The expressed TCR β CDR3 repertoire is dominated by conserved DNA sequences
in channel catfish

R. Craig Findly^{a1}, Frank D. Niagro^a, and Harry W. Dickerson^a

Abstract

We analyzed by high-throughput sequencing T cell receptor beta CDR3 repertoires expressed by $\alpha\beta$ T cells in outbred channel catfish before and after an immunizing infection with the parasitic protozoan *Ichthyophthirius multifiliis*. We compared CDR3 repertoires in caudal fin before infection and at three weeks after infection, and in skin, PBL, spleen and head kidney at seven and twenty-one weeks after infection. Public clonotypes with the same CDR3 amino acid sequence were expressed by $\alpha\beta$ T cells that underwent clonal expansion following development of immunity. These clonally expanded $\alpha\beta$ T cells were primarily located in spleen and skin, a site of infection. Although multiple DNA sequences were expected to code for each public clonotype, each public clonotype was predominately coded by an identical CDR3 DNA sequence in combination with the same J gene in all fish. The processes underlying this shared use of CDR3 DNA sequences are not clear.

Key words: T cell receptor, high-throughput DNA sequencing, CDR3 repertoire, teleost

1. Introduction

The T cell receptor (TCR) is a polymorphic, heterodimeric protein formed by pairing of α and β chains to form the $\alpha\beta$ TCR, or γ and δ chains to form the $\gamma\delta$ TCR. The third complementarity-determining region (CDR3) of the $\alpha\beta$ TCR binds peptides presented by major histocompatibility complex (MHC) class I and II molecules with peptide specificity of the $\alpha\beta$ TCR determined by the sequence and size of the CDR3. The sequence and size of the CDR3 varies for both the α and β chains, and these variations enable TCRs expressed by different $\alpha\beta$ T cells to bind peptides with diverse sequences. It is this diversity in the repertoire of CDR3 sequences expressed by the population of circulating $\alpha\beta$ T cells in an individual that potentially permits one, or more, of the $\alpha\beta$ T cells in the population to recognize and bind any of the possible peptide sequences generated from the array of diverse and unpredictable antigens an individual might encounter (Davis and Bjorkman, 1988; Woodsworth et al., 2013). The CDR3 ranges in size from 7 to 27 amino acids with the potential amino acid sequence diversity of the human TCR β CDR3 chain estimated at $\sim 5 \times 10^{11}$ (Freeman et al., 2009; Robins et al., 2010). The actual expressed diversity of the human TCR β CDR3 repertoire has been estimated by high-throughput sequencing of cDNAs synthesized from RNA isolated from PBL of healthy donors, or from genomic DNA isolated from similar populations of $\alpha\beta$ T cells. This revealed that only a small fraction of the potential repertoire is expressed in $\alpha\beta$ T cells isolated from PBL with approximately 1 to 3×10^6 unique TCR β CDR3 amino acid sequences expressed at any one time, distributed between naïve and memory $\alpha\beta$ T cells (Freeman et al., 2009; Robins et al., 2009; Robins et al., 2010; Wang et al., 2010; Warren et al., 2011). The expressed TCR β CDR3 repertoire is not static, as shown by sequencing of the TCR β CDR3 from PBL isolated from the same individual at different times (Warren et al., 2011). Vaccination also induces changes in the CDR3 repertoire (DeWitt et al. 2016; Covacu et al., 2016).

The extent to which the expressed TCR β CDR3 repertoire changes in an individual after an immunizing infection with a natural pathogen, how the repertoire varies among different tissues of an individual, and how it differs among individuals responding to the same infection, has only begun to be characterized. To investigate the dynamics of TCR β CDR3 repertoires following immunization by a natural infection we used an established model in which channel catfish (*Ictalurus punctatus*) are infected with the ciliated, parasitic protozoan *Ichthyophthirius multifiliis* (Dickerson and Clark, 1998; Dickerson and Findly, 2014). The basic mechanisms

governing hematopoiesis and generation of hypervariable sequence diversity in CDR3 domains of TCRs and immunoglobulins have been evolutionarily conserved from teleosts to mammals, and channel catfish are widely used for studies on the evolution of T and B cell responses (Bengtén et al., 2006; Castro et al., 2011; Lieschke and Trede, 2009; Trede et al., 2004; Weinstein et al., 2009). The ciliate *I. multifiliis* (Ich) is an obligate, single-cell, extracellular parasite that infects virtually all species of freshwater fish. Severe infections result in death, but sub-lethal infections elicit an antibody-mediated immune response that provides immunity (Dickerson and Clark, 1998; Dickerson and Findly, 2014). The trophont, or feeding stage, infects the gill epithelium and the epidermal layer of skin and fin. Both $\alpha\beta$ T cells and B cells are found at the site of infection in skin of channel catfish (Findly et al., 2013; Zhao et al., 2008). As the fins of fish regenerate when damaged, T cell populations in fins, and by extension skin, can be repeatedly sampled from the same fish by taking biopsies of fin (Torini and Poss, 2014). We took biopsies from caudal fin before infection and at three weeks after an immunizing infection with Ich, and collected PBL and tissues samples of skin, spleen and head kidney at seven and twenty-one weeks after infection from outbred channel catfish. We used high-throughput sequencing of cDNAs synthesized from RNA isolated from these samples to determine the expressed diversity of TCR β CDR3 repertoires before infection and during the course of infection and induced immune response.

A partial sequence of the channel catfish TCR β locus defined the genomic sequences for a single D β gene, 29 J β genes and two C β genes, *TCRC β 1* and *TCRC β 2*. *TCRC β 1* is predominantly expressed. Twelve V β genes, distributed in five families, have been identified by cDNA sequencing (Wilson et al., 1998; Zhou et al., 2003). In channel catfish the genomic organization of the genes coding for the α and β chains of the TCR heterodimer has not been established (Liu et al., 2016). In puffer fish, *Tetraodon*, Atlantic salmon (*Salmo salar*), and zebrafish (*Danio rerio*) the organization of the genes coding for the TCR heterodimer resembles that in mammals. The genes are arranged in a translocon organization with large families of V and J genes with heterogeneous sequences, but only one or a few D and C genes (Fischer et al., 2002; Yazawa, et al., 2008; Meeker et al., 2010). In mammals, generation of hypervariable sequence diversity in the $\alpha\beta$ TCR CDR3 occurs at two levels. The TCR α gene is assembled by somatic recombination of V α and J α genes, and the TCR β gene by somatic recombination of V β , D β and J β genes to generate a contiguous V α J α , or V β D β J β , sequence located 5' to the C β ,

or constant region genes. TCR α and β transcripts contain rearranged VJ, or VDJ, gene sequences and a C gene sequence. The diversity contained in the heterogeneous sequences of the multiple V and J genes contributes to the first level of TCR CDR3 sequence diversity. In fish, as in mammals, this genetic diversity is augmented by *rag* dependent random additions and deletions of nucleotide at the junctions of the recombined genes. Deletions of coded sequences and non-templated nucleotide additions are found at the V α -J α and V β -D β and D β -J β junctions (Covacu et al., 2016; Davis and Bjorkman, 1988; Lieschke and Trede, 2009; Woodsworth et al., 2013). This non-templated diversity in CDR3 DNA sequences accounts for the variation in the number and sequence of the amino acids comprising the CDR3 loops of the TCR α and TCR β peptide chains. An $\alpha\beta$ T cell clonotype is defined by its unique CDR3 sequence.

We sequenced cDNAs coding for the TCR β CDR3 of V β 2 transcripts, as $\alpha\beta$ T cells expressing V β 2 underwent clonal expansion after an immunizing infection with Ich. Analysis of V β 2 TCR β CDR3 repertoires expressed by channel catfish revealed public V β 2 clonotypes, expressed by all four fish, which increased in abundance after infection, and accounted for up to 8 % of copies of DNA sequences in a sample. Unexpectedly, we found in each fish that a public clonotype was predominately coded by a single DNA sequence in combination with the same J gene. More strikingly, we found that the identical DNA sequence and J gene was used by all four fish to code for that same clonotype. These dominant DNA sequences were present as only a few copies, or were absent, in fin samples taken before infection and increased in abundance after infection, suggesting that they were expressed by clonally expanded T cell populations responding to Ich antigens. In addition, private clonotypes were also predominately coded for by a single dominant DNA sequence. Due to codon degeneracy, multiple, alternative DNA sequences coding for each of these clonotypes were also present, but these contributed only a few percent of the total copies of sequences coding for a clonotype. It is not clear why $\alpha\beta$ T cells expressing these dominant DNA sequences were favored for clonal expansion over those expressing alternative DNA sequences coding for the same clonotype.

2. Materials and Methods

2.1 Channel catfish and *I. multifiliis* infection and immunization.

Outbred channel catfish (*Ictalurus punctatus*) fingerlings, approximately 4 months of age, were obtained from a State of Georgia hatchery and raised as previously described (Findly et. al.,

2013; Zhao et al., 2008). They had no known exposure to Ich. *I. multifiliis* isolate G14 (serotype D) was passaged on channel catfish and theronts collected using standard protocols (Findly et. al., 2013; Wang, X. et al. 2002; Zhao et al., 2008). Eight fish (20 to 30 g) were tagged and divided into two groups. Biopsies of caudal fin were taken from the four fish in the first group (K353, K358, K362, and K368). All eight fish were then immunized at the same time by direct exposure to a sub-lethal dose of 6×10^3 theronts per fish. Biopsies of caudal fin were again taken from all four fish in the first group at day 22 after infection. Two fish from the first group were euthanized at day 51 (K362 and K368) and day 149 (K353 and K358) after infection, blood was collected from the caudal vein and the fish were dissected to remove head kidney, spleen and skin from the lateral body wall. The four fish in the second group were only used to verify that infection elicited a protective immune response. These fish were challenged at day 49 after infection with a lethal dose of 1.2×10^4 theronts per fish and their immunity to re-infection monitored by visual inspection for an additional 30 days. Naïve controls from the same population were also exposed to the same dose of theronts to insure that this dose was lethal. Infection status was monitored by visual inspection (Findly et. al., 2013; Wang, X. et al. 2002). All animal experiments were approved by the University of Georgia IACUC.

2.2 Tissue Samples and RNA isolation. Fish were anesthetized by immersion in 0.02% Tricaine (MS-222), 0.02% sodium bicarbonate before biopsies of caudal fin were taken. Fish were euthanized in 0.04% Tricaine, 0.04% sodium bicarbonate prior to collection of blood from the caudal vein and dissection to remove head kidney, spleen and skin. Blood samples were diluted 6 fold in RPMI to prevent clotting and RBC and PBL collected by centrifugation at 500 X G for 5 minutes at RT. Tissue samples and the cell pellet from blood samples were homogenized in Trizol according to the manufactures directions and total cell RNA isolated.

2.3 PCR amplification and Illumina Sequencing. The 24 RNA samples were amplified in separate RT-PCR and PCR reactions. RT-PCR was carried out in 25 µl reactions using 1 µg total cell RNA, 0.2 µM Vβ2 forward primer, 0.2 µM TCRCβ1 reverse primer and 0.5 µl Superscript III RT/Taq mix (Invitrogen). Reactions were incubated for 30 min at 55 °C, 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 60 °C for 30 s, and 68s°C for 30 s with a final extension at 68s°C for 2 min in a T100 thermal cycler. The TCRCβ1 primer was complementary to both alleles of TCRCβ1. RT-PCR products were amplified in nested PCR reactions with sequence-specific tags on the 5' termini of the Vβ2 forward primers that designated a specific tissue and

the TCRC β 1 reverse primers that designated a specific fish (Table 1). Sequences of the 5' tags were kindly provide by Dr. T. C. Glenn (Faircloth and Glenn, 2012). PCR was carried out in a final volume of 25 μ l using 0.2 mM dNTPs, 2 mM MgSO₄, 0.2 μ M primers, 2 μ l of the RT-PCR product, and 0.5 units of Platinum Taq HiFi (Invitrogen). Reactions were incubated for 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 63 °C for 30 s, and 68 °C for 30 s with a final extension at 68 °C for 2 min. The nested PCR generated 170 to 190 bp products, which were separated on 1.5% agarose gels, and visualized by ethidium bromide staining. Similar amounts of cDNA products from each of the 24 V β 2 reactions were combined and Illumina forward and reverse sequencing adaptors added. The pooled samples were sequenced on an Illumina MiSeq using a 150 bp paired-end protocol at the University of Georgia Genomics Facility. A Φ X spike-in was added to the library corresponding to $\approx$ 30% of the reads. The unique combination of sequence-specific tags permitted assignment of amplified cDNAs to their original sample after pooling and sequencing.

2.4 CDR3 Sequence Analysis. Illumina forward and reverse reads were merged using FLASH (Magoc and Salzberg, 2011). Assembly used a minimum overlap of 100 bp, a maximum overlap of 150 bp, a maximum mismatch density in the overlapping region of 0.25, and a minimum quality score of Q30. Sequences were required to have complete 5' and 3' tags in order to assign each sequence to the original fish and tissue. Each assembled sequence was required to match the cDNA sequence of the V β 2 gene (Warren et al., 2011). J gene assignments were performed using a string searching algorithm. Nucleotide sequences were translated using EMBOSS Transeq and those sequences with a complete ORF selected (Rice et al., 2000). The CDR3 sequence was defined as the region between the second conserved cysteine of the V β 2 gene and the first conserved phenylalanine in the TCR β J gene (Yousfi et al., 2004). Sample identification for each translated sequence was performed using a string searching algorithm that required exact matches to the sequence-specific tags that defined the tissue and fish from which the sequence originated.

3. Results

Eight outbred channel catfish exposed to a sub-lethal dose of Ich theronts were infected on skin and fins 8 days after exposure, but cleared the infection by day 22. Four of these fish were challenged with a lethal dose of theronts at day 49 after infection and did not become re-

infected, demonstrating that the infection had elicited protective immunity. Biopsies of caudal fin were taken from the other four fish before infection and at day 22 after infection. PBL and tissue from skin, head kidney and spleen were collected from two of these fish at day 51 (K362 and K368) and from the remaining two fish at day 149 (K353 and K358) after infection. We undertook spectratype analysis to determine the size distributions of CDR3 sequences for multiple TCR β V β genes expressed in these tissues. This demonstrated that infection resulted in a shift from a polyclonal to oligoclonal pattern for $\alpha\beta$ T cells expressing V β 2, and to a lesser extent V β 5, which indicated that $\alpha\beta$ T cells expressing V β 2 underwent clonal expansion after infection (unpub).

We sequenced cDNAs coding for the CDR3 of V β 2 transcripts in order to obtain the greatest depth of coverage for clonotypes expressed by $\alpha\beta$ T cells responding to immunization. Primers for V β 2 and TCRC β 1 were used to amplify 170 to 190 bp cDNA products, which were sequenced using an Illumina MiSeq 150 bp paired-end protocol that provided double-stranded coverage of the CDR3 and most of the V β 2 and J gene sequence. V β 2 and *TCRC β 1* primers contained 5' tags that together uniquely designated each tissue and fish, enabling CDR3 sequences to be assigned to a specific sample from each fish. After filtering the raw sequence data we obtained 1.73×10^6 copies of CDR3 DNA sequences yielding 1.02×10^5 unique DNA sequences and 7.4×10^4 amino acid clonotypes (Table S1). In channel catfish, the CDR3 spanned 27 to 60 base pairs, or 9 to 20 amino acids, which was similar in size to human CDR3 (Freeman et al., 2009). Many CDR3 DNA sequences and clonotypes were common to two or more fish. We use the term clonotype to designate a unique CDR3 amino acid sequence and define public clonotypes as those shared among all four fish.

3.1 Public clonotypes responding to Ich infection. We measured the relative abundance of clonotypes to identify those responding to immunization. V β 2 clonotypes that were present in pre-immunization samples in low relative abundance, or which were not present, but which increased in relative abundance after immunization, we inferred corresponded to clonotypes expressed by clonally expanded $\alpha\beta$ T cells responding to Ich infection and immunization. Those clonotypes coded by only a single copy, or a few copies, of DNA sequences represented $\alpha\beta$ T cells present in low numbers. To compare the relative abundance of clonotypes in each sample we calculated for each clonotype (starting with those coded by a single copy of a DNA sequence and proceeding to those coded by multiple copies of a DNA sequence) its cumulative percentage

among all clonotypes in the sample. The cumulative percentage a clonotype represented was plotted against the cumulative percentage of copies of DNA sequences coding for that clonotype in the sample. An uneven or skewed distribution is generated when a small number of clonotypes are expressed by clonally expanded T cells, and consequently are coded by a large percentage of the total copies of DNA sequences in the sample (Venturi et al., 2011).

The number of different clonotypes was similar in all tissues (Table S2). We plotted the cumulative percentages of clonotypes for all 24 samples, which showed a skewed distribution for all samples. The most skewed distributions were seen in skin and spleen samples, suggesting that immunization resulted in clonal expansion of subsets of $\alpha\beta$ T cells that localized to these tissues. We identified 12 public clonotypes that were among the 25 most abundant clonotypes in samples from skin or spleen, but which were present in low relative abundance, or were absent, in pre-infection fin samples. Among these 12 public clonotypes were three clonotypes (CAAEMGGYGNNPAYF, CAAIMGGTQPAYF, and CAARTTGSNPAYF) that were present in low relative abundance in pre-infection fin samples and increased in relative abundance to become the most abundant clonotypes in spleen of fish K362 at day 51 and skin of fish K353 at day 149 after immunization. Clonotype CAAIMGGTQPAYF accounted for 7.9 % of all copies of DNA sequences in the spleen of fish K362, and 5.8 % in skin of fish K353 (Fig. 1). At day 22 after infection the relative abundance of these three clonotypes increased slightly in most fish, although CAAEMGGYGNNPAYF and CAARTTGSNPAYF decreased in fish K362. However, at day 51 and day 149 the relative abundance of the three clonotypes increased in all tissues of the four fish, indicating that they were expressed by clonally expanded $\alpha\beta$ T cell populations (Table S3). Their relative abundance increased after immunization by 9 to 124 fold in spleen and up to 228 fold in skin (Fig 2). The extent of clonal expansion of the 12 public clonotypes varied among individual fish, tissues and time after infection. The clonotypes could be sorted into four basic groups. In group I, clonal expansion was detected at day 22 in fin, and in skin and spleen. In group II, clonal expansion was detected at day 22 in fin and in skin, but not spleen. In group III, there was little difference between pre-infection and day 22 samples, but clonally expanded populations of these clonotypes were found in skin and spleen. In group IV, there was little difference between pre-infection and day 22 samples with expanded populations subsequently found in spleen, but not skin. This pattern differed for some clonotypes in fish K353 (Table S4).

The sites at which clonal expansion of T cells occurs in fish and in which clonally expanded populations of T cells reside are not well defined (Castro et al., 2011). Only small increases in relative abundance of these 12 public clonotypes were seen in head kidney, and appeared associated with increases in skin or spleen, suggesting that head kidney was not a site of clonal expansion or residence of clonally expanded T cells after Ich infection. Increased abundance of clonotypes in PBL mirrored increases in spleen, suggesting migration of $\alpha\beta$ T cell populations expressing these clonotypes from spleen. These results suggest that $\alpha\beta$ T cell populations underwent clonal expansion in spleen, skin, or both, although it is possible these T cell populations initiated clonal expansion at other sites and migrated to spleen and skin.

3.2 DNA sequences coding for the 12 clonally expanded public clonotypes. We determined the repertoire of CDR3 DNA sequences coding for each of these 12 clonally expanded clonotypes in all six tissue samples from the four fish. As multiple DNA sequences can potentially code for each clonotype because of codon degeneracy, we expected to find an array of different DNA sequences coding for each clonotype and that these arrays would differ among the fish. Moreover, it was possible that a clonally expanded clonotype would be coded by different sets of DNA sequences in spleen and skin, suggesting that the clonally expanded $\alpha\beta$ T cell populations expressing these clonotypes arose independently in the two tissues.

In pre-infection samples only low numbers of copies of DNA sequences coding for each of the 12 clonotypes were found. In contrast to our predictions, comparison of the CDR3 DNA sequences in the pre-infection samples from each fish showed that only a single DNA sequence was predominately used to code for each clonotype. Most unexpectedly, we found that each clonotype was predominately coded by an identical DNA sequence in all four fish. In addition, the same J gene was found in combination with the dominant CDR3 DNA sequence in all four fish. In 32 of the 48 samples only the dominant DNA sequence was used to code for a clonotype. In 9 samples in addition to the dominant DNA sequence there were 1 to 4 different alternative sequences, but these were always present in fewer copies and represented only a small fraction of the possible DNA sequences that could code for each clonotype. In only one sample was the dominant DNA sequence, which was used by the other three fish, absent and only an alternative sequence present. In the remaining six samples no DNA sequences coding for a specific clonotype were present (Table S5).

The amino acid sequence “CysAlaAlaArg” corresponds to the amino terminus of 72 % of all clonotypes, and the sequence “CysAlaAla” corresponds to the amino terminus of 95 % of clonotypes. The V β 2 cDNA sequence terminates in “5' TGTGCAGCC 3' ”, coding for the amino acid sequence “CysAlaAla”. The genomic organization of the genes coding for the α and β chains of the TCR heterodimer for *I. punctatus* has not been annotated (Liu et al., 2016). However, the genome includes a V β 2 gene sequence that codes for the amino acid sequence “CysAlaAlaArg” at the predicted 3' terminus of the V β 2 gene. This suggests that “CysAlaAlaArg” is the sequence of the amino terminus for V β 2 clonotypes. These 12 public clonotypes all initiated with the sequence “CysAlaAla” and six with the sequence “CysAlaAlaArg”. The dominant DNA sequences coding for the CDR3 domains of these 12 public clonotypes all initiated with the sequence “TGTGCAGCC” and 6 with the sequence “TGTGCAGCCAGA” (Tables 2 and S6). This suggests that in 6 of the clonotypes there were deletions of two to three nucleotides from the 3' end of V β 2 gene. A complete D gene sequence was not identified in any of the 12 clonotypes. At the junction of the D and J genes there were deletions of both D and J sequences with up to 9 nucleotides deleted from the 5' end of a J gene. The distribution of non-templated nucleotides added at the V-D and D-J junctions showed that adenosine was preferentially inserted (Table 3), although G or C nucleotides are preferentially added by TdT (Motea and Berdis, 2010). In all four fish each dominant CDR3 DNA sequence was paired almost exclusively with the same J gene for each of the 12 clonotypes. This indicates that the selection of J genes was also not random, although there was no apparent preference for specific J genes, and 9 different J genes were utilized among these 12 clonotypes.

Following immunization the same DNA sequences expressed in pre-infection samples were again used almost exclusively to code for these 12 public clonotypes. These dominant DNA sequences contributed on average 97% of copies of DNA sequences coding for the 12 clonotypes. However, at day 22 following infection the number of alternative DNA sequences coding for each of these clonotypes increased from the 1 to 5 different DNA sequences in the pre-infection samples to 11 to 40 sequences, but these alternate sequences were present in only a few copies and comprised only 3% of total copies and decreased in abundance at later time points (Table S7). At day 22 alternative sequences coding for clonotypes CAARTGNSEAYF and CAARTTGSNPAY were more abundant than the dominant DNA sequences in fish K353, but subsequently decreased in abundance. After immunization 31 of the 32 possible

combinations of DNA sequences that can potentially code for clonotype CAAEMGGYGNNPAYF were identified. For clonotype CAAIMGGTQPAYF 48 potential DNA combinations are possible, and 32 were found. A potential 384 DNA combinations could code for clonotype CAARTTGSNPAYF of which 40 were present. The $\alpha\beta$ T cells expressing these alternative DNA sequences were found in the same tissues as those expressing the dominant clonotypes, but only a few copies of each alternative DNA sequence were present with the exception of fish K353 at day 22. Although $\alpha\beta$ T cells expressing the same clonotype coded by alternative DNA sequences were present in these tissues, only those $\alpha\beta$ T cells expressing the dominant DNA sequence underwent clonal expansion. This suggests that their selection was not random.

3.3 Public clonotypes – single CDR3 DNA sequences predominate.

Public clonotypes present in all four fish totaled 5,533, or 7.5% of the 73,857 V β 2 clonotypes. Most public clonotypes were restricted to a single tissue, with almost half in PBL and decreasing numbers in fin and skin, spleen, and head kidney. The percentage of public clonotypes was consistent with previous reports for human TCR β in which the overlap of CDR3 clonotypes among individuals was higher than predicted (Covacu et al., 2016; Robins et al., 2010; Warren et al., 2011). A total of 107 public clonotypes overlapped between skin-fin and spleen samples, 81 among skin-fin, spleen and PBL samples, and just two among skin-fin, spleen and head kidney samples. Fifty one public clonotypes were present in all tissue samples of the four fish (Fig. 3).

We determined the composition of the CDR3 DNA repertoires coding for all 51 public V β 2 clonotypes. Six were included in the set of 12 clonally expanded public clonotypes. For the other 45 clonotypes, 36 were predominantly coded by a single DNA sequence with an average of 17 additional different DNA sequences coding for each clonotype. These single, dominant DNA sequences comprised on average 98% of copies of the sequences coding for a clonotype. The remaining nine clonotypes were coded by two to five dominant DNA sequences.

3.4 Private clonotypes – single CDR3 DNA sequences predominate. The majority of clonotypes were private. We compared the repertoire of DNA sequences coding for 9 private clonotypes that were among the 25 most abundant clonotypes in pre-infection and day 22 fin samples of a single fish. An average of 18 DNA sequences coded for each private clonotype, but the distribution was dominated by single DNA sequences, which accounted for an average of 97% of the copies of DNA sequences, similar to that seen for public clonotypes. As these

clonotypes were abundant in pre-infection samples, they represented clonotypes expressed by $\alpha\beta$ T cells that had responded to stimuli other than Ich infection. An additional 9 clonotypes were selected from among the 25 most abundant clonotypes in spleen or head kidney of single fish. As these clonotypes increased in abundance following infection, this suggested that they represented private clonotypes expressed by $\alpha\beta$ T cells that underwent clonal expansion in response to Ich infection. Between 4 and 14 DNA sequences coded for each clonotype, a reduction in repertoire diversity from that seen in pre-infection and day 22 fin samples. Nevertheless, each clonotype was predominately coded by a single DNA sequence, which comprised on average 98% of copies of the sequences coding for a clonotype.

4. Discussion

Our results provide a detailed characterization of the dynamics of TCR β CDR3 repertoires for V β 2 clonotypes in four outbred fish from before immunization by natural infection until twenty one weeks after immunization. V β 2 repertoires were compared in tissue samples taken from sites of infection in fin and skin, as well as from spleen, head kidney and PBL. Approximately 1×10^5 unique DNA sequences were detected from a total of 1.7×10^6 copies of DNA sequences. We identified $\alpha\beta$ T cells expressing public clonotypes, present in low abundance before infection in all four fish which, following infection and development of immunity, underwent clonal expansion with the number of copies of DNA sequences coding for some clonotypes rising to 6% to 8% of the total number of copies.

High-throughput sequencing of TCR β CDR3 repertoires in PBL from human donors revealed that the overlap of TCR β CDR3 amino acid sequences between individuals was larger than predicted (Robins et al., 2010; Warren et al., 2011). Sequencing of peripheral blood mononuclear cells demonstrated that two donors shared 14.2% and 11.7% of sampled sequences with a third donor at the amino acid level. However, at the nucleotide level there were few shared sequences, and the two donors shared only 1.1% and 0.7% with the third donor. This difference was attributed to antigen driven responses, which resulted in selection of preferred amino acid sequences in the TCR β CDR3 repertoire of $\alpha\beta$ T cells. However, the population of $\alpha\beta$ T cells expressing these preferred sequences used multiple, different nucleotide sequences to code for them, due to the degeneracy of the genetic code (Warren et al., 2011). As the number of possible TCR β CDR3 sequences is theoretically vast, even this level of sharing was unexpected

((Davis and Bjorkman, 1988; Robins et al., 2010; Warren et al., 2011). Because of the degeneracy of the genetic code and the assumption that addition of nucleotides at V-D and D-J junctions is essentially random, we also expected that multiple DNA sequences would code for each of the clonally expanded, public clonotypes in a single fish. In addition, we speculated that a clonotype might be coded by different sets of DNA sequences in skin and spleen, suggesting that different populations of $\alpha\beta$ T cells coding for the clonotype had been clonally expanded in these two tissues. We also expected that the sets of DNA sequences coding for each public clonotype would differ among the fish. Instead, we found that each of these 12 public clonotypes was predominately coded, respectively, by a single DNA sequence in pre-infection fin samples of a fish. Most unexpectedly, we found that all four fish used the same dominant DNA sequence in combination with the same J gene to code for each clonotype. After infection each of these clonally expanded public clonotypes was still predominately coded by the identical DNA sequence found in uninfected fish. However, following infection the number of alternative DNA sequences coding for each public clonotype also increased, although these alternative sequences represented only ~3% of total copies, distributed among 11 to 40 alternative DNA sequences. The presence of these alternative DNA sequences demonstrated that somatic recombination of V, D, and J genes accompanied by deletion and addition of nucleotides at V-D and D-J junctions in $\alpha\beta$ T cells occurred in these fish. For some clonotypes the number of alternative DNA sequences encompassed essentially all possible combinations of sequences that could code for the clonotype. However, even though the same clonotype was expressed by these $\alpha\beta$ T cells, but using alternative DNA sequences to code for it, they did not undergo clonal expansion. At later time points copies of most of these alternative DNA sequences were not present.

The $\alpha\beta$ T cells expressing the dominant DNA clonotypes identified in the pre-infection samples were preferentially expanded after immunizing infection. For these 12 public clonotypes, the number of copies of the dominant DNA sequence averaged 97% of the total copies of all sequences coding for that clonotype after infection. The distribution of these clonally expanded $\alpha\beta$ T cells varied among the clonotypes with some primarily located in skin, some in spleen and some in both tissues. The clonotypes had similar distributions in all fish. We found no preference for J genes, and nine different J genes were used in combination with the 12 clonotypes. The CDR3 DNA sequences for the 12 public clonotypes had nucleotide deletions at the 3' end of the V β 2 gene, at both the 5' and 3' ends of the D gene, and the 5' end of the J genes.

Non-templated nucleotides were added at the V-D and D-J junctions. At both junctions the predominate nucleotide added was A, although TdT has a bias towards addition of G or C nucleotides (Motea and Berdis, 2010).

We compared the repertoire of DNA sequences coding for an additional 45 public clonotypes and of these 36 were coded, respectively, by a dominant DNA sequence in all fish. Although an average of 18 different DNA sequences coded for each of these 36 clonotypes, the dominant sequence represented on average 96 % of copies. Eight of the remaining nine clonotypes were predominantly coded by two or three different DNA sequences. Only one public clonotype was coded by five DNA sequences, but these five sequences were present in all fish. Thus clonotypes that were predominantly coded, respectively, by the same dominant DNA sequence in all four fish included public clonotypes expressed by $\alpha\beta$ T cells that underwent clonal expansion after infection, public clonotypes found in all tissues in all four fish, and private clonotypes.

High-throughput sequencing of the IgM heavy chain CDR3 repertoire in zebrafish, suggested that VDJ use in the primary repertoire of zebrafish at the amino acid level was less random than expected (Jiang et al., 2011). Our results also indicate that the selection of $\alpha\beta$ T cells expressing these public clonotypes was not random, as all fish expressed the same DNA sequences. The mechanisms underlying the processes through which $\alpha\beta$ T cells in different fish preferentially use the same DNA sequence to code for a clonotype are not known. Codon usage for both the dominant and the alternative CRD3 DNA sequences was similar to the codon frequency found in the sequences for *I. punctatus* TCR V β 2, V β 4 and V β 5 cDNAs. These fish could have been previously exposed to a non-lethal infection of a different Ich serotype that was not detected, and which resulted in low level expression of the 12 public clonotypes in our pre-infection samples. However, that does not explain how these 12 public clonotypes were each preferentially coded, respectively, by the same DNA sequence in all four fish. It is possible that copies of rearranged TCR β VDJ gene sequences have been stably integrated into the *I. punctatus* germ line and are preferentially transcribed. A fused IgH VDJ gene is present in the *I. punctatus* germ line at the IgH locus, although it is not known if this IgH VDJ gene is transcribed (Bengten et al., 2006).

We are now analyzing TCR β CDR3 repertoires for V β 5 genes from channel catfish immunized by infection with Ich. In addition, we are analyzing TCR β CDR3 repertoires

following vaccination of channel catfish with purified immobilization antigens (i-antigens) isolated from Ich. I-antigens are a class of highly abundant 40-60 kDa surface membrane proteins covering the cilia and cell body of Ich. They are the primary antigen recognized by channel catfish during infection, and vaccination with them is sufficient to induce protective, serotype-specific immunity to Ich (Wang et al., 2002). Results from these studies will allow us to characterize the composition of TCR β CDR3 repertoires for other V β genes from naturally infected fish, and fish protectively vaccinated against Ich.

Acknowledgements. This work was supported by Agricultural and Food Research Initiative grants 2008-35204-04604 and 1000626 from the USDA National Institute of Food and Agriculture (R.C.F and H.W.D).

References

- Bengten E., et al. 2006. Channel catfish immunoglobulins: repertoire and expression. *Devel. Comp. Immunol.* 30, 77-92.
- Castro, R., et al. 2011. T cell diversity and TCR repertoires in teleost fish. *Fish and Shellfish Immunol.* 31, 644-654.
- Covacu, R., et al. 2016. System-wide analysis of the T-cell response. *Cell. Rep.* 14, 2733–2744.
- Davis, M.M., Bjorkman, P.J. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* 334, 395–402.
- DeWitt, W.S., et al. 2016. Dynamics of the cytotoxic T cell response to a model of acute viral infection. *J. Virol.* 89, 4517-4526.
- Dickerson, H.W., Clark, T. 1998. *Ichthyophthirius multifiliis*: a model of cutaneous infection and immunity in fishes. *Immunol. Rev.* 166, 377-384.
- Dickerson, H.W., Findly, R.C. 2014. Immunity to *Ichthyophthirius* infections in fish: A synopsis *Devel. Comp. Immunol.* 43, 290-299.
- Faircloth, B.C., Glenn, T.C. 2012. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS ONE* 7, e42543.
- Findly, R.C., et al. 2013. B cell memory following infection and challenge of channel catfish with *Ichthyophthirius multifiliis*. *Dev. Comp. Immunol.* 39, 302-311.

- Fischer, C., et al. 2002. Conservation of the T-cell receptor alpha/delta linkage in the teleost fish *Tetraodon nigroviridis*. *Genomics* 79, 241-248.
- Freeman, J.D., et al. 2009. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* 19, 1817-1824.
- Jiang, N., et al. 2011. Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proc. Natl. Acad. Sci. USA* 108, 5348-5353.
- Lieschke, G.J., Trede, N.S. 2009. Fish immunology. *Curr. Biol.* 19, R678-82
- Liu, Z. et al. 2016. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat. Commun.* 7, 11757
- Magoc, T., Salzberg, S. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957-2963.
- Meeker, N.D., et al. 2010. Characterization of the zebrafish T cell receptor beta locus. *Immunogenetics* 62, 23-29.
- Motea, E.A., Berdis, A.J. 2010. Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. *Biochimica et Biophysica Acta* 1804, 1151–1166.
- Rice, P. et al. 2000. EMBOSS: the European molecular biology open software suite. *Trends in Genetics* 16, 276—277.
- Robins, H.S. et al. 2009. Comprehensive assessment of T-cell receptor β - chain diversity in $\alpha\beta$ T cells. *Blood* 114, 4099-4107.
- Robins, H.S. et al. 2010. Overlap and effective size of the human CD8⁺ T cell receptor repertoire. *Sci. Transl. Med.* 2, 1-8.
- Trede, N.S. et al. 2004. The use of zebrafish to understand immunity. *Immunity* 20, 367-379.
- Torini, V.A., Poss, K.D. 2014. Keeping at arm's length during regeneration. *Dev. Cell* 29, 139-145.
- Venturi, V. et al. 2011. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J. Immunol.* 186, 4285-4294.
- Wang, C. et al. 2010. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc. Nat. Acad. Sci. USA.* 107, 1518-1523.
- Wang, X. et al. 2002. Immunisation of channel catfish, *Ictalurus punctatus*, with *Ichthyophthirius*

multifiliis immobilisation antigens elicit serotype-specific protection. Fish Shellfish Immunol.

13, 337-350.

Warren, R.L. et al. 2011. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least

1 million clonotypes. Genome Res. 21, 790-797.

Weinstein, J.A. et al. 2009. High-Throughput sequencing of the zebrafish antibody repertoire. Science 324, 807-810.

Wilson, M.R. et al. 1998. T-cell receptors in channel catfish: structure and expression of TCR α and β genes. Mol. Immunol. 35, 545-557.

Woodsworth, D.J. et al. 2013. Sequence analysis of T-cell repertoires in health and disease. Genome Medicine 5, 98

Yazawa, R. et al. 2008. Striking antigen diversity in the Atlantic salmon T-cell receptor alpha/delta locus. Dev. Comp. Immunol. 32, 204-212.

Yousfi, M.M. et al. 2004. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V- D-J junctions. Bioinformatics 20, i379-i385.

Zhao, X. et al. 2008. Cutaneous antibody secreting cells and B cells in a teleost fish. Devel. Comp.

Immunol 32, 500-508.

Zhou, H. et al. 2003. The T cell receptor β locus of the channel catfish, *Ictalurus punctatus*, reveals

unique features. J. Immunol. 170, 2573-2581.

Figure Legends

Fig. 1. The same three clonotypes were the most abundant in spleen of fish K362 at day 51 and skin of fish K353 at day 149 after infection. Clonotypes CAAIMGGTQPAYF, CAAEMGGYGNNPAYF, and CAARTTGSNPAYF were coded by 7.9 %, 4.6 %, and 3.3 % of all copies of DNA sequences in spleen of fish K362 and 5.8 %, 3.6 %, and 2.6 % of all copies of DNA sequences in skin of fish K353, respectively.

Fig. 2. The fold increase in relative abundance after Ich infection for four fish (K362, K368, K353, K358) of three public clonotypes (CAAEMGGYGNNPAYF, CAAIMGGTQPAYF, and CAARTTGSNPAYF) in five tissues: fin at day 22 and skin, PBL, head kidney and spleen at day 51 (K362, K368) and at day 149 (K353, K358) after infection.

Fig. 3. V β 2 public clonotypes. A total of 73,857 V β 2 clonotypes were identified of which 5,533 were public clonotypes present in all four fish. The number of public clonotypes specific to each tissue and those expressed in multiple tissues are shown. There were 51 public clonotypes expressed in all tissues of the four fish.

Table 1. Primers for CDR3 PCR amplification

RT-PCR Primers			
Vβ2 Forward		TCRCβ1 Reverse	
5 ' ACGGCTACGAGTAAGCCAGA	3 '	5 ' CAGTCTGCTGCTGATGGTGT	3 '
Nested PCR Primers			
Tissue	Forward Primers		
Fin: Pre-infection	5 ' TGCATAC	TGACCATCTCTAAACTCCTCCA	3 '
Fin: Day 22	5 ' ACCGAC	TGACCATCTCTAAACTCCTCCA	3 '
Skin	5 ' CCTATTGG	TGACCATCTCTAAACTCCTCCA	3 '
PBL	5 ' AACACCAG	TGACCATCTCTAAACTCCTCCA	3 '
Head	5 ' ACAAGGC	TGACCATCTCTAAACTCCTCCA	3 '
Kidney			
Spleen	5 ' CGGTT	TGACCATCTCTAAACTCCTCCA	3 '
Fish	Reverse Primers		
K362	5 ' ATCTGG	TTTAGCACGGTGACAGTTGG	3 '
K368	5 ' GAAGTCTA	TTTAGCACGGTGACAGTTGG	3 '
K353	5 ' ACGTC	TTTAGCACGGTGACAGTTGG	3 '
K358	5 ' CTAGG	TTTAGCACGGTGACAGTTGG	3 '

The 5' sequence-specific tags on the nested PCR primers that specified a tissue or fish are shown in bold. The RT-PCR primers correspond to Vβ2 bases 203 to 222 and to TCR1B bases 242 to 261 of AF038159 and AF410786, respectively. The nested primers correspond to Vβ2 bases 285 to 306 and to TCR1B bases 58 to 77 of AF038159 and AF410786, respectively.

Table 2. CDR3 Sequences of 3 public clonotypes responding to infection

Vβ2 5' ...TGTGCAGCCAGA 3'	D 5' GGGACAGGGGGC 3'	J Gene	Nucleotide Deletions Vβ2 Gene	Nucleotide Deletions J Gene
CAAEMGGYGNNPAYF		21	3	1
TGTGCAGCCG AAATGGGGGG TTATGGAAACAATCCAGCTTACTTT				
CAAIMGGTQPAYF		24	2	2
TGTGCAGCCATA ATGGGTGG CACTCAGCCTGCATACTTT				
CAARTTGSNPAYF		28	0	8
TGTGCAGCCAGAA CCACAGGGAG CAACCCAGCATACTTT				

The CDR3 amino acid sequence for 3 public clonotypes and the dominant DNA sequence coding for each of these clonotype are shown. The genomic DNA sequences for the 3' end of the Vβ2 gene and TCRβ D gene are shown above. The number of nucleotides deleted from the 3' end of the Vβ2 gene and the 5' end of the J genes are indicated. The D gene sequence in a CDR3 DNA sequence and N nucleotide additions at the V-D and the D-J gene junctions are shown in bold.

Table 3. Distribution of nucleotide additions at the V-D and the D-J gene junctions

Total nucleotides added	A	C	G	T
56	26 (46 %)	10 (18 %)	9 (16 %)	11 (20 %)

The total number of N nucleotides added at the V-D and the D-J gene junctions and the percentage each nucleotide represents for the 12 public clonotypes shown in Tables 2 and S6.

Figure

Figure