Supporting Information

Edholm et al. 10.1073/pnas.1309840110

SI Materials and Methods

Experimental Animals. Transgenic *Xenopus laevis* was generated using I-SecI meganuclease as previously described (1). Briefly, the I-SecI-*Xenopus* nonclassical gene 10 (XNC10) shRNA-GFP expression vector was constructed by cloning the GFP reporter flanked by the 18-bp I-SecI recognition sites into the I-SceIpBSIISk+ vector (provided by R. Grainger, University of Virginia, Charlottesville, VA). Subsequently, the XNC10 shRNA under the control of the hU6 Pol III promoter was cloned into the I-SecI-GFP vector. Before microinjection, females were primed with human chronic gonadotropin (Sigma), and eggs were fertilized, dejellied, and injected with 80 pg I-SceI meganuclease-digested XNC10shRNA-GFP expression vector as previously described (1). After injection, all embryos were incubated at 13 °C for 4 h to delay cell division, transferred to 0.3× modified Barth's saline supplemented with 5 μg/mL gentamicin, and reared at 18 °C until hatching. Transgenic larvae were screened for GFP expression at developmental stage 56 using an SMZ1500 Nikon stereomicroscope.

Construction and Production of b2m-XNC10 Tetrameric Complexes. Beta 2 microglobulin (b2m) was engineered with a 23-aa Gly-rich C-terminal flexible linker (GGGGSGGSGSGGGSSGTELGSGS). Soluble XNC10 encompassing the α1-α3 domains and lacking the transmembrane segment and cytoplasmic domain was engineered with a substrate sequence (GLNDIFEAQKIEWH) for BirA site-specific biotinylation at the end of the α3 domain. The b2m-linker-XNC10 construct was generated by sequential cloning into the pMIBV5-HisA expression vector (Invitrogen) with a polyhistidine region and a V5 epitope tag at the C terminus. It was transfected into Sf9 insect cells using the Cellfectin reagent (Invitrogen) according to the manufacturer's protocol. After 2 wk of Blasticidin selection, stably transfected cells were expanded for large-scale protein production. Monomeric b2m-linker-XNC10 (b2m-linker-XNC10) was purified by Ni-NTA-Agarose Chromatography (Qiagen) using high-stringency conditions (50 mM NaH_2PO_4, 500 mM NaCl, 100 mM Imidazole, 0.5% Tween, pH 8.0), eluted in 50 mM NaH_2PO_4, 500 mM NaCl, and 250 mM Imidazole, pH 8.0, dialyzed against 10 mM Tris·HCl, pH 7.4, and concentrated to 1 μg/μL using Amicon Ultra Centrifugal Filter (Millipore). BirA enzymatic biotinylation was performed for 16–18 h at 30 °C according to the manufacturer's protocol (Avidity), and the purified biotinylated proteins were extensively dialyzed against amphibian PBS, pH 7.5, to remove any unbound biotin. The specificity and degree of biotinylation were verified by Western blot analysis using HRP-conjugated streptavidin (Pierce). XNC10 tetramers were generated by incubating b2m-linker-XNC10 with fluorochrome-labeled streptavidin at a 5:1 ratio at room temperature for 4 h before use in flow cytometry analysis.

Conditions for Flow Cytometry. For flow cytometry, cells (0.25×10^6) were stained with 5 μg XNC10-tetramer (XNC10-T) -allophycocyanin (APC) for 30 min at 4 °C followed by *Xenopus* anti-CD8 (AM22), -CD5 (pan T-cell marker 2B1), or -MHC class II (AM20) mAbs and FITC-conjugated goat anti-mouse secondary Abs (Southern Biotech); 50,000 events were collected with an FACSCanto II (BD Bioscience) and analyzed using FlowJo software (Tree Star Inc.). Dead cells were excluded using Propidium iodide (BD Pharmingen). Splenocytes were FACS sorted (F-Aria-18; BD Bioscience), and subsamples from each sorted fraction were analyzed by flow cytometry to verify sorting purity.

Conditions for Quantitative PCR, RT-PCR, and 5′-RACE. RNA was prepared using TRIzol Reagent (Invitrogen) and treated with DNase (Ambion; Life Technologies) according to the manufacturer's protocol; 500 ng total RNA were transcribed into cDNA with iScript reverse transcriptase using a mix of oligo-dT and random hexamer primers (Bio-Rad). Typical parameters for RT-PCR were 5 min at 94 °C followed by 30 cycles of 94 °C for 30 s, 59 °C for 30 s, and 72 °C for 1 min with a final extension cycle of 72 °C for 10 min. For nested reactions, 1 μL initial PCR reaction was used as a template in the subsequent PCR. Quantitative PCR parameters were 2 min at 95 °C followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Quantitative PCR gene expression was performed using the ΔΔCT method with the ABI 7300 Real-Time PCR System and PerfeCta SYBR Green FastMix ROX. Expression analyses of XNC10 and invariant T cell receptor (iTCR)α chain were normalized to the GAPDH endogenous control against the lowest observed tissue expression. Analysis was performed using the ABI sequence detection system software, and all primers were validated before use (Table S7). PCR products were cloned into pGEM-T (Promega) and sequenced. The IMGT/DomainGapAlign TOOL (http://imgt.cines.fr) or when possible, IMGT/V-QUEST (2) was used to determine complementarily determining regions (CDRs), framework, and strand delimitations. For 5′-rapid amplification of cDNA ends (RACE) analysis, RNA was isolated with the SMARTer RACE cDNA Amplification Kit (Clontech Laboratories) and amplified with either a T cell receptor (TCR)α or TCR constant region-specific primers.

TCRα Repertoire Sequence Analysis. The 454 GS junior reads were split by barcode identification into three populations using the fastx_barcode_splitter from fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Any reads <200 bp in length were removed using the fastq_quality_trimmer from the fastx-toolkit with the following parameter: -t 0 –l 200. The remaining reads were screened for the anchoring Cα region using BLASTn version 2.2.27 (3) with the following parameters: -outfmt 6 -evalue 1e-5 -perc_identity 95. To determine the sequence identity, known Vα and Jα sequences were used as the query in a BLAST search with the following parameters: -outfmt 6 -evalue 1e-10 -perc_identity 95, where the Cα-anchored region and size-selected reads were the subject database. Initial clustering of the selected sequences was performed using uSEarch version 6.0.307 with a >85% cutoff (4). A Perl script was used to correlate the Vα and Jα BLAST identification of each sequence and the clusters obtained with uSearch; therefore, the frequency of each Vα and Jα region was recorded for each cluster. All major clusters were then reclustered with 100% and >95% identify thresholds and correlated with the Vα and Jα BLAST identification using the same Perl script described.

1. Nedelkovska H, Edholm ES, Haynes N, Robert J (2013) Effective RNAi-mediated β2-microglobulin loss of function by transgenesis in Xenopus laevis. *Biol Open* 2(3):335–342.
2. Lefranc MP, et al. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27(1):55–77.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
4. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460–2461.

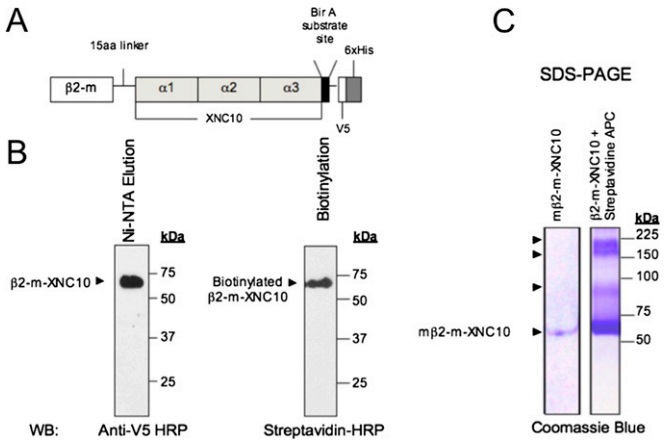

Fig. S1. Design and production of an *X. laevis* class Ib XNC10 tetramer. (*A*) Schematic representation of the *b2m*-linker-XNC10 construct used to generate 2-m-XNC10 monomers. (*B*) *b2m*-XNC10 protein was purified from concentrated SF9 insect culture supernatant under native conditions using Ni-NTA-agarose; the eluted proteins were separated by reducing 10% (vol/vol) SDS/PAGE, transferred to nitrocellulose membranes, and visualized by Western blot using anti–V5-HRP. Eluted 2-m-XNC10 protein was dialyzed against 10 mM Tris·HCl, subjected to Bir A enzymatic biotinylation, dialyzed extensively against amphibian PBS, and analyzed, except where visualized using streptavidine-HRP. (*C*) Monomeric biotinylated 2-m-XNC10 was concentrated to 1 g/L, incubated with streptavidin-APC at a 5:1 molar ratio, and run on an 8% SDS/PAGE gel under nonreducing conditions. Proteins were visualized using Coomassie Blue stain. Arrowheads indicate appropriately sized 2-m-XNC10 proteins. WB, western blot.

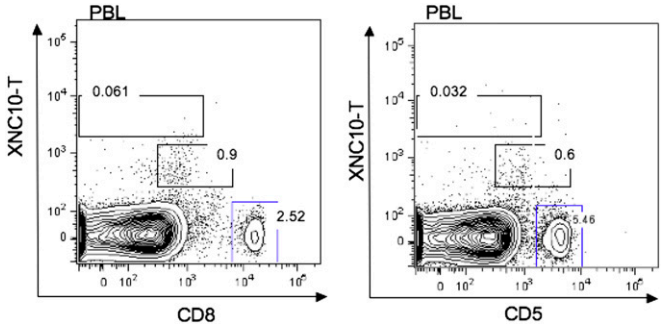

Fig. S2. Identification of XNC10-T$^+$ T-cell populations in the peripheral blood leukocytes (PBL) of adult *X. laevis*. Flow cytometry of live PBL isolated from adult unprimed *X. laevis* and double stained with streptavidine-APC–conjugated XNC10-T and CD8 or CD5-specific mAbs. Percentage of positive staining cells is indicated in the scatterplot. Data are representative of three independent experiments.

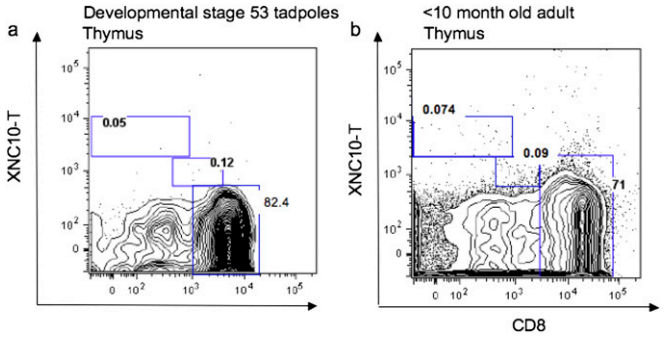

Fig. S3. No detectable XNC10-T$^+$–cell populations in the thymus of either larval or adult *X. laevis*. Flow cytometry of thymocytes either (*A*) pooled from four stage 53 tadpoles or (*B*) isolated from a single adult unprimed *X. laevis*. Cells were double stained with streptavidine-APC–conjugated XNC10-T and CD8-specific mAb and gated on live cells. Percentage of positive staining cells is indicated in the scatterplot. Data are representative of three independent experiments.

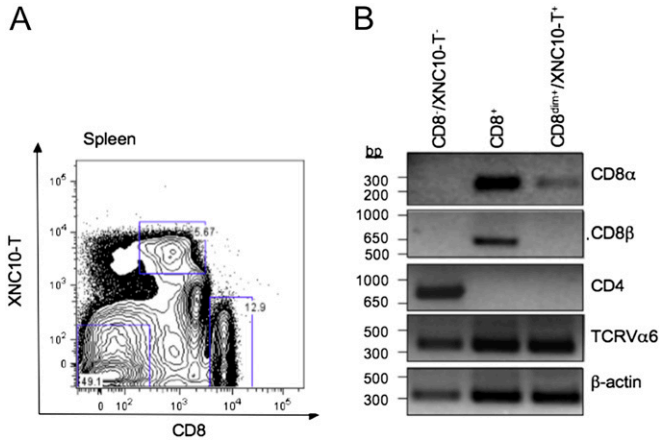

Fig. S4. Identification of CD8 expressing type II XNC10-T^{+} cells. (*A*) Flow cytometry of spleen leukocytes isolated from adult unprimed *X. laevis* (1 y), gated on live cells, and double stained with streptavidine-phycoerythrin–conjugated XNC10-T and CD8 mAb. Sorting gates and percentages of cells are indicated in the plot. (*B*) Gene expression profiles of sorted XNC10-T$^{+/}$CD8^{dim+} cells, CD8^{+} and XNC10-T^{-} CD8^{-} double negative cells from spleen leukocytes, actin, and loading control: CD8α, CD8β, CD4, and TCRVα6.

```
                                                  CDR1                                CDR2
Vb6-a   ------SAQQVKVIQDQKFLAVTLGSSVKVQCSHTGPSYYAMLWYQQKLGQGLKLMVLSSDT--
Vb6-b   ------SAQQVKVIQDQKFLAVTLGSSVKVQCSHTGPSYYAMFWYQQKLGQGLKLMVLSSDT--
Vb6-c   ------SAQQVKVIQDQKFLAVTLGSSVKVQCSHTGPSYYAMLWYQQKLGQGLKLMVLSSDT--
Vb7-a   ------LTQLINVTQDYRFLAVPERSPVKLHCSHSASGFYSMFWYQQKLGQGLTLMVFSSDT--
Vb7-b   ------LTQLINVTQDYRFLAVPERSPVKLHCSHSASGFYSMFWYQQKLGQGLTLMVFSSDT--
Vb10    ---------GVVVTQEKKLIVAHHGKSLAIPCQHDATDYLNMLWYQQRPGKELKLLAISAGKG-
Vb13    ------VCQGVLVKQEKRLVLVPFDQSADIPCNQDQSTHLNMYWYQQNPGEGLKLMIYSTNEK-
vb14    -------CLAAPVVQYPTVQLSHPGDNVTVQCEHNDGSYTRKYWYRQENNGPLTLLGSTYTTQE
vb19    ----SGLCQDGRVVQVPLLLVARVGAEMSLRCRVTDSSLSNTLWYRQGGDGPLTLIGHTYRGGT
                  * *                 *          ** *     * *

                                                                  CDR3       frequency
Vb6-a   KPGNMEEGFKE--WSMERPDTQTSDLSISQVGAEHSAVYFCAASEG----QDYQAYFGDG 60%
Vb6-b   KPGNMEEGFKE--WSMERPDTQKSDLSISQVGAEHSAVYFCAASDIGGQNSGYAQTFGDG 2.8%
Vb6-c   KPGNMEEGFKE--WSMERPDTQASDLSISQVGAEHSAVYFCAASDGG--PLTERQYFGDG 2.8%
Vb7-a   KPGNMEEGFKE--WSMERPDTQTSDLSISQVGAEHSAVYFCAASEGQ----DYQAYFGDG 2.8%
Vb7-b   KPGNMEEGFKE--WSMERPDTQNSYLSISQVGSEHSAVYFCAASPGG----VDRQYFGEG 2.8%
Vb10    -DTTIENEFKGK-WTMERPDVLNSTLNRDKVDGEDSAVYYCASSHRQ--LSGNKVTFGEG 2.8%
Vb13    -EEKMEKEYEER-WSLNRPDIYNSVLRLKEAKMEDAAQYFCASQSSQ--GN-TAAYFGEG 20%
vb14    --ATVEKDFVGGRIIIQHKTAQGAVLIILRVSPQDSALYYCASS-GR--EGQNRLYFGDG 2.8%
vb19    --PSIESSFSGTNITIRAESTDSSLLLISGVSGQHSGTYFCAGSIDR--GNAAVQYFGDG 2.8%
            *                 *           * **              ** *
```

Fig. S5. TCR rearrangements isolated from XNC10-T^{+} populations. Amino acid alignment of TCR rearrangements isolated from XNC10-T^{+} populations sorted from adult spleen. CDRs are underlined and indicated at the top of the sequence. Conserved Cys-23 and -104 are highlighted in gray; the glycine bulge motif in the J region is bold, and residues conserved in all sequences are indicated by asterisks. The frequency of each V-D-J rearrangement in the XNC10-T^{+} population is indicated at the end of the sequence.

a

```
                 Vα6
19-iVα6          TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
123              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
203              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
143              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
173              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
131              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
151              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
31               TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
181              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
21               TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
162              TTCCATCTCTGTAAAGGAGAAGCCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCA
                 ************************************************************
                  < J
19-iVα6          AG--TGATACTGGGGGTTCCACC--AAACTCACATTTGGGAAAGGAACAAAACTGACAGT
123              A-----GTGCTGT-----CAAC---AAGCTTATGTTTGGCAAAGGGACCCAGCTAATTGT
203              AG--TACTGGA-ACAG-GTGGCTTAAAAAATATTTTTGGCAGGGGAACAAAGCTTAATGT
143              AG------TGCTG----TTAA-CAAGCT--TATGTTTGGCAAAGGGACCCAGCTAATTGT
173              AG------TGCTG----TCAA-CAAGCT--TATGATTGGCAAAGGGACCCAGATAATTGT
131              AGGATGAATATGGATCTTCAGTA--TAACTTGCCTTTGGAAAAGGAACAACGCTGTCAGT
151              AGGATGAATATGGATCTTCGGTA--TAACTTGCCTTTGGAAAAGGAACAACGCTGTAAGT
31               ATACTGGTTCAGG-----CAACTGGAGAATTATATTTGGCAGGGGCACTAGACTTTCTGT
181              ATACTGGTTCAGG-----CAACTGGAGAATTATATTTGGCAGGGGCACTAGACTTTCTGT
21               ATACTGGTTCAGG-----CAACTGGAGAATTATATTTGGCAGGGGCACTAGACTTTCTGT
162              AG----------------------------------------------------------
                 *
                         < C region
19-iVα6          TCTACCAAACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
123              GATGCCAAACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
203              AATTGCAGACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
143              GATGCCAAACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
173              GATGCCAAACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
131              GACAGCTGACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
151              GACAGCTGACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
31               GACTCCAGACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
181              GACTCCAGACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGA-
21               GACTCCAGACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
162              --------ACATTAAAGATGTAGAACCTTCAATGTATCGCCTCAAAGCTGAC
                         ******************************************
```

b

```
24-iVα6          GAGAATTTACAG-TGAA---AATTGAAAAGAAAGAT--TTCCATCTCTGTAAAGGAGAAG
203              GAGAATTTACAG-TGAA---AATTGAAAAGAAAGAT--TTCCATCTCTGTAAAGGAGAAG
231              GAGAATTTACAG-TGAA---AATTGAAAAGAAAGAT--CTCCATCTCTGTAAAGGAGAAG
261              GAGAATTTACAG-TGAA---AATTGAAAAGAAAGAT--TTCCATCTCTGTAAAGGAGAAG
274              AAGGATTTTCAGCTGAGCACAAGAGAAGTGAAACATCTTTCCATCTGAAAAAAGACAAAG
184              AAGGATTTTCAGCTGAGCACAAGAGAAGTGAAACATCTTTCCATCTGAAAAAAGACAAAG
                 ** **** *** ***    **  ***  **** **   *******    ****   ***

24-iVα6          CCAAAGTGACAGACTCTGCCGCGTATTTCTGTGCTGCAAGTGATACTGGGGGTTCCACCA
203              CCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCAAGTGATGGTTCGGGTTATAACA
231              CCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCAA----TACCAATAACTACA---
261              CCAAAGTGACAGACTCTGCCGTGTATTTCTGTGCTGCAA----TACCAATAACTACA---
274              CTGAACTGCAAGACTCCGGGGGTGTATTTCTGTGCTGTCAGTGGT---TCGGGTTATAACA
184              CTGAACTGCAAGACTCCGGGGGTGTATTTCTGTGGTGTCAGTGATATACCAACTTG-GCTA
                 *  ** **  ****** *  * *********** **  *    *         *
```

Fig. S6. (Continued)

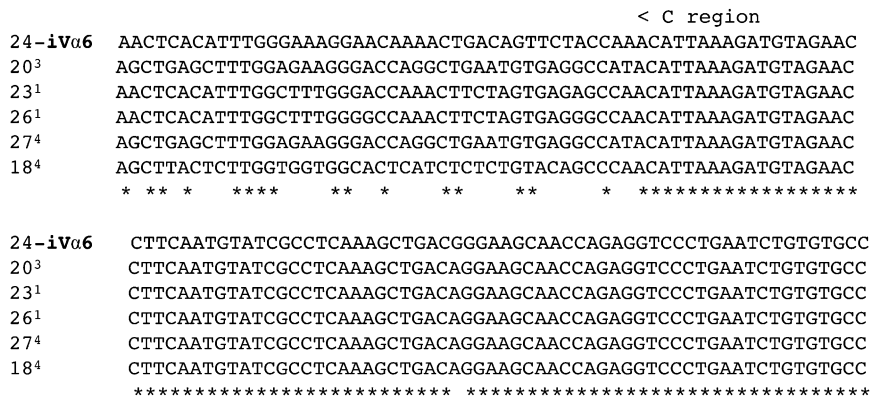

Fig. S6. Nested RT-PCR using Vα6-specific forward primers combined with TCRα C-region reverse primers revealed a high number of nonproductive rearrangements in the spleen of developmental stage 53 XNC10shRNA Tg tadpoles. Nucleotide alignment of sequences from (*A*) XNC10Tg-1 and (*B*) XNC10Tg-2. Vα6 sequence is shown in bold, and nucleotides conserved in all sequences are indicated by asterisks. J and C regions are indicated above the sequence. Nonproductive rearrangements because of [1]insertion or deletion of nucleotide resulting in a frame shift in the CDR3a junctional region, [2]no detectable J region, or [4]pseudononfunctional Va6 gene are indicated. [3]Productive rearrangement of Vα6 to a J region other than Jα1.43. Control sequences (n = 20) from two age-matched dejellied controls were all productive, with 30%/15% iVα6-Jα1.43 and 70%/85% Vα6 rearranged to other Jα regions in outbred animals 1 and -2, respectively.

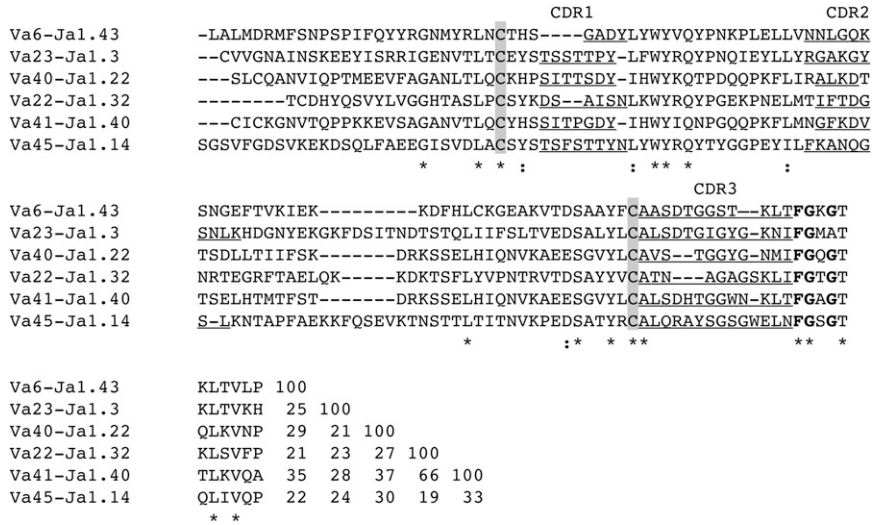

Fig. S7. The dominant invariant TCR⟨ rearrangements are distinct. Amino acid alignment of dominant TCR⟨ rearrangements isolated from $CD8^{-}$ and $CD8^{dim+}$ populations sorted from stage 50 tadpoles. CDRs are underlined and indicated at the top of the sequence. Conserved Cys-23 and -104 are highlighted in gray, the glycine bulge motif in the J region is bold, residues conserved in all sequences are indicated by asterisks, and conserved substitutions are indicated at the bottom of the sequence. Percent amino acid identity based on pairwise alignments is indicated at the end of the sequences.

Table S1. TCR Vα-Jα and Vβ-Dβ-Jβ genes use by XNC10-T^{dim+} type II cells

Vα	n-nucleotides		Jα	CDR3	Amino acids	Clones
TCR Vα-Jα gene use by XNC10-T^{dim+} type II cells						
Vα6						
CAAS	—		DT**G**GST**KL**TFGKGTKLTVL	Jα1.43	12	15
Vα15.2						
CAVS	—		YS**G**NNL**KL**IFGSGTQLIVH	Jα1.20	12	2
Vα3.5						
CAV	—		EFS**GSKL**IFGSGTELIVQ	Jα1.10	10	1
Vα28						
CAVS	—		YS**G**SGWE**L**NFGSGTQLIVQ	Jα1.14	12	1
Vα14						
CGVR	*g*		Y**G**ASD**KL**VFGKGTILSVTA	Jα1.47	11	1
Vα46.1						
CAVS	*gt*		T**G**GWKYVFGSGTRLKVL	Jα1.11	11	1
CAFRDE	*ggg*		AGSVG**KL**IFGSGTRLIIQ	Jα1.18	14	3
Vα3.1						
CAVSA	*gca*		YGS**KL**TMGLGTVLFVR	Jα1.53	11	2
Vβ	n-Dβ-n		Jβ	CDR3	Amino acids	Clones
TCR Vβ-Dβ-Jβ gene use by XNC10-T^{dim+} type II cells						
Vβ1.1						
CAATS	*cgggacaggggggg*	Dβ2	TNERQYFGDGTILTVL	Jβ8	15	1
Vβ1.2						
CAA	*ccactggag*	Dβ1	TAAYFGEGTVLTVL	Jβ3	9	1
Vβ2						
CAWS	*agaggacaggggga*	Dβ2	NYQAYFGDGTILTVL	Jβ13	13	
Vβ3						
CAWS	*cgcgacaggc*	Dβ2	TADRLYFGDGTILTVL	Jβ9	12	4
Vβ6						
CAASE	*gggacagg*	Dβ2	DYQAYFGDGTILTVL	Jβ13	11	9
Vβ9						
CASSH	*ccgacaact*	Dβ2	SGNKVTFGEGTILHVL	Jβ2	13	3
CAVS	*gagacagggtggg*	Dβ2	DRQYFGEGTILTVL	Jβ8	12	1
Vβ16						
CAASK	*gggacat*	Dβ2	TTERQYFGDGTILTVL	Jβ6	12	2
Vβ18						
CTSS	*acgacaga*	Dβ2	NYQAYFGDGTILTVL	Jβ13	11	2
CTSS	*taggggggggtg*	Dβ2	ANTAAYFGEGTVLTVL	Jβ3	13	1
CTSS	*taggggggggtg*	Dβ2	AAYFGEGTVLTVF	Jβ3	13	1
Vβ20						
CAAS	*gggggggggg*	Dβ2	DYQAYFGDGTILTVL	Jβ13	11	1
CASSI	*cgcgacaggc*	Dβ2	TERQYFGDGTILTVL	Jβ6	13	1

RNA from sorted type II XNC10-T^{dim+} cells was subjected to 5′-RACE using TCRα and -β constant region primers; 26 TCRα clones and 30 TCRβ clones were sequenced. Nongerm line-encoded nucleotides are shown. Residues in the CDR3α region conserved with the invariant Vα6-Ja1.43 rearrangement (Gly-110, Lys-115, and Leu-116) are bold. The CDR3 region is defined according to the International Immunogenetics numbering (1). Number of clones signifies the number of sequences with identical nucleotide sequences.

1. Lefranc MP, et al. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27(1):55–77.

Table S2. TCR Vα-Jα gene usage by CD8$^+$ cells

Vα	n-nucleotides	Jα	CDR3	(aa)	Clones
Vα25					
CALRS	-	DGYKYTFGSGTQLIVI	Jα1.16	10	2
CALRS	*ct*	NYGASNKLIFGPGTKINVK	Jα1.2	14	1
CALRR	-	DTGGRALTFGKGTQINVM	Jα1.27	12	1
CALRR	-	TGGRALTFGKGTQINVM	Jα1.27	11	1
CALRR	-	ATGTVGKLIFGSGTRLIIQ	Jα1.18	13	1
Vα46.1					
CALSD	*cgc*	NAGFSKLVFGKGTQLLIT	Jα1.17	13	2
CALSDA	*g*	SYASKVLFGTGTLLTVRS	Jα1.4	12	1
CALSDA	-	AYSGSGWELNFGSGTQLIVQ	Jα1.14	15	1
CALTDAA	-	TGYSKLTFGKGTQLVIL	Jα1.13	13	1
Vα39					
CALRE	*a*	AYTGTVGKLIFGSGTRLIIQ	Jα1.18	14	1
CALR	-	DFVTDKLTFGIGTILLVN	Jα1.1	11	1
CAL	*ag*	SAATKLIFGEGTQLNVR	Jα1.26	10	1
Vα23					
CALSD	*ggg*	DTAGRTLTFGKGTQINVM	Jα1.24	13	1
CALRDA	*ag*	NWGTGNKIIFGGGTFLTVL	Jα2.1	15	1
Vα18.1					
CAVS	*t*	VEQSNYRLHFGIGTRLLVR	Jα1.49	12	1
Vα3.2					
CAV	*cccc*	YGGGNYKLTFGSGTKLIVT	Jα1.8	13	1
Vα32					
CATD	-	DAGAGSKLIFGTGTKLSVF	Jα1.32	12	1
Vα12					
CAASD	*tgg*	TGYGSKLTMGLGTVLFVR	Jα1.53	13	1
Vα28					
CAVS	-	DGGGNYKLTFGSGTKLIVT	Jα1.8	12	1
Vα41					
CALSD	-	AGAGLKQTFGTGTKLSVF	Jα1.32	12	1
Vα42					
CAVVA	*ag*	AAGRTLTFGKGTQINVM	Jα1.24	11	1
Vα3.4					
CAV	*taa*	SYTAGSKVIFGKGTILTVN	Jα1.19	12	1
Vα29					
CALRP	*g*	ANAGAGLKQTFGTGTKLSVF	Jα1.32	14	1
Vα17					
CAMR	*t*	SYTAGSKIIFGKGTILTVN	Jα1.19	13	1
Vα54					
CALSA	-	DVGGFKILFGSGTKLNVK	Jα1.9	12	1
Vα3.3					
CAVR	*gg*	ASYASKVLFGTGTLLTVRS	Jα1.4	11	1
Vα31					
CAVS	*ag*	DGYKYTFGSGTQLIVI	Jα1.16	10	1
Vα24					
CALSD	*g*	GAGFGKLVFGKGTQLLIT	Jα1.17	12	1
Vα3.1					
CAVS	*ccc*	NNYGKLTFGKGTILFVK	Jα1.3	11	1

RNA from sorted CD8+ cells was subjected to 5′-RACE using TCRα constant region primer; 31 TCRα clones were sequenced. Each Vα family and subfamily is boldfaced. Non-germline encoded nucleotides are shown in italic. Number of clones signifies number of sequences with identical nucleotide sequences. The CDR3 region is defined as according to the International IMGT (1).

1. Lefranc MP, et al. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27(1):55–77.

Table S3. *X. laevis* Jα gene segments

Scaffold	Position	Jα	Amino acid translation
8495	123′399	Jα1.53	ETGYGSKLTMGLGTVLFVR
8495	124′980	Jα1.52	EYGSSV*LAFGKGTTLSVT
8495	125′409	Jα1.51	TGSGXNWRIIFGRGTRLSVT
8495	126′533	Jα1.50	SFGGAEKLVFGRGTTLSIV
8495	128′025	Jα1.49	EQSNYRLHFGIGTRLLVR
8495	128′875	Jα1.48	AGSSDKLVFGSGTTLNIV
8495	129′416	Jα1.47	KYGASDKLVFGKGTILSVT
8495	131′632	Jα1.46	FTANHGMSKIVFGKGTRLNIR
8495	132′188	Jα1.45	QSTGYNKLIFGKGTILTVQ
8495	133′487	Jα1.44	NTNKYIFGAGTQLSVL
8495	135′324	Jα1.43	TGGSTKLTFGKGTKLTVL
8495	136′349	Jα1.42	NNNNVKLTFGSGTSLRVL
8495	137′382	Jα1.41	GYTKLTFGKGTKLTVL
8495	137′763	Jα1.40	GGWNKLTFGAGTTLKVQ
8495	139′026	Jα1.39	SKTLNNLIFEKGTHKTK-
8495	139′286	Jα1.38	GSGYNKLSFGEGTRLNVR
8495	139′633	Jα1.37	NTYSKLIFGQGTMLYVI
8495	142′664	Jα1.36	KGSTWGKLLFGKGTKLQVL
8495	143′064	Jα1.35	NNVGNKITFGKGTQLNVS
8495	144′036	Jα1.34	QSGGYKLTFGGGTELTVQ
8495	144′576	Jα1.33	ENNLKLVFGKGTHLIIL
8495	145′242	Jα1.32	AGAGSKLIFGTGTKLSVF
8495	145′903	Jα1.31	TGSYKLTFGSGTNLVVK
8495	146′326	Jα1.30	SSSNYGKLVFGQGTKLNVN
8495	148′420	Jα1.29	SGSGVGKLIFGTGTKLKVT
8495	149′193	Jα1.28	NAANKLMFGKGTQLNVM
8495	149′710	Jα1.27	TGGRALTFGKGTQINVM
8495	149′833	Jα1.26	SAATKLIFGEGTQLNVR
8495	151′494	Jα1.25	NAAKKLMFGKGTQLNVM
8495	151′983	Jα1.24	TAGRTLTFGKGTQINVM
8495	154′067	Jα1.23	SGGNRKLTFGTGTELSVV
8495	155′012	Jα1.22	STGGYGNMIFGQGTQLKVN
8495	155′836	Jα1.21	TTGVNKLIFGKGTNLIVK
8495	157′950	Jα1.20	YSGNNLKLIFGSGTQLIVH
8495	158′386	Jα1.19	SYTAGSKIIFGKGTILTVN
8495	159′448	Jα1.18	YTGSVGKLIFGSGTRLIIQ
8495	161′325	Jα1.17	NAGFSKLVFGKGTQLLIT
8495	161′919	Jα1.16	NDGYKYTFGSGTQLIVI
8495	162′609	Jα1.15	TNSNWKIIFGKGTKLNVY
8495	163′504	Jα1.14	YSGSGWELNFGSGTQLIVQ
8495	164′298	Jα1.13	ANTGYSKLTFGKGTQLVIL
8495	165′622	Jα1.12	YAGSSYGKLSFGSGTKLMVI
8495	166′114	Jα1.11	TGGWKYVFGSGTRLKVL
8495	166′886	Jα1.10	EFSGSKLIFGSGTELIVQ
8495	167′579	Jα1.9	DVGGFKILFGSGTKLNVK
8495	168′579	Jα1.8	YGGGNYKLTFGSGTKLIVT
8495	169′459	Jα1.7	TAAGGYRLIFGLGTKLIVT
8495	172′110	Jα1.6b	YGGSKMIFGKGTILTVT
8495	173′484	Jα1.6	NWGGAGNKLIFGTGTQLEVI
8495	174′052	Jα1.5	TYNTGKLTFGHGTKLLVK
8495	175′428	Jα1.4	SYASKVLFGTGTLLTVR
8495	178′246	Jα1.3	TGIGYGKNIFGMATKLTVK
8495	178′815	Jα1.2	NNNYGKLTFGKGTILFVK
8495	179′105	Jα1.1	NYGASNKLIFGPGTKINVK
8495	179′913	Jα1.0	FVTDKLTFGIGTILLVN
4741	76′544	Jα2.1	NWGTGNKIIFGGGTFLTVL
4741	79′648	Jα2.2	YGGGNDKLTFGSGTKLIVK
4741	81′980	Jα2.3	WKYVFGSGTKLNVL
4741	84′437	Jα2.4	NTGYAKLTFGTGTQLVVL
4741	86′016	Jα2.5	TSSNWKILFGKGTKLNIY
4741	94′192	Jα2.6	HNAAKKLMFGKGTQLNVM
4741	98′656	Jα2.7	ENNLKWVFGKGTHLTVL
4741	107′100	Jα2.8	LCATGGYNKLTFGKGTKLKVL

Table S4. TCR Vα-Jα and Vβ-Dβ-Jβ gene use by type I XNC10-T$^+$ cells

Vα	n-nucleotides	Dβ	Jα	CDR3	Amino acids	Clones
TCR Vα-Jα gene use by type I XNC10-T$^+$ cells						
Vα6						
CAASD	—		**TG**GST**KL**TFGKGTKLTVL	Jα1.43	12	35
Vβ	n-Dβ-n		Jβ	CDR3	Amino acids	Clones
TCR Vβ-Dβ-Jβ gene use by type I XNC10-T$^+$ cells						
Vβ6						
CAASE	*gggacagg*	Dβ2	DYQAYFGDGTILTVL	Jβ13	11	21
CAASD	*atcggaggacaaa*		NSGYAQTFGDGTLLTVL	Jβ10	15	1
CAASD	*ggggggcccct*		TERQYFGDGTILTVL	Jβ6	13	1
Vβ7						
CAASE	*gggacagg*	Dβ2	DYQAYFGDGTILTVL	Jβ13	11	1
CAAS	*ccaggggggt*	Dβ2	DRQYFGEGTILTVL	Jβ8	11	1
Vβ13						
CASQS	*gccagggg*	Dβ2	NTAAYFGEGTVLTVL	Jβ3	12	7
Vβ10						
CASSH	*ccgacaact*	Dβ2	SGNKVTFGEGTILHVL	Jβ2	13	1
Vβ14						
CASS	*ggcaggga*	Dβ2	GQNRLYFGDGTIVTVL	Jβ11	12	1
Vβ19						
CAGS	*tcgacagggg*	Dβ2	NAAVQYFGDGTVLTVL	Jβ5	13	1

RNA from sorted type I XNC10-T$^+$ cells was subjected to 5′-RACE using TCRα and -β constant region primers, and 35 clones from each reaction were sequenced. CDR3α residues conserved in the type II cells (Gly-110, Lys-115, and Leu-116) are bold. Representative TCRα and -β repertoires from sorted type II XNC10-T^{dim+} and CD8$^+$ cells are shown in Tables S1 and S2. CDR3 regions are determined according to the IMGT numbering for TCR V domains (1). Number of clones signifies the number of sequences with identical nucleotide sequences.

1. Lefranc MP, et al. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27(1):55–77.

Table S5. TCR Vβ-Dβ-Jβ gene use by CD8$^+$ cells

Vβ	n-Dβ-n	Dβ	Jβ	CDR3	Amino acids	Clones
Vβ1.1						
CAA	*tggaca*	Dβ2	SANTAAYFGEGTVLTVL	Jβ3	11	1
CAA	*ccccgggacaggggggg*	Dβ2	ERQYFGDGTILTVL	Jβ7	12	3
Vβ1.2						
CAASD	*ggggggagg*	Dβ1	TNTAAYFGEGTVLTVL	Jβ3	12	2
Vβ1.3						
CAA	*aaaaat*		TGQAQYFGDGTVLTVL	Jβ4	10	1
CAVSI	*tcggca*	Dβ2	NDRQYFGEGTILTVL	Jβ3	11	1
Vβ4						
CAAS	*acaggggt*	Dβ2	NDRQYFGEGTILTVL	Jβ8	11	1
Vβ5						
CAASI	*gatggggg*	Dβ1	YQAYFGDGTILTVL	Jβ13	11	1
Vβ6						
CAASD	*gggacaggggg*	Dβ2	GYAQTFGDGTLLTVL	Jβ10	13	1
CAASE	*gggacagg*	Dβ2	DYQAYFGDGTILTVL	Jβ13	11	1
Vβ7						
CAASE	*agg*		NERQYFGDGTILTVL	Jβ7	10	1
CAAS	*ccacgggacaatt*	Dβ2	NAAVQYFGDGTILTVL	Jβ5	13	1
CAA	*gggacaggggg*	Dβ2	TTERQYFGDGTILTVL	Jβ6	12	1
Vβ8						
CAVSI	*atcgacaagt*		TGQAQYFGDGTVLTVL	Jβ4	13	1
CAVSI	*tcggcag*		NDRQYFGEGTILTVL	Jβ8	11	1
Vβ9						
CAVSM	*ggactgggggagcctc*	Dβ1	PNTAAYFGEGTVLTVL	Jβ3	15	1
CAV	*gggacatct*	Dβ2	SSGNKVTFGEGTILHVL	Jβ2	12	1
Vβ13						
CASQS	*ccgggacagt*	Dβ2	TNDRQYFGEGTILTVL	Jβ8	13	1
CASQ	*gaggacagggag*	Dβ2	NDRQYFGEGTILTVL	Jβ8	12	2
CASQ	*ccccgacaggggg*	Dβ2	AYSAAYFGEGTIVSVL	Jβ12	13	1
CASQS	*tacagggg*	Dβ2	TYSAAYFGEGTIVSVL	Jβ12	13	1
CASQ	*cccgggacagggggg*	Dβ2	DERQYFGDGTILTVL	Jβ7	13	1
Vβ14						
CASSI	*agggggggagcg*	Dβ1	AYFGEGTVLTVL	Jβ3	10	1
CASS	*tatcgggggggcg*		DYQAYFGDGTILTVL	Jβ13	12	1
CVSS	*cccgac*		SDRAQYFGEGTRLLVL	Jβ1	11	1
CASSI	*ggggggggcg*		DYQAHFGDGTILTVL	Jβ13	12	1
Vβ15						
CSASL	*aaacaggcaat*	Dβ2	YSGYAQTFGDGTLLTVL	Jβ10	14	1
Vβ16						
CAAS	*ccgggacagggggg*	Dβ2	AGQTQYFGDGTVLTVL	Jβ4	13	1
CASQ	*ccgggacagt*	Dβ2	TNDRQYFGEGTILTVL	Jβ7	13	1
Vβ17						
CASST	*gggcagat*		NDRQYFGEGTILTVL	Jβ8	12	1
CASST	*gggggcccggg*		AYFGEGTVLTVL	Jβ3	10	1
Vβ19						
CAGST	*gggggagg*		AAVQYFGDGTILTVL	Jβ5	12	1
CAGST	*gacagt*	Dβ2	TNERQYFGDGTILTVL	Jβ7	12	1

RNA from sorted XNC10-T$^-$/CD8$^+$ cells was subjected to 5′-RACE using TCRβ constant region primers; 36 TCRβ clones were sequenced clones. Nongerm line-encoded nucleotides are shown. The CDR3 region is defined according to the IMGT numbering (1). The number of clones signifies the number of sequences with identical nucleotide sequences.

1. Lefranc MP, et al. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 27(1):55–77.

Table S6. TCRα repertoire in CD8$^+$ from stage 50 tadpoles

Vα family	No. of sequences	Total population (%) identified Vα segments
Vα6	171	0.8590806331
Vα22	167	0.838985181
Vα23	93	0.4672192921
Vα40	183	0.9193669931
Vα41	79	0.3968852051
Vα45	210	1.0550113041
IGH	13	0.0653102241
Vα1	10	0.0502386341
Vα2	15	0.075357951
Vα3	911	4.5767395137
Vα4	6	0.030143181
Vα5	14	0.0703340871
Vα7	59	0.2964079381
Vα9	63	0.3165033911
Vα10	900	4.5214770162
Vα11	32	0.1607636271
Vα12	1,560	7.8372268272
Vα13	1,093	5.4910826432
Vα14	70	0.3516704351
Vα15	351	1.7633760362
Vα16	261	1.3112283352
Vα17	353	1.7734237631
Vα18	2,210	11.102738013
Vα19	1	0.0050238631
Vα20	127	0.6380306461
Vα21	167	0.838985181
Vα24	6	0.030143181
Vα25	704	3.5367997991
Vα26	372	1.8688771671
Vα27	64	0.3215272541
Vα29	201	1.0097965341
Vα30	110	0.5526249691
Vα31	18	0.090429541
Vα33	110	0.5526249691
Vα34	97	0.4873147451
Vα36	147	0.7385079131
Vα37	120	0.6028636021
Vα38	65	0.3265511181
Vα39	573	2.87867371
Vα42	80	0.4019090681
Vα43	46	0.2310977141
Vα44	140	0.7033408691
Vα46	1,659	8.3345892993
Vα55	185	0.929414721
Vα56	48	0.2411454411
Vα58	346	1.7382567191
Vδ	2	0.0100477271
Vv59	95	0.4772670181
Vα60	11	0.0552624971
Vα61	236	1.1856317511
Vα62	15	0.075357951
Vα63	129	0.6480783721
Vα64	247	1.2408942481
Vα65	566	2.8435066571
Vα66	294	1.4770158251
Vα67	212	1.065059031
Vα68	441	2.2155237381
Vα69	1,010	5.0741019841
Vα70	44	0.2210499871
Vα71	36	0.1808590811
Vα72	196	0.9846772171
Vα73	79	0.3968852051

Table S6. Cont.

Vα family	No. of sequences	Total population (%) identified Vα segments
Vα74	88	0.4420999751
Vα75	60	0.3014318011
Vα76	316	1.5875408191
Vα77	27	0.135644311
Vα78	675	3.3911077621
Vα79	156	0.7837226831
Vα80	543	2.72795781
Vα81	120	0.6028636021
Vα82	14	0.0703340871
Vα83	10	0.0502386341
Other Vα		
36 different with >10 sequences/cluster	73	0.366742025
Total	19,905	100

Table S7. Primers used in the study

Primer identification	5′-3′ sequence
β2-m F	*tcaaagcttgg*agagtaacattagtcccccg*
β2-m R	*acagaattc*gaaaatgtccagtctgtactt*
XNC10α1 F	*tcaaagcttgg*gcagtcacattttccaatacagcacaacactggtttcagatc*
XNC10α3 R	*acagaattcgtgccattcgattttctgagcttcgaagatgtcg ttcaaaccacc*aaatggtacaatcaggggattctc*
CD8α F	aagccacctacgactaccaccaa†
CD8α R	ccgttcttctcagtctcaggcaca†
CD8β F	tcatcatctctttctggggc†
CD8β R	aattcagtgggtgcttcctg†
CD4 F	tccatctctgacatcccctc†
CD4 R	tcaccagacacacgtccatt†
TCRαC R2	agtgaccaggcacacgagttcagg†‡
TCRαC R1	gtcagctttgaggcgatacattg§
TCRβC R	caagtagccatggtcagtctgttcc†
Vα6-F1	ggctcttatggacagaatgttcag†
GAPDH F	acccccttcatcgacttggac†
GAPDH R	ggagccagacagtttgtagtg†
β-actin F	ggtgtcatggttggaatgg†
β-actin R	tgggttacaccatcacctgag†
TCRαC A12	ccatctcatccctgcgtgtctccgactcag*agacgtct*caagctttgcggcgatacattgaag¶
TCRαC D12	ccatctcatccctgcgtgtctccgactcag*ctgactgac*aagctttgaggcgatacattgaag¶
TCRαC D12	ccatctcatccctgcgtgtctccgactcag*gtctagtgc*aagctttgaggcgatacattgaag¶
XNC10 F	cgccatcgcattcgtctttc‖
XNC10 R	tcttcaacaccagtcttgttc‖
Vα6 F2	caacccgtcacccatctttc¶‖
Jα1.43R	gagtttggtggaacccccagt¶‖
GAPDH F	gacatcaaggccgccattaagact‖
GAPDH R	agatggaggagtgagtgtcaccat‖

*Expression of β2-m-XNC0.
†RT-PCR.
‡5′-RACE.
§Nested RT-PCR
¶Deep sequencing.
‖Quantitative PCR.