A Model for the Evolution of the Mammalian T-cell Receptor α/δ and μ Loci Based on Evidence from the Duckbill Platypus

Zuly E. Parra,[1] Mette Lillie,[2] and Robert D. Miller*,[1]

[1]Department of Biology, Center for Evolutionary & Theoretical Immunology, University of New Mexico
[2]Faculty of Veterinary Science, University of Sydney, Sydney, Australia
*Corresponding author: E-mail: rdmiller@unm.edu.
Associate editor: Yoko Satta

Abstract

The specific recognition of antigen by T cells is critical to the generation of adaptive immune responses in vertebrates. T cells recognize antigen using a somatically diversified T-cell receptor (TCR). All jawed vertebrates use four TCR chains called α, β, γ, and δ, which are expressed as either a αβ or γδ heterodimer. Nonplacental mammals (monotremes and marsupials) are unusual in that their genomes encode a fifth TCR chain, called TCRμ, whose function is not known but is also somatically diversified like the conventional chains. The origins of TCRμ are also unclear, although it appears distantly related to TCRδ. Recent analysis of avian and amphibian genomes has provided insight into a model for understanding the evolution of the TCRδ genes in tetrapods that was not evident from humans, mice, or other commonly studied placental (eutherian) mammals. An analysis of the genes encoding the TCRδ chains in the duckbill platypus revealed the presence of a highly divergent variable (V) gene, indistinguishable from immunoglobulin heavy (IgH) chain V genes (VH) and related to V genes used in TCRμ. They are expressed as part of TCRδ repertoire (VHδ) and similar to what has been found in frogs and birds. This, however, is the first time a VHδ has been found in a mammal and provides a critical link in reconstructing the evolutionary history of TCRμ. The current structure of TCRδ and TCRμ genes in tetrapods suggests ancient and possibly recurring translocations of gene segments between the IgH and TCRδ genes, as well as translocations of TCRδ genes out of the TCRα/δ locus early in mammals, creating the TCRμ locus.

Key words: T-cell receptors, monotremes, comparative genomics.

Introduction

T lymphocytes are critical to the adaptive immune system of all jawed vertebrates and can be classified into two main lineages based on the T-cell receptor (TCR) they use (Rast et al. 1997; reviewed in Davis and Chein 2008). The majority of circulating human T cells are the αβT cell lineage which use a TCR composed of a heterodimer of α and β TCR chains. αβT cells include the familiar T cell subsets such as CD4+ helper T cells and regulatory T cells, CD8+ cytotoxic T cells, and natural killer T (NKT) cells. T cells that are found primarily in epithelial tissues and a lower percentage of circulating lymphocytes in some species express a TCR composed of γ and δ TCR chains. The function of these γδ T cells is less well defined and they have been associated with a broad range of immune responses including tumor surveillance, innate responses to pathogens and stress, and wound healing (Hayday 2009). αβ and γδ T cells also differ in the way they interact with antigen. αβTCR are major histocompatibility complex (MHC) "restricted" in that they bind antigenic epitopes, such as peptide fragments, bound to, or "presented" by, molecules encoded in the MHC. In contrast, γδTCR have been found to bind antigens directly in the absence of MHC, as well as self-ligands that are often MHC-related molecules (Sciammas et al. 1994; Hayday 2009).

The conventional TCR chains are composed of two extracellular domains that are both members of the immunoglobulin (Ig) domain super-family (reviewed in Davis and Chein 2008) (fig. 1). The membrane proximal domain is the constant (C) domain, which is largely invariant amongst T-cell clones expressing the same class of TCR chain, and is usually encoded by a single, intact exon. The membrane distal domain is called the variable (V) domain and is the region of the TCR that contacts antigen and MHC. Similar to antibodies, the individual clonal diversity in the TCR V domains is
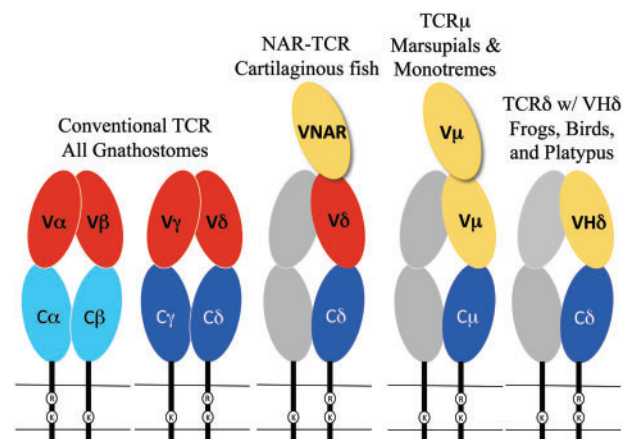

Fig. 1. Cartoon diagram of the TCR forms found in different species. Oblong circles indicate Ig super-family domains and are color coded as C domains (blue), conventional TCR V domains (red), and VHδ or Vμ (yellow). The gray shaded chains represent the hypothetical partner chain for TCRμ and TCRδ using VHδ.

generated by somatic DNA recombination (Tonegawa 1983). The exons encoding TCR V domains are assembled somatically from germ-line gene segments, called the V, diversity (D), and joining (J) genes, in developing T cells, a process dependent upon the enzymes encoded by the recombination activating genes (RAG)-1 and RAG-2 (Yancopoulos et al. 1986; Schatz et al. 1989). The exons encoding the V domains of TCR β and δ chains are assembled from all three types of gene segments, whereas the α and γ chains use only V and J. The different combinations of V, D, and J or V and J, selected from a large repertoire of germ-line gene segments, along with variation at the junctions due to addition and deletion of nucleotides during recombination, contribute to a vast TCR diversity. It is this diversity that creates the individual antigen specificity of T-cell clones.

The TCR genes are highly conserved among species in both genomic sequence and organization (Rast et al. 1997; Parra et al. 2008, 2012; Chen et al. 2009). In all tetrapods examined, the TCRβ and γ chains are each encoded at separate loci, whereas the genes encoding the α and δ chains are nested at a single locus (TCRα/δ) (Chien et al. 1987; Satyanarayana et al. 1988; reviewed in Davis and Chein 2008). The V domains of TCRα and TCRδ chains can use a common pool of V gene segments, but distinct D, J, and C genes.

Diversity in antibodies produced by B cells is also generated by RAG-mediated V(D)J recombination and the TCR and Ig genes clearly share a common origin in the jawed-vertebrates (Flajnik and Kasahara 2010; Litman et al. 2010). However, the V, D, J, and C coding regions in TCR have diverged sufficiently over the past >400 million years (MY) from Ig genes that they are readily distinguishable, at least for the conventional TCR. Recently, the boundary between TCR and Ig genes has been blurred with the discovery of non-conventional TCRδ isoforms that have been found that use V genes that appear indistinguishable from Ig heavy chain V (VH) (Parra et al. 2010, 2012). Such V genes have been designated as VHδ and have been found in both amphibians and birds (fig. 1). In the frog Xenopus tropicalis, and a passerine bird, the zebra finch Taeniopygia guttata the VHδ are located within the TCRα/δ loci where they co-exist with conventional Vα and Vδ genes (Parra et al. 2010, 2012). In galliform birds, such as the chicken Gallus gallus, VHδ are present but located at a second TCRδ locus that is unlinked to the conventional TCRα/δ (Parra et al. 2012). VHδ are the only type of V gene segment present at the second locus and, although closely related to antibody VH genes, the VHδ appear to be used exclusively in TCRδ chains. This is true as well for frogs where the TCRα/δ and IgH loci are tightly linked (Parra et al. 2010).

The TCRα/δ loci have been characterized in several eutherian mammal species and at least one marsupial, the opossum Monodelphis domestica, and VHδ genes have not been found to date (Satyanarayana et al.1988; Wang et al. 1994; Parra et al. 2008). However, marsupials do have an additional TCR locus, unlinked to TCRα/δ, that uses antibody-related V genes. This fifth TCR chain is called TCRμ and is related to TCRδ, although it is highly divergent in sequence and structure (Parra et al. 2007, 2008). A TCRμ has also been found in the duckbill platypus and is clearly orthologous to the marsupial genes, consistent with this TCR chain being ancient in mammals, although it has been lost in the eutherians (Parra et al. 2008; Wang et al. 2011). TCRμ chains use their own unique set of V genes (Vμ) (Parra et al. 2007; Wang et al. 2011). Trans-locus V(D)J recombination of V genes from other Ig and TCR loci with TCRμ genes has not been found. So far, TCRμ homologues have not been found in non-mammals (Parra et al. 2008).

TCRμ chains are atypical in that they contain three extra-cellular IgSF domains rather than the conventional two, due to an extra N-terminal V domain (fig. 1) (Parra et al. 2007; Wang et al. 2011). Both V domains are encoded by a unique set of Vμ genes and are more related to Ig VH than to conventional TCR V domains. The N-terminal V domain is diverse and encoded by genes that undergo somatic V(D)J recombination. The second or supporting V domain has little or no diversity. In marsupials this V domain is encoded by a germ-line joined, or pre-assembled, V exon that is invariant (Parra et al. 2007). The second V domain in platypus is encoded by gene segments requiring somatic DNA recombination; however, only limited diversity is generated partly due to the lack of D segments (Wang et al. 2011). A TCR chain structurally similar to TCRμ has also been described in sharks and other cartilaginous fish (fig. 1) (Criscitiello et al. 2006; Flajnik et al. 2011). This TCR, called NAR-TCR, also contains three extracellular domains, with the N-terminal V domain being related to those used by IgNAR antibodies, a type of antibody found only in sharks (Greenberg et al. 1995). The current working model for both TCRμ and NAR-TCR is that the N-terminal V domain is unpaired and acts as a single, antigen binding domain, analogous to the V domains of light-chainless antibodies found in sharks and camelids (Flajnik et al. 2011; Wang et al. 2011).

Phylogenetic analyses support the origins of TCRμ occurring after the avian–mammalian split (Parra et al. 2007; Wang et al. 2011). Previously, we hypothesized the origin of TCRμ being the result of a recombination between ancestral IgH and TCRδ-like loci (Parra et al. 2008). This hypothesis, however, is problematic for a number of reasons. One challenge is the apparent genomic stability and ancient conserved synteny in the region surrounding the TCRα/δ locus; this region has appeared to remain stable over at least the past 350 MY of tetrapod evolution (Parra et al. 2008, 2010). The discovery of VHδ genes inserted into the TCRα/δ locus of amphibians and birds has provided an alternative model for the origins of TCRμ; this model involves both the insertion of VH followed by the duplication and translocation of TCR genes. Here we present the model along with supporting evidence drawn from the structure of the platypus TCRα/δ locus, which is also the first analysis of this complex locus in a monotreme.

Materials and Methods

Identification and Annotation of the Platypus TCRα/δ Locus

The analyses were performed using the platypus (Ornithorhynchus anatinus) genome assembly version 5.0.1 (http://www.ncbi.nlm.nih.gov/genome/guide/platypus/).

The platypus genome was analyzed using the whole-genome BLAST available at NCBI (www.ncbi.nlm.nih.gov/) and the BLAST/BLAT tool from Ensembl (www.ensembl.org). The V and J segments were located by similarity to corresponding segments from other species and by identifying the flanking conserved recombination signal sequences (RSS). V gene segments were annotated 5′ to 3′ as Vα or Vδ followed by the family number and the gene segment number if there were greater than one in the family. For example, Vα15.7 is the seventh Vα gene in family 15. The D segments were identified using complementarity-determining region-3 (CDR3) sequences that represent the V–D–J junctions, from cDNA clones using VHδ. Platypus TCR gene segments were labeled according to the IMGT nomenclature (http://www.imgt.org/). The location for the TCRα/δ genes in the platypus genome version 5.0.1 is provided in supplementary table S1, Supplementary Material online.

Confirmation of Expression of Platypus VHδ

Reverse transcription PCR (RT–PCR) was performed on total splenic RNA extracted from a male platypus from the Upper Barnard River, New South Wales, Australia. This platypus was collected under the same permits as in Warren et al. (2008). The cDNA synthesis step was carried out using the Invitrogen Superscript III-first strand synthesis kit according to the manufacturer's recommended protocol (Invitrogen, Carlsbad, CA, USA). TCRδ transcripts containing VHδ were targeted using primers specific for the Cδ and VHδ genes identified in the platypus genome assembly (Warren et al. 2008). PCR amplification was performed using the QIAGEN HotStar HiFidelity Polymerase Kit (BD Biosciences, CLONTECH Laboratories, Palo Alto, CA, USA) in total volume of 20 μl containing 1× Hotstar Hifi PCR Buffer (containing 0.3 mM dNTPs), 1 μM of primers, and 1.25U Hotstar Hifidelity DNA polymerase. The PCR primers used were 5′-GTACCGCCAACCACCAGGGAAAG-3′ and 5′-CAGTTCACTGCTCCATCGCTTTCA-3′ for the VHδ and Cδ, respectively. A previously described platypus spleen cDNA library constructed from RNA extracted from tissue from a Tasmanian animal was also used (Vernersson et al. 2002).

PCR products were cloned using TopoTA cloning® kit (Invitrogen). Sequencing was performed using the BigDye terminator cycle sequencing kit version 3 (Applied Biosystems, Foster City, CA, USA) and according to the manufacturer recommendations. Sequencing reactions were analyzed using the ABI Prism 3100 DNA automated sequences (PerkinElmer Life and Analytical Sciences, Wellesley, MA, USA). Chromatograms were analyzed using the Sequencher 4.9 software (Gene Codes Corporation, Ann Arbor, MI, USA). Sequences have been archived on GenBank under accession numbers JQ664690–JQ664710.

Phylogenetic Analyses

Nucleotide sequences from FR1 to FR3 of the V genes regions, including CDR1 and CDR2, were aligned using BioEdit (Hall 1999) and the accessory application ClustalX (Thompson et al. 1997). Nucleotide alignments analyzed were based on amino acid sequence to establish codon position (Hall 1999). Alignments were corrected by visual inspection when necessary and were then analyzed using the MEGA Software (Kumar et al. 2004). Neighbor joining (NJ) with uncorrected nucleotide differences (p-distance) and minimum evolution distances methods were used. Support for the generated trees was evaluated based on bootstrap values generated by 1000 replicates. GenBank accession numbers for sequences used in the tree construction are in supplementary table S2, Supplementary Material online.

Results and Discussion

The TCRα/δ locus was identified in the current platypus genome assembly and the V, D, J, and C gene segments and exons were annotated and characterized (fig. 2). The majority of the locus was present on a single scaffold, with the remainder on a shorter contig (fig. 2). Flanking the locus were SALL2, DAD1 and several olfactory receptor (OR) genes, all of which share conserved synteny with the TCRα/δ locus in amphibians, birds, and mammals (Parra et al. 2008, 2010, 2012). The platypus locus has many typical features common to TCRα/δ loci in other tetrapods (Satyanarayana et al. 1988; Wang et al. 1994; Parra et al. 2008, 2010, 2012). Two C region genes were present: a Cα that is the most 3′ coding segment in the locus, and a Cδ oriented 5′ of the Jα genes. There is a large number of Jα gene segments ($n = 32$) located between the Cδ and Cα genes. Such a large array of Jα genes are believed to facilitate secondary Vα to Jα rearrangements in developing $\alpha\beta$T cells if the primary rearrangements are nonproductive or need replacement (Hawwari and Krangel 2007). Primary TCRα V–J rearrangments generally use Jα segments towards the 5′-end of the array and can progressively use downstream Jα in subsequent rearrangements. There is also a single Vδ gene in reverse transcriptional orientation between the platypus Cδ gene and the Jα array that is conserved in mammalian TCRα/δ both in location and orientation (Parra et al. 2008).

There are 99 conventional TCR V gene segments in the platypus TCRα/δ locus, 89 of which share nucleotide identity with Vα in other species and 10 that share identity with Vδ genes. The Vδ genes are clustered towards the 3′-end of the locus. Based on nucleotide identity shared among the platypus V genes they can be classified into 17 different Vα families and two different Vδ families, based on the criteria of a V family sharing >80% nucleotide identity (not shown, but annotated in fig. 2). This is also a typical level of complexity for mammalian Vα and Vδ genes (Giudicelli et al. 2005; Parra et al. 2008). Also present were two Dδ and seven Jδ gene segments oriented upstream of the Cδ. All gene segments were flanked by canonical RSS, which are the recognition substrate of the RAG recombinase. The D segments were asymmetrically flanked by an RSS containing at 12 bp spacer on the 5′-side and 23 bp spacer on the 3′-side, as has been shown previously for TCR D gene segments in other species (Carroll et al. 1993; Parra et al. 2007, 2010). In summary, the overall content and organization of the platypus TCRα/δ locus appeared fairly generic.
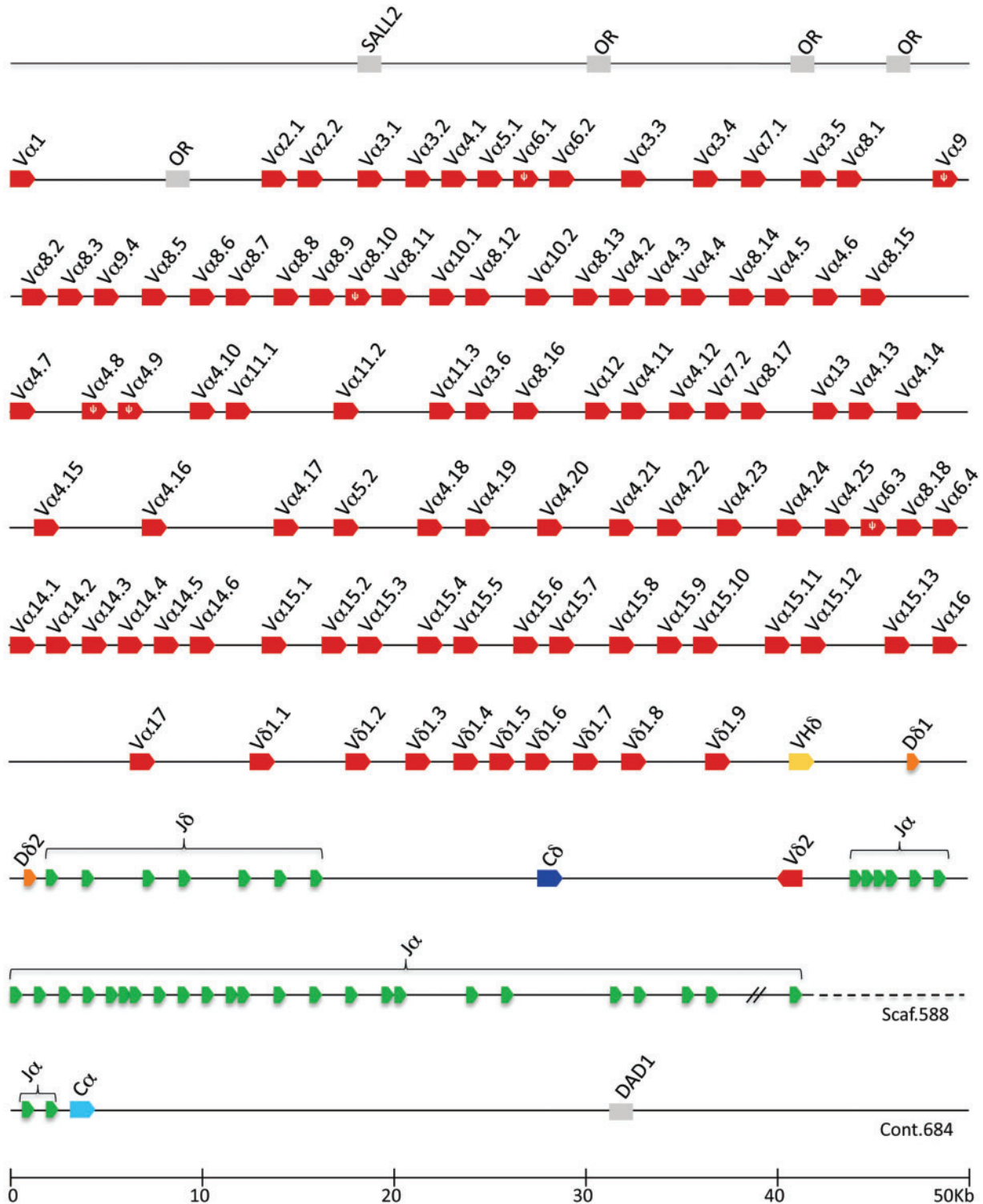
Fig. 2. Annotated map of the platypus *TCRα/δ* locus showing the locations of the Vα and Vδ (red), VHδ (yellow), Dδ (orange), Jα and Jδ (green), Cδ (dark blue), and Cα (light blue). Conserved syntenic genes are in gray. The scaffold and contig numbers are indicated.

What is atypical in the platypus *TCRα/δ* locus was the presence of an additional V gene that shared greater identity to antibody VH genes than to TCR V genes (figs. 2 and 3). This V gene segment was the most proximal of the V genes to the D and J genes and was tentatively designated as VHδ. VHδ are, by definition, V genes indistinguishable from Ig VH genes but used in encoding TCRδ chains and have previously been

found only in the genomes of birds and frogs (Parra et al. 2008, 2010, 2012).

VH genes from mammals and other tetrapods have been shown to cluster into three ancient clans and individual species differ in the presence of one or more of these clans in their germ-line IgH locus (Tutter and Riblet 1989; Ota and Nei 1994). For example, humans, mice, echidnas, and frogs have
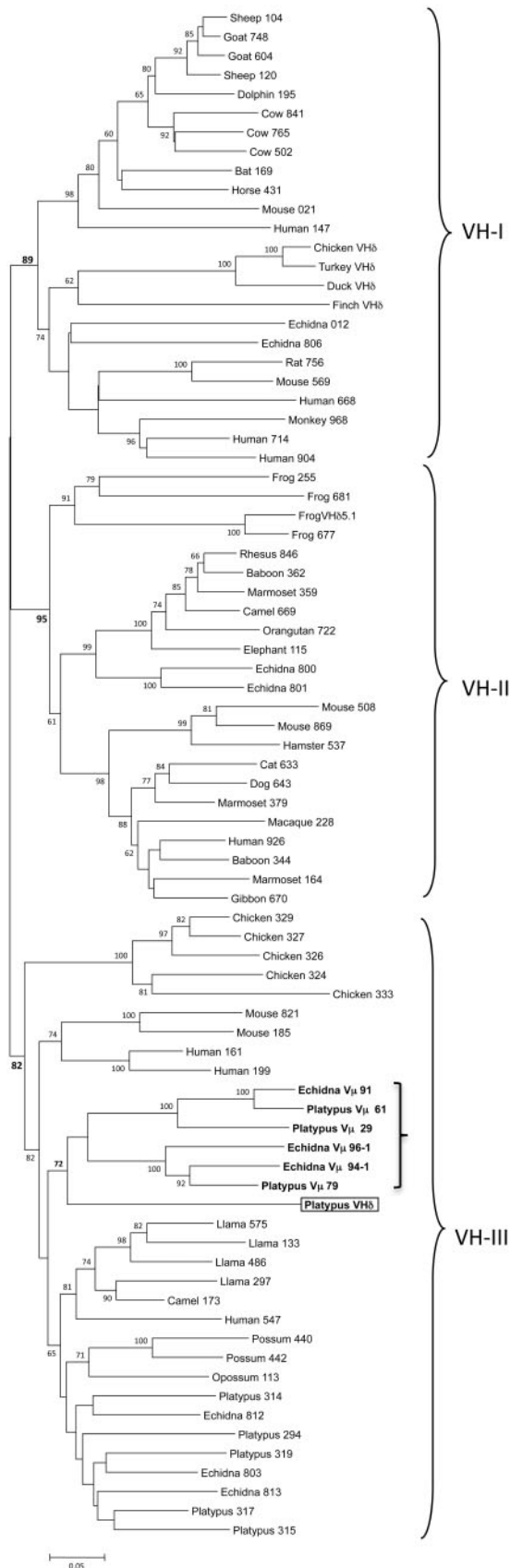
Fig. 3. Phylogenetic tree of mammalian VH genes including the platypus VHδ and monotreme Vμ. The three major VH clans are bracketed. The platypus VHδ is boxed and the clade containing platypus VHδ along with platypus and echidna Vμ is in bold and indicated by a smaller bracket in VH clan III. The three-digit numbers following the VH gene labels are the last three digits of the GenBank accession number referenced in supplementary table S2, Supplementary Material online. The numbers following the platypus and echidna Vμ labels are clone numbers. The tree presented was generated using the Minimum Evolution method. Similar topology was generation using the Neighbor Joining method.

VH genes from all three clans (Schwager et al. 1989; Ota and Nei 1994; Belov and Hellman 2003), whereas rabbits, opossums, and chickens have only a single clan (McCormack et al. 1991; Butler 1997; Johansson et al. 2002; Baker et al. 2005). In phylogenetic analyses, the platypus VHδ was most related to the platypus Vμ genes found in the TCRμ locus in this species (fig. 3). Platypus VHδ, however, share only 51–61% nucleotide identity (average 56.6%) with the platypus Vμ genes. Both the platypus Vμ and VHδ clustered within clan III (fig. 3) (Wang et al. 2011). This is noteworthy given that VH genes in the platypus IgH locus are also clan III and, in general, clan III VH are the most ubiquitous and conserved lineage of VH (Johansson et al. 2002; Tutter and Riblet 1989). Although clearly related to platypus VH, the VHδ gene share only 34–65% nucleotide identity (average 56.9%) with the *bona fide* VH used in antibody heavy chains in this species.

It was necessary to rule out that the VHδ gene present in the platypus TCRα/δ locus was not an artifact of the genome assembly process. One piece of supporting evidence would be the demonstration that the VHδ is recombined to downstream Dδ and Jδ segments and expressed with Cδ in complete TCRδ transcripts. PCR using primers specific for VHδ and Cδ was performed on cDNA synthesized from splenic RNA from two different platypuses, one from New South Wales and the other from Tasmania. PCR products were successfully amplified from the NSW animal and these were cloned and sequenced. Twenty clones, each containing unique nucleotide sequence, were characterized and found to contain the VHδ recombined to the Dδ and Jδ gene segments (fig. 4A). Of these 20, 11 had unique V, D, and J combinations that would encode 11 different complementarity-determining regions-3 (CDR3) (fig. 4B). More than half of the CDR3 (8 out of 11) contained evidence of using both D genes (VDDJ) (fig. 4B). This is a common feature of TCRδ V domains where multiple D genes can be incorporated into the recombination due to the presence of asymmetrical RSS (Carroll et al. 1993). The region corresponding to the junctions between the V, D, and J segments, contained additional sequence that could not be accounted for by the germ-line gene segments (fig. 4B). There are two possible sources of such sequence. One are palindromic (P) nucleotides that are created during V(D)J recombination when the RAG generates hairpin structures that are resolved asymmetrically during the re-ligation process (Lewis 1994). The second are non-templated (N) nucleotides that can be added by the enzyme terminal deoxynucleotidyl transferase (TdT) during the V(D)J recombination process. An unusual feature of the platypus VHδ is the presence of a second cysteine encoded near the 3′-end of the gene, directly next to the cysteine
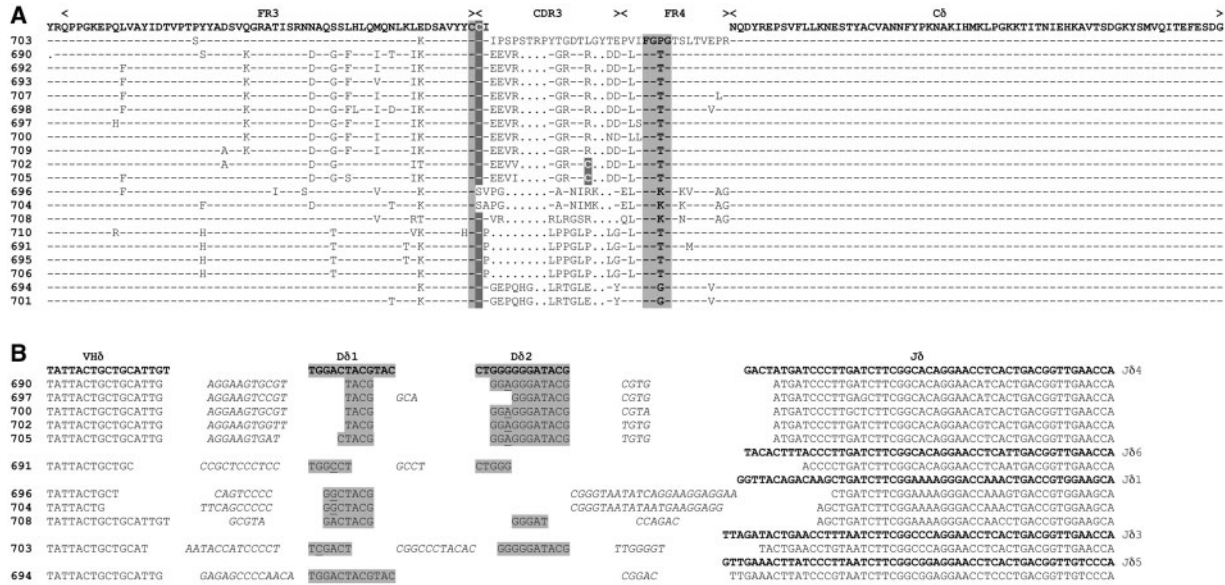
Fig. 4. (A) Alignment of predicted protein sequence of transcripts containing a recombined VHδ gene isolated from platypus spleen RNA. The individual clones are identified by the last three digits of their GenBank accession numbers (JQ664690–JQ664710). Shown is the region from FR3 of the VHδ through the beginning of the Cδ domain. The sequence in bold at the top of the alignment is the germ-line VHδ and Cδ gene sequence. The double cysteines at the end of FR3 and unpaired cysteines in CDR3 are shaded, as is the canonical FGXG in FR4. (B) Nucleotide sequence of the CDR3 region of the eleven unique V(D)J recombinants using VHδ described in the text. The germ-line sequence of the 3′-end of VHδ, the two Dδ, are shown at the top. The germ-line Jδ sequences are shown on the right-hand side of the alignment interspersed amongst the cDNA sequences using each. Nucleotides in the junctions between the V, D, and J segments, shown italicized, are most likely N-nucleotides added by TdT.

predicted to form the intra-domain disulfide bond in Ig domains (fig. 4A). Additional cysteines in the CDR3 region of VH domains have been thought to provide stability to unusually long CDR3 loops, as has been described for cattle and the platypus previously (Johansson et al. 2002). The CDR3 of TCRδ using VHδ are only slightly longer than conventional TCRδ chains (ranging 10–20 residues) (Rock et al. 1994; Wang et al. 2011). Furthermore, the stabilization of CDR3 generally involves multiple pairs of cysteines, which were not present in the platypus VHδ clones (fig. 4A). Attempts to amplify TCRδ transcripts containing VHδ from splenic RNA obtained from the Tasmanian animal were unsuccessful. As a positive control, TCRδ transcripts containing conventional Vα/δ were successfully isolated, however. It is possible that Tasmanian platypuses, which have been separated from the mainland population at least 14,000 years either have a divergent VHδ or have deleted this single V gene altogether (Lambeck and Chappell 2001).

Although there is only a single VHδ in the current platypus genome assembly, there was sequence variation in the region corresponding to FR1 through FR3 of the V domains (fig. 4A and sequence data not shown but available in GenBank). Some of this variation could represent two alleles of a single VHδ gene. Indeed, the RNA used in this experiment is from a wild-caught individual from the same population that was used to generate the whole-genome sequence and was found to contain substantial heterozygosity (Warren et al. 2008). There was greater variation in the transcribed sequences, however, than could be explained simply by two alleles of a single gene (fig. 4A). Two alternative explanations are the occurrence of somatic mutation of expressed VHδ genes or

allelic variation in gene copy number. Somatic mutation in TCR chains is controversial. Nonetheless, it has been invoked to explain the variation in expressed TCR chains that exceeds the apparent gene copy number in sharks, and has also been postulated to occur in salmonids (Yazawa et al. 2008; Chen et al. 2009). Therefore, it does not seem to be out of the realm of possibility that somatic mutation is occurring in platypus VHδ. Indeed, the mutations appear to be localized to the V region with no variation in the C region (fig. 4A). This may be due to its relatedness of VHδ to Ig VH genes where somatic hyper-mutation is well documented. Such somatic mutation contributes to overall affinity maturation in secondary antibody responses (Wysocki et al. 1986). The pattern of mutation seen in platypus VHδ however, is not localized to the CDR3, which would be indicative of selection for affinity maturation, but was also found in the framework regions. Furthermore, in the avian genomes where there is also only a single VHδ, there was no evidence of somatic mutation in the V regions (Parra et al. 2012). The contribution of mutation to the platypus TCRδ repertoire, if it is occurring, remains to be determined. Alternatively, the sequence polymorphism may be due to VHδ gene copy number variation between individual TCRα/δ alleles.

Irrespective of the number of VHδ genes in the platypus TCRα/δ locus, the results clearly support TCRδ transcripts containing VHδ recombined to Dδ and Jδ gene segments in the TCRα/δ locus (fig. 4). A VHδ gene or genes in the platypus TCRα/δ locus in the genome assembly, therefore, does not appear to be an assembly artifact. Rather it is present, functional and contributes to the expressed TCRδ chain repertoire. The possibility that some platypus TCRα/δ loci contain

more than a single VHδ does not alter the principal conclusions of this study.

Previously, we hypothesized the origin of TCRμ in mammals involving the recombination between and ancestral TCRα/δ locus and an IgH locus (Parra et al. 2008). The IgH locus would have contributed the V gene segments at the 5′-end of the locus, with the TCRδ contributing the D, J, and C genes at the 3′-end of the locus. The difficulty with this hypothesis was the clear stability of the genome region surrounding the TCRα/δ locus. In other words, the chromosomal region containing the TCRα/δ locus appears to have remained relatively undisrupted for at least the past 360 million years (Parra et al. 2008, 2010, 2012). The discovery of VHδ genes within the TCRα/δ loci of frog and zebra finch is consistent with insertions occurring without apparently disrupting the local syntenic region. In frogs, the IgH and TCRα/δ loci are tightly linked, which may have facilitated the translocation of VH genes into the TCRα/δ locus (Parra et al. 2010). However, close linkage is not a requirement since the translocation of VH genes appears to have occurred independently in birds and monotremes, due to the lack of similarity between the VHδ in frogs, birds, and monotremes (Parra et al. 2012). Indeed, it would appear is if the acquisition of VH genes into the TCRα/δ locus occurred independently in each lineage.

The similarity between the platypus VHδ and V genes in the TCRμ locus is, so far, the clearest evolutionary association between the TCRμ and TCRδ loci in one species. From the comparison of the TCRα/δ loci in frogs, birds, and monotremes, a model for the evolution of TCRμ and other TCRδ forms emerges (fig. 5), which can be summarized as follows:

1) Early in the evolution of tetrapods, or earlier, a duplication of the D–J–Cδ cluster occurred resulting in the presence of two Cδ each with its own set of Dδ and Jδ segments (fig. 5A).

2) Subsequently, a VH gene or genes was translocated from the IgH locus and inserted into the TCRα/δ locus, most likely to a location between the existing Vα/Vδ genes and the 5′-proximal D–J–Cδ cluster (fig. 5B). This resulted in the configuration like that which currently exists in the zebra finch genome (Parra et al. 2012).

3) In the amphibian lineage there was an inversion of the region containing VHδ–Dδ–Jδ–Cδ cluster and an expansion in the number of VHδ genes (fig. 5C). Currently, X. tropicalis has the greatest number of VHδ genes, where they make up the majority of V genes available in the germ-line for use in TCRδ chains (Parra et al. 2010).

4) In the galliform lineage (chicken and turkey), the VHδ–Dδ–Jδ–Cδ cluster was trans-located out of the TCRα/δ locus where it currently resides on another chromosome (fig. 5D). There are no Vα or Vδ genes at the site of the second chicken TCRδ locus and only a single Cδ gene remains in the conventional TCRα/δ locus (Parra et al. 2012).

5) Similar to galliform birds, the VHδ–Dδ–Jδ–Cδ cluster was trans-located out of the TCRα/δ locus in presumably the last common ancestor of mammals, giving rise to TCRμ (fig. 5E). Internal duplications of the VHδ–Dδ–Jδ genes gave rise to the current [(V–D–J) — (V–D–J) — C] organization necessary to encode TCR chains with double V domains (Parra et al. 2007, Wang et al. 2011). In the platypus, the second V–D–J cluster, encoding the supporting V, has lost its D segments and generates V domains with short CDR3 encoded by direct V to J recombination (Wang et al. 2011). The whole cluster appears to have undergone additional tandem duplication as it exists in multiple tandem copies in the opossum and also likely in the platypus (Parra et al. 2007, 2008; Wang et al. 2011).

6) In the therian lineage (marsupials and placentals), the VHδ was lost from the TCRα/δ locus (Parra et al. 2008). In placental mammals, the TCRμ locus was also lost (Parra et al. 2008). The marsupials retained TCRμ, however the second set of V and J segments, encoding the supporting V domain in the protein chain, were replaced with a germ-line joined V gene, in a process most likely involving germ-line V(D)J recombination and retro-transposition (fig. 5F) (Parra et al. 2007, 2008).

TCR forms such as TCRμ, which contain three extracellular domains, have evolved at least twice in vertebrates. The first was in the ancestors of the cartilaginous fish in the form of NAR-TCR (Criscitiello et al. 2006) and the second in the mammals as TCRμ (Parra et al. 2007). NAR-TCR uses an N-terminal V domain related to the V domains found in IgNAR antibodies, which are unique to cartilaginous fish (Greenberg et al. 1995; Criscitiello et al. 2006), and not closely related to antibody VH domains. Therefore, it appears that NAR-TCR and TCRμ are more likely the result of convergent evolution rather than being related by direct descent (Parra et al. 2007; Wang et al. 2011). Similarly, the model proposed in fig. 5 posits the direct transfer of VH genes from an IgH locus to the TCRα/δ locus. But it should be pointed out the VHδ found in frogs, birds, and monotremes are not closely related (fig. 3); indeed, they appear derived each from different, ancient VH clans (birds, VH clan I; frogs VH clan II; platypus VH clan III). This observation would suggest that the transfer of VHδ into the TCRα/δ loci occurred independently in the different lineages. Alternatively, the transfer of VH genes into the TCRα/δ locus may have occurred frequently and repeatedly in the past and gene replacement is the best explanation for the current content of these genes in the different tetrapod lineages. The absence of VHδ in marsupials, the highly divergent nature of Vμ genes in this lineage, and the absence of conserved synteny with genes linked to TCRμ in the opossum, provide little insight into the origins of TCRμ and its relationship to TCRδ or the other conventional TCR (Parra et al. 2008). The similarity between VH, VHδ, and Vμ genes in the platypus genome, which are all clan III, however is striking. In particular, the close relationship between the platypus VHδ and Vμ genes lends greater support for the model presented in fig. 5E, with TCRμ having been derived from TCRδ genes.
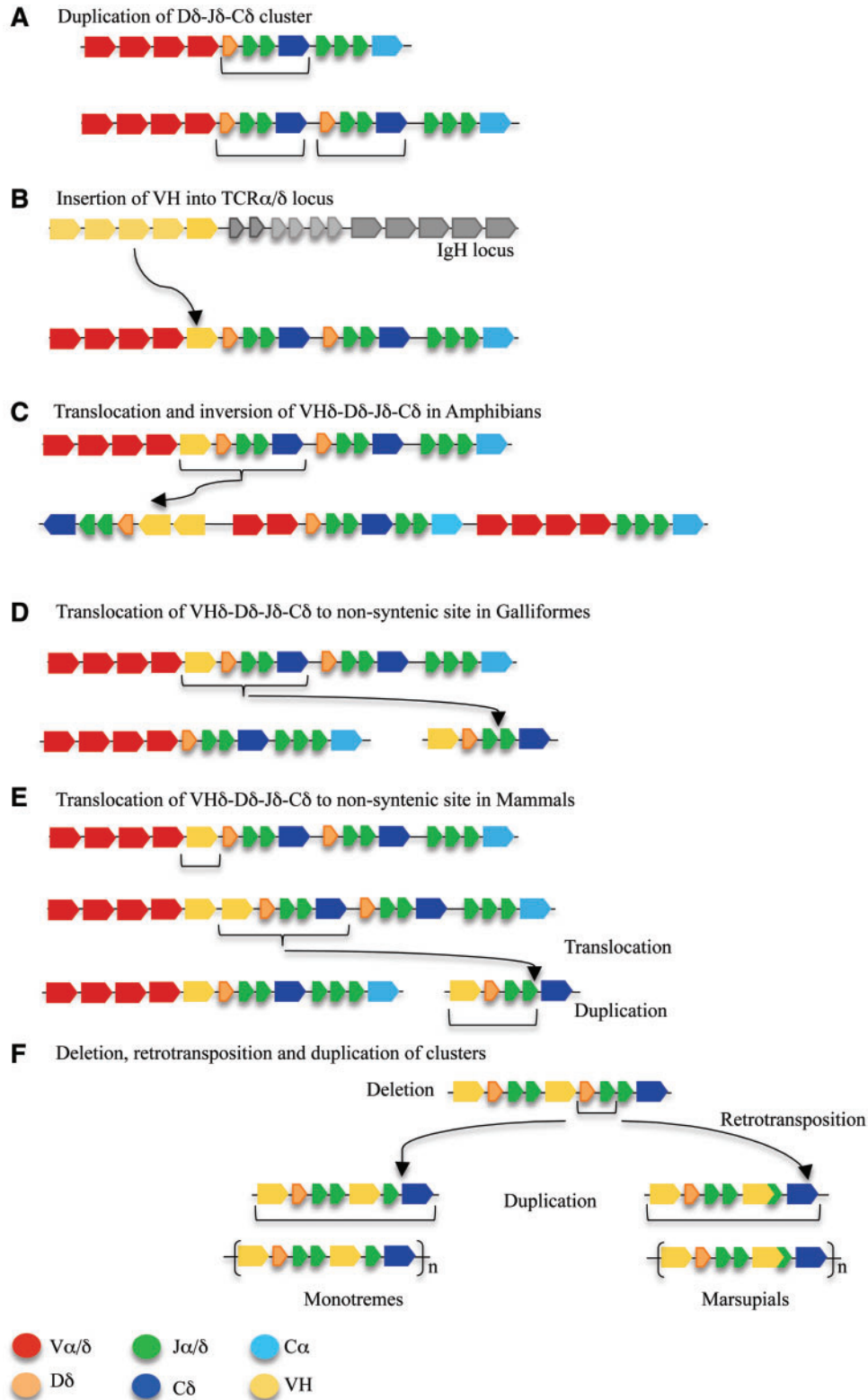
Fig. 5. A model of the stages of evolution of the *TCRα/δ* loci in tetrapods and the origins of TCRµ in mammals. A color key of the gene segments is presented at the bottom. (*A*) Depiction of the Dδ-Jδ-Cδ duplication in an ancestral *TCRα/δ* locus that provides a second Cδ gene found in frogs and zebra finch. (*B*) Depiction of the insertion of a VH gene into the *TCRα/δ* locus producing a current organization as it is found in zebra finch. (*C*) Depiction of the inversion/translocation and VHδ gene duplication that yielded the current organization found in frogs. (*D*) Depiction of the translocation of a VHδ–Dδ–Jδ–Cδ cluster to a location outside the *TCRα/δ* locus generating a second TCRδ locus as it is currently found in chicken and turkey. (*E*) Depiction the translocation that took place in mammals giving rise to the TCRµ locus. (*F*) Loss of TCRµ in placental mammals, loss of D gene segments in cluster encoding the support V domain, retro-transpostion to form a germ-line joined V in marsupials, and duplication of TCRµ clusters in both monotremes and marsupials.

The presence of TCR chains that use antibody like V domains, such as TCRδ using VHδ, NAR-TCR or TCRμ are widely distributed in vertebrates with only the bony fish and placental mammals missing. In addition to NAR-TCR, some shark species also appear to generate TCR chains using antibody V genes. This occurs via trans-locus V(D)J recombination between IgM and IgW heavy chain V genes and TCRδ and TCRα D and J genes (Criscitiello et al. 2010). This may be possible, in part, due to the multiple clusters of Ig genes found in the cartilaginous fish. It also illustrates that there has been independent solutions to generating TCR chains with antibody V domains in different vertebrate lineages. In the tetrapods, the VH genes were trans-located into the TCR loci where they became part of the germ-line repertoire. Whereas in cartilaginous fish something equivalent may occur somatically during V(D)J recombination in developing T cells. Either mechanism suggests there has been selection for having TCR using antibody V genes over much of vertebrate evolutionary history.

The current working hypothesis for such chains is that they are able to bind native antigen directly. This is consistent with a selective pressure for TCR chains that may bind or recognize antigen in ways similar to antibodies in many different lineages of vertebrates. In the case of NAR-TCR and TCRμ, the N-terminal V domain is likely to be unpaired and bind antigen as a single domain (fig. 1), as has been described for IgNAR and some IgG antibodies in camels (recently reviewed in Flajnik et al. 2011). This model of antigen binding is consistent with the evidence that the N-terminal V domains in TCRμ are somatically diverse, while the second, supporting V domains have limited diversity with the latter presumably performing a structural role rather than one of antigen recognition (Parra et al. 2007; Wang et al. 2011). There is no evidence of double V domains in TCRδ chains using VHδ in frogs, birds, or platypus (fig. 1) (Parra et al. 2010, 2012). Rather, the TCR complex containing VHδ would likely be structured similar to a conventional γδTCR with a single V domain on each chain. It is possible that such receptors also bind antigen directly, however this remains to be determined.

A compelling model for the evolution of the Ig and TCR loci has been one of internal duplication, divergence and deletion; the so-called birth-and-death model of evolution of immune genes promoted by Nei and colleagues (Ota and Nei 1994; Nei et al. 1997). Our results in no way contradict that the birth-and-death mode of gene evolution has played a significant role in shaping these complex loci. However, our results do support the role of horizontal transfer of gene segments between the loci that has not been previously appreciated. With this mechanism T cells may have been able to acquire the ability to recognize native, rather than processed antigen, much like B cells.

Supplementary Material

Supplementary tables S1 and S2 are available at *Molecular Biology and Evolution* (http://www.mbe.oxfordjournals.org/).

References

Baker ML, Belov K, Miller RD. 2005. Unusually similar patterns of antibody V segment diversity in distantly related marsupials. *J Immunol.* 174:5665–5671.

Belov K, Hellman L. 2003. Immunoglobulin genetics of *Ornithorhynchus anatinus* (platypus) and *Tachyglossus aculeatus* (short-beaked echidna). *Comp Bioch Phys Part A.* 136:811–819.

Butler JE. 1997. Immunoglobulin gene organization and the mechanism of repertoire development. *Scand J Immunol.* 45:455–462.

Carroll AM, Slack JK, Mu X. 1993. V(D)J recombination generates a high frequency of nonstandard TCR D delta-associated rearrangements in thymocytes. *J Immunol.* 150:2222–2230.

Chen H, Kshirsagar S, Jensen I, Lau K, Covarrubias R, Schluter SF, Marchalonis JJ. 2009. Characterization of arrangement and expression of the T cell receptor gamma locus in the sandbar shark. *Proc Natl Acad Sci U S A.* 106:8591–8596.

Chien YH, Iwashima M, Kaplan KB, Elliott JF, Davis MM. 1987. A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation. *Nature* 327:677–682.

Criscitiello MF, Ohta Y, Saltis M, McKinney EC, Flajnik MF. 2010. Evolutionarily conserved TCR binding sites, identification of T cells in primary lymphoid tissues, and surprising trans-rearrangements in nurse shark. *J Immunol.* 184:6950–6960.

Criscitiello MF, Saltis M, Flajnik MF. 2006. An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. *Proc Natl Acad Sci U S A.* 103:5036–5041.

Davis MM, Chein YH. 2008. T cell antigen receptors. In: Paul WE, editor. Fundamental immunology, 6th ed. Philadelphia: Lippincott. p. 313–345.

Flajnik MF, Deschacht N, Muyldermans S. 2011. A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? *PLoS Biol.* 9:e1001120.

Flajnik MF, Kasahara M. 2010. Origin and evolution of adaptive immune system: genetic events and selective pressures. *Nat Rev Gen.* 11: 47–59.

Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. 1995. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* 374: 168–173.

Giudicelli V, Chaume D, Lefranc MP. 2005. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* 33:D256–D261.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.

Hawwari A, Krangel MS. 2007. Role for rearranged variable gene segments in directing secondary T cell receptor alpha recombination. *Proc Natl Acad Sci U S A.* 104:903–907.

Hayday AC. 2009. gd T cells and the lymphoid stress-surveillance response. *Immunity* 31:184–196.

Johansson J, Aveskogh M, Munday B, Hellman L. 2002. Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus*

anatinus): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. *J Immunol.* 168:5155–5162.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 5:150–163.

Lambeck K, Chappell J. 2001. Sea level change through the last glacial cycle. *Science* 292:679–686.

Lewis SM. 1994. P nucleotide insertions and the resolution of hairpin DNA structures in mammalian cells. *Proc Natl Acad Sci U S A.* 91: 1332–1336.

Litman GW, Rast JP, Fugmann SD. 2010. The origins of vertebrate adaptive immunity. *Nat Rev Immunol.* 10:543–553.

McCormack W, Tjoelker L, Thompson C. 1991. Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion. *Ann Rev Immunol.* 9:219–241.

Nei M, Gu X, Sitnikova T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci U S A.* 94:7799–7806.

Ota T, Nei M. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol.* 11:469–482.

Parra ZE, Baker ML, Hathaway J, Lopez AM, Trujillo J, Sharp A, Miller RD. 2008. Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica. BMC Genomics* 9:111.

Parra ZE, Baker ML, Schwarz RS, Deakin JE, Lindblad-Toh K, Miller RD. 2007. A unique T cell receptor discovered in marsupials. *Proc Natl Acad Sci U S A.* 104:9776–9781.

Parra ZE, Mitchell K, Dalloul RA, Miller RD. 2012. A second TCRδ locus in Galliformes uses antibody-like V domains: insight into the evolution of TCRδ and TCRμ genes in tetrapods. *J Immunol.* 188: 3912–3919.

Parra ZE, Ohta Y, Criscitiello MF, Flajnik MF, Miller RD. 2010. The dynamic TCRδ: TCRδ chains in the amphibian *Xenopus tropicalis* utilize antibody-like V genes. *Eur J Immunol.* 40: 2319–2329.

Rast JP, Anderson MK, Strong SJ, Luer C, Litman RT, Litman GW. 1997. Alpha, beta, gamma, and delta T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* 6:1–11.

Rock EP, Sibbald PR, Davis MM, Chien YH. 1994. CDR3 length in antigen-specific immune receptors. *J Exp Med.* 179:323–328.

Satyanarayana K, Hata S, Devlin P, Roncarolo MG, De Vries JE, Spits H, Strominger JL, Krangel MS. 1988. Genomic organization of the human T-cell antigen-receptor alpha/delta locus. *Proc Natl Acad Sci U S A.* 85:8166–8170.

Schatz DG, Oettinger MA, Baltimore D. 1989. The V(D)J recombination activating gene, RAG-1. *Cell* 59:1035–1048.

Schwager J, Burckert N, Courtet M, Du Pasquier L. 1989. Genetic basis of the antibody repertoire in Xenopus: analysis of the VH diversity. *EMBO J.* 8:2989–3001.

Sciammas R, Johnson RM, Sperling AI, Brady W, Linsley PS, Spear PG, Fitch FW, Bluestone JA. 1994. Unique antigen recognition by a herpesvirus-specific TCR-gamma delta cell. *J Immunol.* 152: 5392–5397.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24:4876–4882.

Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* 302: 575–581.

Tutter A, Riblet R. 1989. Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. *Proc Natl Acad Sci U S A.* 86:7460–7464.

Vernersson M, Aveskogh M, Munday B, Hellman L. 2002. Evidence for an early appearance of modern post-switch immunoglobulin isotypes in mammalian evolution (II); cloning of IgE, IgG1 and IgG2 from a monotreme, the duck-billed platypus, Ornithorhynchus anatinus. *Eur J Immunol.* 32:2145–2155.

Wang K, Klotz JL, Kiser G, Bristol G, Hays E, Lai E, Gese E, Kronenberg M, Hood L. 1994. Organization of the V gene segments in mouse T-cell antigen receptor alpha/delta locus. *Genomics* 20:419–428.

Wang X, Parra ZE, Miller RD. 2011. Platypus TCRμ provides insight into the origins and evolution of a uniquely mammalian TCR locus. *J Immunol.* 187:5246–5254.

Warren WC, Hillier LW, Marshall Graves JA, et al. 2008. (78 co-authors). (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–183.

Wysocki L, Manser T, Gefter ML. 1986. Somatic evolution of variable region structures during an immune response. *Proc Natl Acad Sci U S A.* 83:1847–1851.

Yancopoulos GD, Blackwell TK, Suh H, Hood L, Alt FW. 1986. Introduced T cell receptor variable region gene segments recombine in pre-B cells. Evidence that B and T cells use a common recombinase. *Cell* 44:251–259.

Yazawa R, Cooper GA, Beetz-Sargent M, Robb A, McKinnel L, Davidson WS, Koop BF. 2008. Functional adaptive diversity of the Atlantic salmon T-cell receptor gamma locus. *Mol Immunol.* 45:2150–2157.