a duplication of TCR\(\delta\) genes, it likely occurred after the separation of mammals from birds and reptiles (4). However, some insight into the origins of TCR\(\alpha\) may come from recent work on the genetics of amphibian TCR\(\delta\)-chains (29). The TCR\(\alpha\)/\(\delta\) locus in the frog *Xenopus tropicalis* contains two C\(\delta\) genes, one of which, C\(\delta\)1, is expressed with V genes called VH\(\delta\). These frog VH\(\delta\) are indistinguishable from clan II Ig VH genes, and, although the *X. tropicalis* TCR\(\alpha\)/\(\delta\) and Igh loci are closely linked, the VH\(\delta\) genes appear to be dedicated for use in TCR\(\delta\)-chains and are not used in IgH chains (29). This close linkage, however, may have facilitated insertion of VH genes among the TCR\(\delta\) genes in amphibians. The region of the frog TCR\(\alpha\)/\(\delta\) locus containing C\(\delta\)1 and multiple VH\(\delta\) genes is distinct and, in an inverted transcriptional orientation from the rest of the TCR\(\alpha\)/\(\delta\) genes, functioning almost as a separate minicluster (29). Amphibians, therefore, appear to be another vertebrate lineage that uses TCR\(\delta\)-chains containing Ab-like V genes. Unlike TCR\(\alpha\) and NAR-TCR, frog TCR\(\delta\)-chains are not expressed with two V domains, however. Rather, *X. tropicalis* TCR\(\delta\)-chains using VH\(\delta\) are structured like conventional two-domain TCR chains.

It is possible, and seems likely, that the TCR\(\alpha\) locus evolved from genome duplication and translocation of an ancestral region of the TCR\(\alpha\)/\(\delta\) locus similar to the C\(\delta\)1 region in frogs. Indeed, the discovery of VH genes in the *X. tropicalis* TCR\(\alpha\)/\(\delta\) locus is consistent with their presence in the TCR\(\delta\) locus prior to the

**FIGURE 6.** Diagrams of the predicted platypus TCR\(\alpha\) gene organization, transcripts, and protein structure. A, Representative TCR\(\alpha\) scaffolds containing TCR\(\alpha\) coding sequences. Closed or open triangles flanking the V\(\mu\), D\(\mu\), and J\(\mu\) gene segments indicate the presence of 23- or 12-bp spacer RSS, respectively. The L sequence, CP, TM-CT, and 3′ UTR exons are indicated. B, Predicted TCR\(\alpha\) germline DNA and rearranged DNA structure and primary TCR\(\alpha\) mRNA transcript structure. Conserved R and K residues in the TM region are indicated in the predicted cell surface TCR protein structure. C, Comparison of a representative opossum TCR\(\alpha\) cluster with the predicted platypus homolog.

**Table III.** Comparison of the features of TCR\(\mu\), shark NAR-TCR, and mammalian conventional TCR\(\alpha\)/\(\delta\)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Model Species</th>
<th>C Gene</th>
<th>Double V Leader Sequence</th>
<th>No. of D Segments Used</th>
<th>CDR3 Lengtha (Mean)</th>
<th>V Nature</th>
<th>Reference No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCR(\mu)</td>
<td>Platypus</td>
<td>C(\mu) Yes Yes Yes Yes No No 2-4 0? 9–22(14) 9–12(11) VH clan III VH clan III This study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Opossum</td>
<td>C(\mu) Yes Yes Yes Yes No No 1-3 NA 9–25(16) 9–27(16) VH-related VH-related</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nurse shark</td>
<td>C(\delta) Yes Yes Yes Yes No No 1-2 NA 7–20(13) NA V-NAR V(\delta)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCR(\alpha)/(\delta)</td>
<td><em>Xenopus</em></td>
<td>C(\delta) Yes NA Yes NA Yes NA 1-2 NA 7–20(13) NA V(\alpha), V(\delta), VH(\delta) (VH clan II) NA (29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCR(\alpha)/(\delta)</td>
<td>Human</td>
<td>C(\delta) No Yes NA Yes NA 2-3 NA 8–12(15) NA V(\alpha), V(\delta)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCR(\alpha)/(\delta)</td>
<td>Mouse</td>
<td>C(\delta) No Yes NA Yes NA 2 NA 6–19(13) NA V(\alpha), V(\delta)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aRange in codons.

bThe C proximal V in marsupial TCR\(\mu\) is a germline joined V.

cFused to the V domain exon as the result of retrotransposition.

---

By guest on April 20, 2017 http://www.jimmunol.org/ Downloaded from http://www.jimmunol.org/ on April 20, 2017
evolution of TCRµ. Internal duplications of clusters of V, D, and J segments within the TCRµ locus, as hypothesized previously, would then give rise to the double V organization in mammals (2). What remains puzzling is the variation in the source of VH genes used in each lineage. The VH6 in X. tropicalis are apparently derived from clan II VH, the platypus VHµ genes are clan III VH, and, although the marsupial VHµ genes are more closely related to VH than TCR V genes, they fall outside the clan I, II, and III designations. These observations suggest that the VH genes used in TCRδ- or TCRµ-chains have been replaced over time with different VH lineages, even within the mammals. If the platypus TCRµ locus is indeed organized as tandem clusters similar to what has been shown in opossum (4), such gene clusters may facilitate gene replacement and duplication that is not easily achieved by the translocon organization of the conventional TCR genes.

The lack of TCRµ in commonly studied mammals such as humans and mice no doubt contributed to it remaining undiscovered for nearly a quarter of a century following that of the conventional TCRα, β, γ, and δ (4, 29–33). Determining why placental mammals may have lost this TCR chain will require first determining what function(s) TCRµ T cells perform in those species in which they are found.

Disclosures
The authors have no financial conflicts of interest.

References