IMMUNOLOGY

Antigen receptor repertoires of one of the smallest known vertebrates

Orlando B. Giorgetti[1]*, Prashant Shingate[2*†], Connor P. O'Meara[1], Vydianathan Ravi[2], Nisha E. Pillai[2‡], Boon-Hui Tay[2], Aravind Prasad[2§], Norimasa Iwanami[1‖], Heok Hui Tan[3], Michael Schorpp[1], Byrappa Venkatesh[2¶], Thomas Boehm[1¶]

The rules underlying the structure of antigen receptor repertoires are not yet fully defined, despite their enormous importance for the understanding of adaptive immunity. With current technology, the large antigen receptor repertoires of mice and humans cannot be comprehensively studied. To circumvent the problems associated with incomplete sampling, we have studied the immunogenetic features of one of the smallest known vertebrates, the cyprinid fish *Paedocypris* sp. "Singkep" ("minifish"). Despite its small size, minifish has the key genetic facilities characterizing the principal vertebrate lymphocyte lineages. As described for mammals, the frequency distributions of immunoglobulin and T cell receptor clonotypes exhibit the features of fractal systems, demonstrating that self-similarity is a fundamental property of antigen receptor repertoires of vertebrates, irrespective of body size. Hence, minifish achieve immunocompetence via a few thousand lymphocytes organized in robust scale-free networks, thereby ensuring immune reactivity even when cells are lost or clone sizes fluctuate during immune responses.

INTRODUCTION

During the early stages of vertebrate evolution, the emergence of lymphocytes as a new cell type in adaptive immune systems was followed by the invention of somatic diversification of antigen receptors and their clonal expression (*1*, *2*). Somatic diversification has the potential to generate an enormous number of structurally distinct receptors from a small set of germline-encoded building blocks and is a defining and essential characteristic of vertebrate immunity (*3*). Because effective immunity depends on large and diverse repertoires of antibodies [immunoglobulin (Ig)] (*4*) and T cell receptors (TCRs) (*5*), numerous studies have examined the diversity of antigen receptor repertoires under physiological and pathological conditions. However, the rules underlying the structure of antigen receptor repertoires are not yet fully defined (*6*, *7*), despite their enormous importance for the understanding of adaptive immunity in general and the natural history of clinically relevant immune disorders in particular (*8*). Recently, the development of powerful sequencing technologies has led to renewed interest in this biological problem (*6*, *7*), although the sheer magnitude of the repertoires (*9*–*15*), and the complex anatomy pose a considerable challenge to defining their size and structure (*16*), particularly for animals with billions of lymphocytes distributed throughout the whole body.

Notwithstanding the inevitable sampling problems, studies of human, mouse, and zebrafish immune systems have revealed that despite their extraordinary diversity, the repertoires of different individuals partially overlap and that the frequency distributions of clonotypes contained in the sampled repertoires follow a power law (*9*–*15*, *17*, *18*). Moreover, these studies have uncovered intriguing aspects of immune system maturation, heritable contributions, and the effects of immune responses on sequence compositions (*9*–*15*, *17*–*20*). However, because it is unclear whether the samples subjected to analysis are representative of the total lymphocyte populations of the entire animal, a considerable degree of uncertainty remains about the generality of these properties. For instance, if the observed power-law distributions of clonotype frequency were indeed representative properties, then it would suggest that antigen receptor repertoires are organized as self-similar or fractal systems (*21*). Fractal systems exhibit similar topological patterns at increasingly small scales and thus have a series of desirable properties for immune systems, the most important of which is their robustness to changes in the frequencies or even total loss of individual components (*22*).

To circumvent the inevitable sampling problems encountered with large vertebrates, such as humans, and the enormous size of their antigen receptor repertoires (*9*–*15*, *17*–*20*), we have studied the immunogenetic features of one of the smallest known vertebrates. The cyprinid fish *Paedocypris* sp. "Singkep" ("minifish") (*22*, *23*) is known to mature at approximately 8 mm in standard length. So far, minifish were examined for adaptations of genome structure and developmental trajectories associated with miniaturization (*22*, *23*); by contrast, its immune system has not yet been studied. We reasoned that owing to its small body size and the correspondingly small number of lymphocytes, it should be possible to achieve near-complete coverage of clonotype sequences, a previously unattainable goal. Here, we show that self-similarity is a fundamental property of antigen receptor repertoires of vertebrates, irrespective

[1]Department of Developmental Immunology, Max Planck Institute of Immunobiology and Epigenetics, Stuebeweg 51, 79108 Freiburg, Germany. [2]Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore 138673, Singapore. [3]Lee Kong Chian Natural History Museum, National University of Singapore, Singapore 117377, Singapore.
*These authors contributed equally to this work.
†Present address: Genome Institute of Singapore, A*STAR, 60 Biopolis Street, Singapore 138672, Singapore.
‡Present address: 10x Genomics Pte Ltd., Singapore, Singapore.
§Present address: Population Health and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Australia.
‖Present address: Center for Bioscience Research and Education, Utsunomiya University, Utsunomiya, Japan.
¶Corresponding author. Email: boehm@ie-freiburg.mpg.de (T.B.); mcbbv@imcb.a-star.edu.sg (B.V.)

of their body size, and illustrate that scale-free networks of antigen receptor specificities allow minifish to achieve immunocompetence with a few thousand lymphocytes.

RESULTS

Genome assembly of minifish

Our initial analysis of the immunogenome of minifish focused on the structure of antigen receptor gene loci. Because studies in minifish are limited to small numbers of preserved specimens of this uncommon species, we relied on DNA and RNA sequence information. To this end, we first assembled comprehensive transcriptomes and then established high-quality genome assemblies to be able to determine the numbers, positions, and order of individual genetic of immune-related genes. We sequenced genomic DNAs extracted from two individuals, a male and a female minifish. The assemblies indicated identical overall genome sizes of 403 Mb (contig N50, 42.8 kb; scaffold N50, 7.3 Mb) and 404 Mb (contig N50, 36.3 kb; scaffold N50, 11.0 Mb), respectively, with an estimated completeness of about 95% (table S1), similar to other species of *Paedocypris* (*24*); approximately 27% of both genomes were found to contain repetitive sequences (table S2). The transcriptomes of a further pair of animals were established to support the gene annotation efforts. A total of 20,013 and 18,003 protein-coding genes were predicted in the male and female minifish genome assemblies, respectively, in line with other cyprinids (*24*).

Immune-related organs and lymphocyte numbers

The presence of immune-related organs has not yet been investigated in minifish. However, with respect to the thymus, a primary lymphoid organ, it is known from studies of other teleosts that two paralogous transcription factor genes, *foxn1* and *foxn4*, both contribute to thymopoiesis (*25*); both genes are found in the minifish (see the Supplementary Materials). We conclude that minifish have a functional thymic microenvironment that is known to be required for T cell development. With respect to secondary lymphoid tissues, we focused our attention on the spleen, since teleosts do not have lymph node structures (*26*). Studies in mammals and zebrafish have shown that the formation of the splenic primordium requires the activity of the *tlx1* transcription factor gene (*27*, *28*), which sets the stage for subsequent organ formation. We found that the minifish genome contains an intact *tlx1* gene (see the Supplementary Materials), suggesting that the spleen is formed normally in minifish. Likewise, no information is available on the number of lymphocytes in minifish. Under the assumption that the cyprinid body plan and the general structure of the hematopoietic tissues are conserved between zebrafish and minifish, we measured the number of T lymphocytes in zebrafish of about 3 weeks of age; at this time point, zebrafish are similar in size and body weight to minifish. To specifically mark T lineage cells, we constructed an *lck:GFP* reporter strain and found that the number of T cells in 3-week-old zebrafish corresponds to about 37,000 cells (see Materials and Methods). In zebrafish, the number of B cells is approximately twice that of T cells (*29*), indicating that minifish may have in the order of 75,000 B cells.

The genes encoding antigen receptors and their signaling components

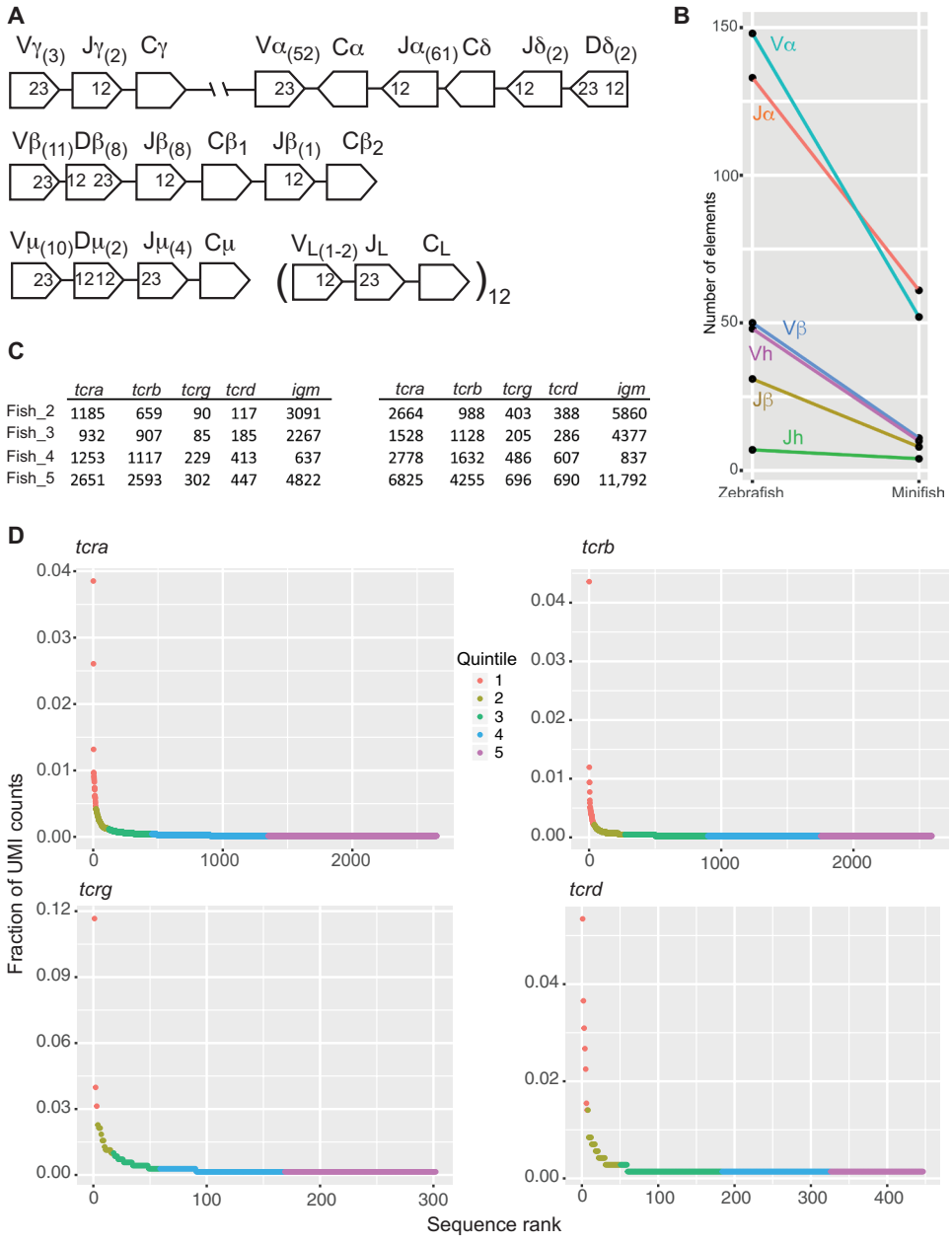
Although we had considered the possibility that minifish may not require all lymphocyte lineages that constitute the canonical adaptive immune system in larger animals, we found that minifish has the complete genetic machinery to generate antibodies and the two principal TCRs. The *igh* locus has a structure similar to that of other teleost genomes (*30*) but lacks exons encoding the constant region of *igz* (Figs. 1A and 2); six translocon elements each for two families of *igl* genes (Fig. 1A) complete the components of the canonical antibody generating system of minifish, in line with the presence of genes encoding key elements of the B cell receptor (BCR) signaling complex (*cd79a* and *cd79b*) (see the Supplementary Materials). With respect to the TCR genes, we found that the *tcra/d* locus conforms to the typical teleost structure (Figs. 1A and 2) (*31*). The same is true (*32*) for the *tcrb* locus (Figs. 1A and 2). As for the phylogenetically closely related cyprinid *Danio rerio* (zebrafish) (see www.ensembl.org/Danio_rerio), the *tcrg* locus is closely linked to the *tcra/d* locus. Collectively, our analysis suggests that all known somatically diversifying antigen receptor gene loci are present in minifish. However, in contrast to the situation of protein-coding genes (*24*), it appears that the miniaturization of body size is associated with a marked reduction of the numbers of V, D, and J elements, substantially constraining the magnitude of combinatorial diversity during somatic diversification of antigen receptors; this reduction occurs in all antigen receptor loci when compared to zebrafish (Fig. 1B). The reduction of elements is essentially random, as exemplified by the 52 variable genes in the *tcra/d* locus in comparison to their counterparts in zebrafish (fig. S1A). As expected, minifish has genes encoding key elements of the TCR signaling complex (*cd3e*, *cd3gd*, and two paralogs of *cd3z*) (see the Supplementary Materials).

The diversity of antigen receptor repertoires

The small body size of minifish offered the unprecedented opportunity to examine the diversity of expressed antigen receptor genes in much greater depth than would be possible with larger species, including zebrafish (*17*, *18*). To this end, we extracted total RNA from whole bodies of four fish and used the equivalent of $\sim^1/_3$ of total RNA each to establish an unbiased representation of *igm* and *tcr* clonotypes after complementary DNA (cDNA) synthesis and multiplex amplification; the read statistics are presented in table S3. Our sequencing strategy not only minimizes the sampling problem but also includes the repertoires expressed by all lymphocytes, irrespective of whether they are situated in primary lymphopoietic organs or peripheral tissue sites, hence comprising receptors before and after selection. In this work, clonotypes are primarily defined as unique nucleotide sequences across the entire V, D (if appropriate), and J segments, rather than just CDR3 sequences. However, in the subsequent network analysis, which is carried out using conceptually translated protein sequences, a clonotype may be derived from one to many nucleotide sequences that all have the same CDR3 protein sequence irrespective of variations in V and J segments; to distinguish them from the primary clonotypes, we refer to them as CDR3 clonotypes. We have chosen to use Shannon's entropy theorem to examine diversity of both nucleotide and protein sequences; moreover, it can be used to estimate a minimum number of different sequences that a system of entropy H can generate (see Materials and Methods).

Characteristics of *igm* gene assemblies

As shown in Fig. 1C, we detected up to about 5000 different *igm* sequences in minifish individuals. Considering the fraction of RNA sequenced in this experiment, *igh* clonotypes may reach a total of



	tcra	tcrb	tcrg	tcrd	igm
Fish_2	1185	659	90	117	3091
Fish_3	932	907	85	185	2267
Fish_4	1253	1117	229	413	637
Fish_5	2651	2593	302	447	4822

	tcra	tcrb	tcrg	tcrd	igm
	2664	988	403	388	5860
	1528	1128	205	286	4377
	2778	1632	486	607	837
	6825	4255	696	690	11,792

Fig. 1. Characteristics of germline and rearranged antigen receptors of *Paedocypris* sp. Singkep. (**A**) Germline structure of immune antigen receptor genes. The numbers of elements are indicated in parentheses; the spacer lengths of recombination signal sequences are indicated by numbers inside the cartoons. (**B**) General reduction of genetic elements in minifish compared to zebrafish. (**C**) Numbers of antigen receptor clonotypes (left table) and corresponding complementary DNA (cDNA) molecules (right table) in four unrelated individuals; these numbers were determined by subjecting one-third of total RNA to sequencing (cf., Materials and Methods). (**D**) Clonal distributions of TCR chains from a single individual (fish no. 5) represented in quintiles; these distributions follow a power-law indicative of the fractal nature of the repertoires.

about 15,000 per fish. Under the assumption that minifish harbor about 75,000 B cells, this would correspond to an average clone size of approximately 5 cells per clonotype. In addition to contributions by palindromic (P) nucleotides, the nucleotide sequences of CDR3 regions provided clear evidence of nontemplated (N) nucleotide additions at the junctions (fig. S2A), in line with the presence of a functional terminal deoxynucleotidyl transferase ortholog (see the Supplementary Materials); the length distribution of CDR3 sequences assumed a Gaussian shape with a mean value of 12.4 ± 1.5 (means ± SD) amino acid residues (fig. S2B). Entropy analysis based on amino acid sequences indicated that the contributions of the V and J regions amount to 3.03 and 1.78 bits, respectively, with the internal segment of the CDR3 regions (comprising of P, N nucleotides, and Dh element sequences) additionally contributing 14.92 bits (~76% of the total). These results indicate that the *igm* locus can generate a minimum of ($2^{19.73}$~) 860,000 different igm heavy chains.

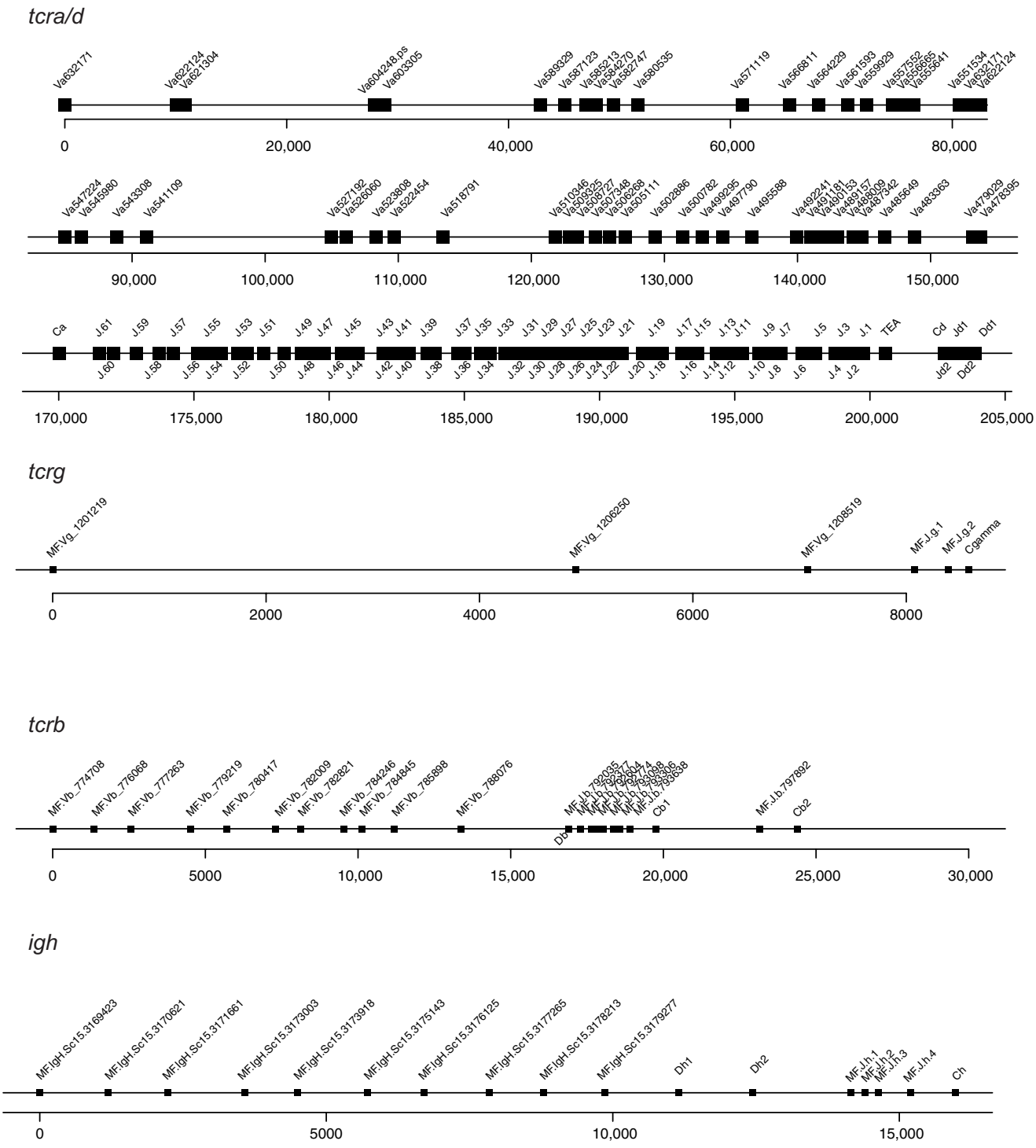


Fig. 2. Structure of antigen receptor loci in minifish. In the *tcra/d* locus, one Cδ region, two Jδ, and two Jδ elements are present and are arranged in tandem to 61 Jα elements and the constant region of the *tcra* gene; the V region cluster comprises 52 Vα/δ elements and is situated in opposite orientation downstream of *tcra* constant region gene. As observed for other teleosts, this configuration necessitates inversional rearrangements to generate functional variable regions (VαJα for *tcra* and VδDδJδ for *tcrd* genes) but allows for the possibility of lineage-modifying secondary rearrangements (*31*). The *tcrg* locus is closely linked to the *tcra/d* locus on the same scaffold, although it consists of a mere three Vγ, two Jγ, and one constant region. In the *tcrb* locus, 11 Vβ elements are associated with two constant regions; however, only 1 of the Cβ genes is preceded by a D element (and 8 J elements), whereas the second constant region is preceded by 1 J element only. The *igh* locus has a structure typical of teleost genomes comprising 10 tandemly arranged variable (V_H), 2 diversity (D_H), 4 joining (J_H), and 1 constant region (Cμ) elements. Exons encoding the constant region of *igz* were not detected. Physical maps of the indicated loci were derived from the following scaffolds (sc): *tcrg* and part of *tcra/d* on female sc0015; remainder of *tcra/d* on female sc0030; *tcrg* on male sc0017 and *tcra/d* on male sc0032; *tcrb* on female sc0010 and male sc0072; and *igh* on female sc0014 and male sc0015. The genes encoding Ig light chains are not shown.

The repertoire of *igm* clonotypes is characterized by a small fraction of prevalent clones, whereas most of clonotypes are of low frequency (fig. S2C). Although we have detected an intact *aicda* gene in the genome and transcriptome sequences (see the Supplementary Materials), the comparatively small number of sequences available for analysis precluded a definitive conclusion about the presence of substoichiometric (that is somatically mutated) variants of germline V sequences in the transcriptome. Although the *igl* light chain gene repertoires were not studied here, it is possible that the antibody specificities may go beyond the 15,000 clonotypes estimated from the analysis of the heavy chain assemblies at the *igm* locus, which would further reduce the average clone size.

The two T cell lineages exhibit diverse antigen receptor repertoires

We estimated the relative proportion of the two principal T cell lineages based on the number of clonotypes. We found that the numbers of different clonotypes of *tcrg* and *tcrd* are much smaller than those of *tcra* and *tcrb* (Fig. 1C), suggesting that only about 13.6 ± 6.7% of T cells belong to the γδ T cell lineage; this finding is in line with recent work using tcrg- and tcrd-specific antisera in adult zebrafish (*33*), further emphasizing the similar immune system structures of zebrafish and minifish. On the basis of the clonotype numbers of *tcrb* and *tcrd* assemblies in fish no. 5 (Fig. 1C), a minifish individual may have at least about 8000 αβ T cells and 1100 γδ T cells; given that *trca* and *tcrg* assemblies also contribute to diversification of antigen specificities in the TCR heterodimers, these numbers must be considered a lower bound. On the basis of *tcrb* and *tcrd* clonotype numbers alone, the average clone size is in the order of ~4, a number consistent with the estimated number of T cells in zebrafish of the same body size. Despite the small overall number of cells in the T cell compartment, we found that minifish has a complete set of expressed co-receptor genes *cd8a*, *cd8b*, *cd4-1*, and *cd4-2* (see the Supplementary Materials). Although it was not possible to determine the relative proportions of presumptive cytotoxic and helper lineages, these findings suggest that the two canonical sublineages of αβ T cells are maintained in this small vertebrate; moreover, the presence of *foxp3a*- and *foxp3b*-related genes (see the Supplementary Materials) suggests further functional subdivisions among helper subsets. Collectively, these results indicate that the canonical diversity of teleost T cell lineages is maintained in minifish and suggest that immune homeostasis can be established even if each of the functional sublineages comprises at most a few thousand cells.

Characteristics of *tcr* gene assemblies

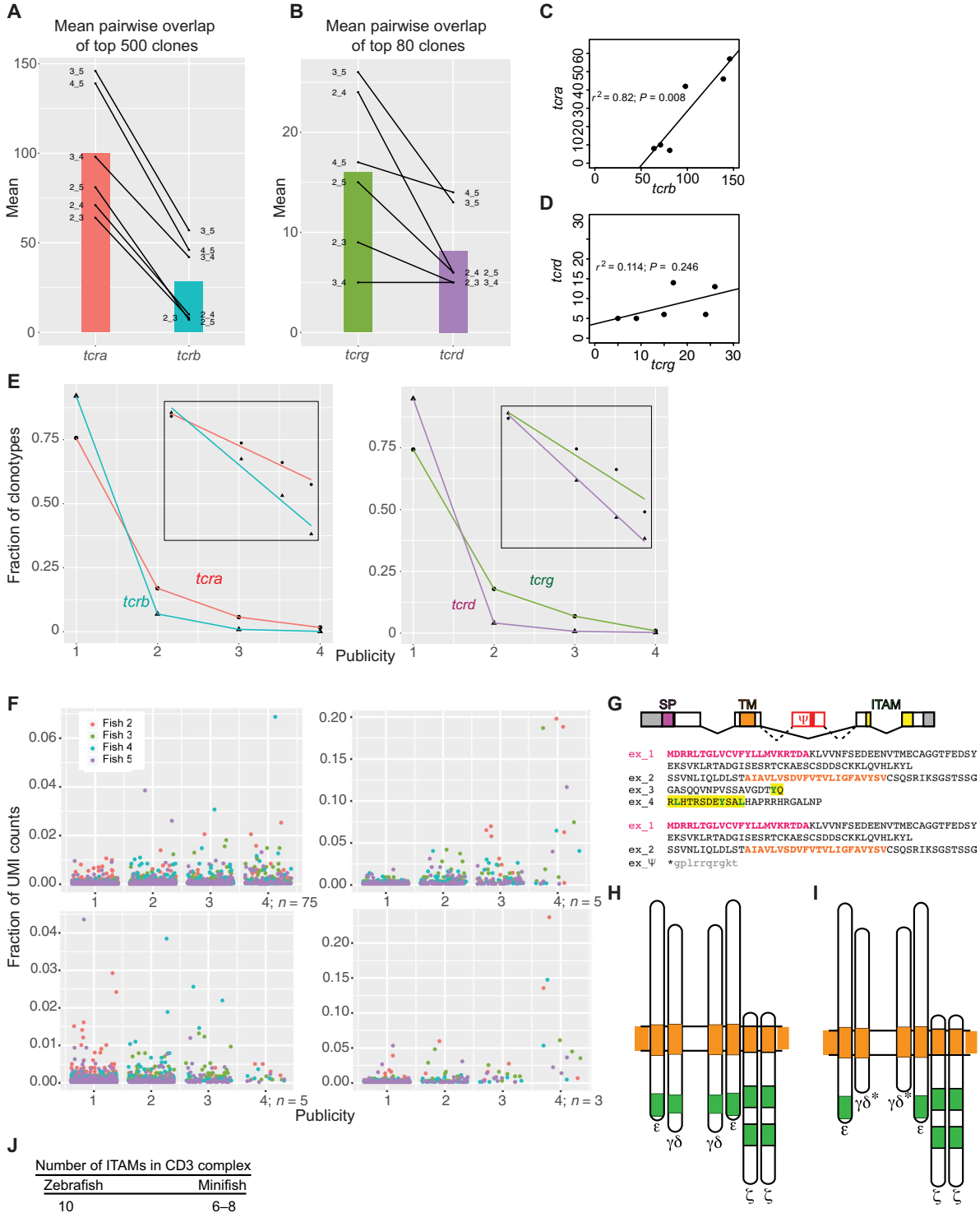
Detailed inspection of *tcrg* and *tcrd* sequences exhibits P nucleotides and N-region additions at the coding joints (fig. S3), substantially increasing the limited combinatorial diversity (Figs. 1A and 2) of these chains. The length distributions of CDR3 regions of both chains are heavily skewed, particularly when the number of molecules is taken into consideration (fig. S4). A total of 4 of 52 V elements in the Vα/δ gene cluster were exclusively found in functionally assembled *tcrd* transcripts, in addition to an additional 4 elements that were predominantly (ratio of *tcrd*/*tcra* usage, >10) associated with this chain (fig. S1B). This indicates that ~(8/52=) 15% of Vα/δ elements are associated with *tcrd* assemblies, similar to the estimated proportion of γδ T cells. The low numbers of *tcrg* and *tcrd* clonotypes precluded a meaningful entropy analysis.

The length distribution of the CDR3 regions in *tcrb* assemblies assumes a Gaussian shape, with a mean value of 13.3 ± 1.3 (means ± SD) amino acid residues (fig. S4). Entropy analysis based on amino acid sequences indicated that the contributions of the V and J regions amount to 2.87 and 2.66 bits, respectively, with the internal segment of the CDR3 regions (two regions of P and N nucleotides and one D region; see fig. S3) contributing an additional 12.5 bits (~70% of total entropy). The total entropy H of tcrb sequences amounts to 18.05 bits and is similar to that estimated for the igm repertoire; in analogy, this figure suggests that minifish can generate a minimum of ($2^{18.05}$~) 270,000 different tcrb clonotypes; since this number likely exceeds the number of αβ T cells in these animals, the full tcrb repertoire can only be realized on a population basis.

The frequencies with which individual Vα and Jα elements are used in *tcra* assemblies were found to consistently vary across the locus (fig. S5A), as observed for *tcrb* assemblies (fig. S5B). A total of 44 Vα/δ elements that are exclusively or predominantly used in *tcra* assemblies (fig. S1) combine with 61 Jα elements (Figs. 1A and 2), generating a total of 2684 possible VαJα combinations. Among the *tcra* assemblies of the four fish analyzed here, approximately 55% of these combinations were found. Despite variable degrees of usage of the two elements (fig. S5A), the V-J combinations are essentially random; this can be deduced from the low value of their mutual information (0.39 bits) in comparison to their joint entropy (9.93 bits) (fig. S5C). The overall length of the CDR3 region of *tcra* assemblies was found to be 13.1 ± 1.2 (means ± SD) amino acid residues (fig. S4), identical in size to that of *tcrb* chains. Since ~75% of *tcra* chains exhibited neither P nor N nucleotides at the junctions (fig. S3), combinatorial diversity is the dominant mechanism of diversity generation, with additional contributions to diversity by nucleotide deletions at the V-J junctions. Accordingly, entropy analysis based on amino acid sequence indicated that the CDR3 region contributed only 3.4 bits (~25% of total entropy) to the 4.8 and 5.8 bits of entropy furnished by V and J segments, respectively. The total entropy H of 14 bits for *tcra* chains suggests that minifish can generate a minimum diversity of (2^{14}~) 16,000 different *tcra* clonotypes, close to the number of αβ T cells in this animal. This result indicates that in contrast to the situation of *tcrb* clonotypes, the T cells in each minifish fish express a large fraction of the entire *tcra* repertoire that can be generated in this species' immune system.

High degree of publicity in the *tcr* repertoire

To gain insight into the composition of the *tcr* repertoires, we determined—at the nucleotide level—the frequencies with which individual clonotypes were recovered by sequencing. The *tcr* repertoires of minifish are dominated by a small number of frequent clonotypes, whereas most other clonotypes are of low frequency, a typical feature of a power-law distribution (Fig. 1D). Hence, we expect that additional clonotypes that were not recovered by our sequencing strategy likely will belong to the low-frequency class. Next, we determined the degree of overlap in the *tcr* repertoires among the four individuals analyzed here. Sequences found in at least two individuals of a population are commonly defined as public clonotypes (*34*). Pairwise comparisons of nucleotide sequences indicated that, on average, about 100 clonotypes (range, 70 to 145) of the 500 most frequent *tcra* clonotypes and about 25 clonotypes (range, 7 to 50) of the 500 most frequent *tcrb* clonotypes are shared (Fig. 3A). For the *tcrg* and *tcrd* repertoires, we found that 16 (range, 5 to 26) and 8



J

Number of ITAMs in CD3 complex	
Zebrafish	Minifish
10	6–8

Fig. 3. Publicity in the TCR repertoires. (**A** and **B**) Pairwise comparisons of the top 500 clonotypes each of *tcra* and *tcrb* (A) and *tcrg* and *tcrd* (B). (**C** and **D**) Correlation of shared clonotypes for six two-way comparisons of the four fish for tcra/trcb (C) and tcrg/tcrd (D). (**E**) Proportion of unique and clonotypes shared among two, three, or four individuals. Inset: log-log plot of data. The slopes are indicative of the fractal dimensions. (**F**) Prevalence of unique and clonotypes shared among two, three, or four individuals, identified by their origin in color code. The number of clonotypes that are present in all individuals is indicated (see table S3). *tcra* (top left), *tcrb* (bottom left), *tcrg* (top right), *tcrd* (bottom right). (**G**) Schematic of the *cd3gd* gene structure with coding exons, poison exon (ψ), splicing patterns, and functional protein domains indicated. SP, signal peptide; TM, transmembrane domain; ITAM, immune receptor tyrosine-based activation motif. (**H**) Schematic of the cognate minifish cd3 protein complex comprising eight ITAM motifs, modeled according to the octameric structure in 1:1:1:1 stoichiometry of TCRαβ:CD3γε:CD3δε:CD3ζζ (*38*) of the human TCR-CD3 complex. TM domains are indicated by orange squares, ITAM motifs by green squares, the cell membrane is indicated by two straight lines. (**I**) Schematic of the alternative minifish cd3 complex with six ITAM motifs; the variant cd3gd protein is highlighted by asterisk (*). (**J**) Number of ITAM motifs in CD3 complexes of zebrafish and minifish.

(range, 5 to 14) of the 80 most frequent clonotypes, respectively, are shared by any two individuals (Fig. 3B).

The extents of shared clonotypes for *tcra* and *tcrb* in two-way comparisons between different pairs of individuals are highly correlated (Fig. 3, A and C); moreover, the usage of V and J elements of the two chains is nearly identical among all individuals (fig. S5, A and B). Since the CDR3 regions of *tcra* sequences exhibit few random nucleotide additions, a substantial degree of overlap of clonotypes between individuals is observed; the same is true for the CDR3 regions of *tcrb*, despite the presence of a D element (Fig. 3, A and C). Hence, the degree of publicity of both *tcra* and *tcrb* clonotypes is likely determined by the respective constellation of *mhc* genes (fig. S6) of each individual. The strong correlation ($r^2 = 0.82$) of shared clonotypes in the six two-way comparisons of *tcra* and *tcrb* assemblies (Fig. 3C) illustrates the strong impact of peptide–major histocompatibility complex (MHC) complexes on the composition of the αβ TCR repertoire (*35*, *36*). As expected for the lack of MHC restriction in the γδ T cell lineage, a weak, if any, correlation for shared clonotypes of *tcrg* and *tcrd* assemblies was found (Fig. 3, B and D).

A comparison of nucleotide sequences of overlapping clonotypes among the four fish indicates that the patterns of publicity fall into two groups; *tcra* and *tcrg* sequences both have high publicity, whereas *tcrb* and *tcrd* sequences exhibit lower degrees of publicity (Fig. 3E). This finding suggests that two different sets of rules govern the generation of the repertoires of *tcra* and *tcrg* and *tcrb* and *tcrd*, respectively. These characteristics are best represented by the corresponding fractal dimensions, expressed in similar slopes of the log-transformed rank/frequency distributions for *tcra* and *tcrg* and *tcrb* and *tcrd*, respectively (Fig. 3E, insets). Collectively, these results suggest that αβ and γδ heterodimers exhibit a similar overall structural design. In the assemblies of all antigen receptor genes, public sequences tend to be associated with higher molecule counts than private clonotypes (Fig. 3F). However, the two types of TCR heterodimers differ by the frequencies with which individual clonotypes are represented in the repertoires of individual fish; the frequencies of fully public clonotypes of *tcrg* and *tcrd* are almost always higher than those of private clonotypes (Fig. 3F and table S4). Although we cannot distinguish whether this is due to preferential generation of certain assemblies or their subsequent selection, this result demonstrates that the γδ TCR lineage is dominated by a small number of prevalent clones that are identical for all fish (fig. S4).

Structure of the CD3 signaling complex

Our analysis of *tcr* assemblies suggests that αβ and γδ heterodimers are both composed of one chain with restricted diversity (encoded by *tcra* and *tcrg*), whereas the other chain is much more variable (encoded by *tcrb* and *tcrd*). In analogy to the situation of semi-invariant TCRs described in mammals, such as those characterizing invariant natural killer T (iNKT) cells (*37*), we considered the possibility that the unusual properties of the T cell repertoire of minifish may be associated with the recognition of restricted sets of antigens. In this scenario, one would expect a substantial degree of receptor cross-reactivity, possibly necessitating further adaptations, for instance, in the components of the signal transduction pathway(s) to fine-tune the antigen response. To this end, we focused on the CD3 signaling apparatus of the TCR (*38*). The minifish cd3ε chain exhibits the characteristic single immunoreceptor tyrosine-based activation motif (ITAM), whereas the two paralogs encoding the cd3ζ component both encode only two ITAMs (Fig. 3, G and H, and see the Supplementary Materials), instead of the more common three ITAM/two ITAM constellation in the closely related cyprinid *D. rerio* (*39*). In addition to this hard wired modification, further studies led to the discovery of an unusual splicing event in the *cd3gd* gene (Fig. 3G), which represents the evolutionary ancestor of the distinct *CD3G* and *CD3D* genes in mammals. In addition to the canonical transcript, we recovered an alternatively spliced version, incorporating a cryptic "poison" exon (Fig. 3G). The conceptual translation of this variant transcript reveals an in-frame stop codon and predicts a variant cd3gd protein that retains the transmembrane domain but lacks the characteristic ITAM motif (Fig. 3I). As a result, instead of 10 ITAMs per typical cyprinid CD3/TCR complex (Fig. 3H), conditional splicing events make it possible to adjust the number of ITAMs to between six and eight (Fig. 3, I and J), a constellation that would allow the titration of the strength of downstream signal transmission after TCR engagement (*40*).

Structure of the antigen receptor repertoires

The small numbers of minifish antigen receptor clonotypes offer an unprecedented opportunity to achieve a near complete description of their network structure. Following previous studies, we focused on the CDR3 regions of individual clonotypes. To this end, the conceptually translated sequences of individual clonotypes were collapsed into one node, when their CDR3 sequences were identical, to which we refer as a CDR3 clonotype. Pairs of nodes were then connected by an edge, when they were separated by one amino acid difference [Levenshtein distance of 1 (*41*)], that is, by replacement, deletion, or addition of one amino acid. The networks of all five antigen receptor chains thus constructed formed clusters of nodes typically containing many V segments (Fig. 4A) but only one or two J segments (Fig. 4B); the insets in Fig. 4 (A and B) illustrate this phenomenon for the largest cluster of tcrb CDR3 clonotypes. This structure emerges as a result of the fact that the sequence diversity of the CDR3 region is dominated (but not exclusively determined) by the distinct sequences found in the 5′ ends of each J segment as opposed to the relatively uniform amino acid sequences that are present at the 3′ ends of V elements. This phenomenon was previously also described for mouse and human networks (*42*), suggesting that it represents a fundamental design principle of the immune system. However, since antigen receptor genes differ in the number of V and J segments, the number of individual clonotypes, and the overall structure of the CDR3 region [particularly the presence or absence of D segment(s) and the extent of addition of P and nontemplated N nucleotides], the resulting network architectures differ among the antigen receptor genes (Fig. 4, C to F, and table S5). In all four fish, the igh network is dominated by one giant component that contains three of the four J elements and connects 63.51 ± 13.47% ($n = 4$ fish; means ± SEM) of all nodes (range, 43.3 to 71.2%) (Fig. 4C); accordingly, the cluster sizes for the igm network show a marked bimodal distribution (Fig. 4D and table S5). Overall, this results in a situation very similar to what has been described in mouse networks (*7*). The average degree of connectivity, that is, the number of edges connected to a node, varies between 1.9 and 4.6 across the four fish, whereas the corresponding maximum degree of connectivity in the network varies between 18 and 44 (table S5).

Owing to the low sequence diversity of tcrg CDR3 clonotypes, only a minority remains unconnected in the network; individual nodes are connected by one of the two J segments and organized in

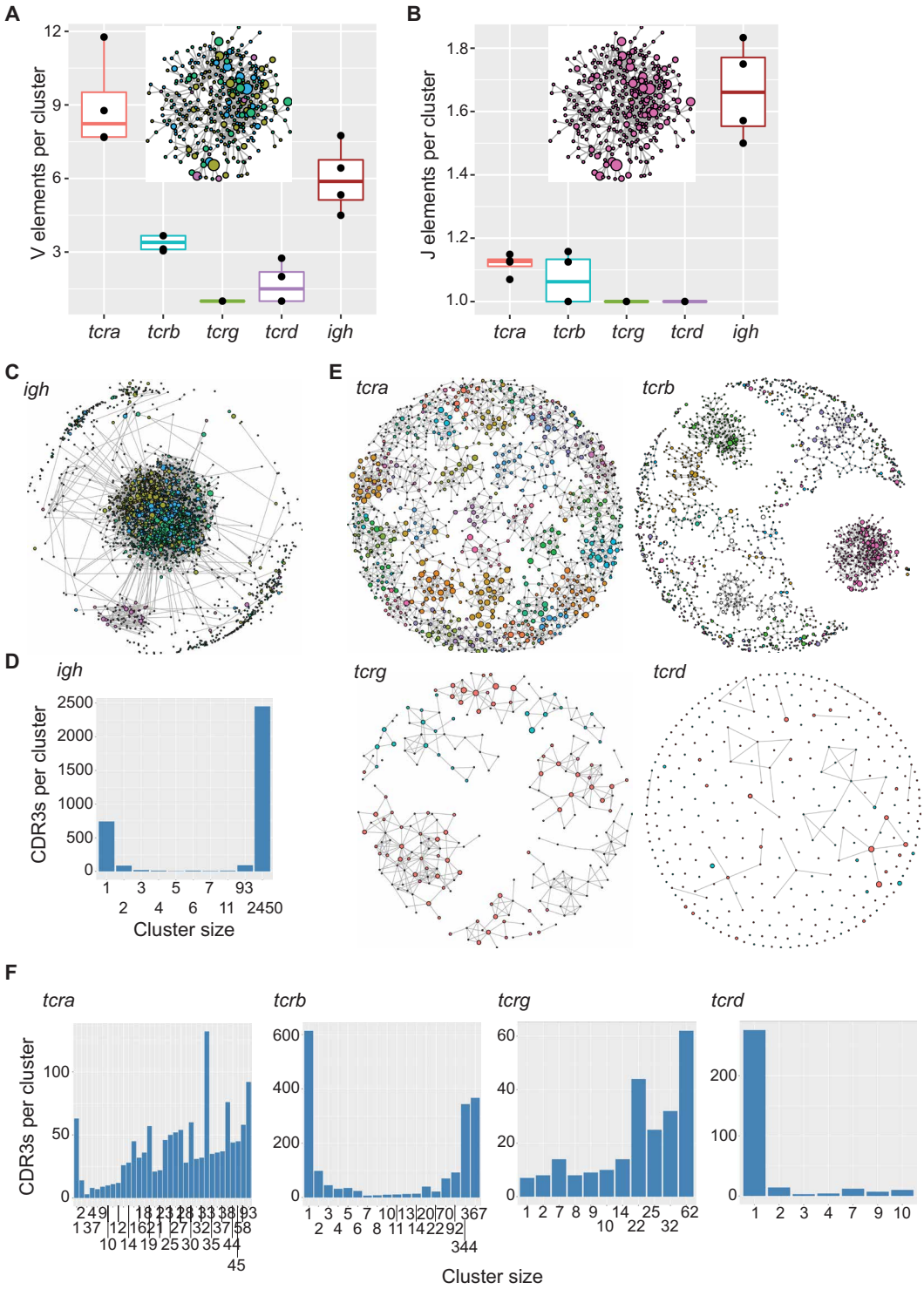


Fig. 4. Structure of antigen receptor networks. (**A**) Numbers of variable genes (V) per cluster of connected CDR3 sequences in four fish. (**B**) Numbers of joining genes (J) per cluster of connected CDR3 sequences in four fish. In (A) and (B), individual values are indicated by dots. The box plot indicates the mean and 25 and 75 percentiles. (**C**) Network of connected igh CDR3 sequences. (**D**) Distribution of the number of igh CDR3 sequences according to cluster size (indicated at the bottom). (**E**) Networks of connected CDR3 sequences of the four tcr assemblies. (**F**) Distribution of the number of CDR3 sequences of the four tcr assemblies according to cluster size (indicated at the bottom). In the display of tcrb network, 300 nodes situated far away from the central nodes are not shown. In (C) and (E), the size of the dot indicates the degree of publicity; unconnected nodes are small, and fully public clones are indicated by the largest diameter; individual J elements are indicated by different colors.

several distinct clusters that do not coalesce (Fig. 4, E and F). This peculiar archipelago-like structure is not seen with tcrd CDR3 clonotypes; here, the network comprises mostly unconnected nodes as a result of the vast potential diversity of trcd assemblies (Fig. 4, E and F). The network of tcra CDR3 clonotypes is again composed of distinct clusters, mostly determined by one J element. However, as a consequence of a general lack of P and N nucleotides at the junctions and the dominance of particular V-J recombinations, these clusters rarely coalesce (Fig. 4, E and F). In this regard, the network structures of tcra and tcrg are similar, reinforcing the conclusion that they are built according to similar rules. Moreover, the comparable network organization suggests that in their respective heterodimeric constellation, they are expected to make a smaller contribution to the capacity of antigen discrimination than their partner chains. The network of tcrb CDR3 clonotypes exhibits the most complex structure, combining features seen in other networks. Cluster sizes follow a bimodal distribution, with a substantial fraction of unconnected nodes and contributions of several large clusters that are dominated by a single J element each (Fig. 4, E and F).

The central position of public clonotypes in networks

Next, we considered the position of public CDR3 clonotypes in the networks of the antigen receptor assemblies. Irrespective of the variable distributions of cluster sizes observed for the five genes, in the respective networks, public CDR3 clonotypes are universally associated with the larger clusters (Fig. 5A). This apparent centrality of public sequences was previously observed for mammalian antigen receptor gene repertoires (*7*, *42*) and may thus represent a general design principle of antigen receptor repertoires. The increase in node connectivity associated with publicity was most pronounced for tcrb and igh assemblies (Fig. 5B). This trend is further illustrated by the observation that, for fish no. 5, all 118 nodes that are present in all fish (publicity degree 4; red dots in Fig. 5A) are found in networks clusters, identifiable as the large circles in the networks shown in Fig. 3 (C and E); for the other three fish, a maximum of four of these 118 nodes are unconnected.

The stability of networks

Next, we addressed the stability of the networks. In a first set of simulations, we removed all public CDR3 clonotypes from the networks of all five antigen receptor genes and examined the changes in the distributions of the degrees of connectivity. For instance, in the case of the igm network of fish no. 5, this led to the removal of 596 of 3440 nodes (~17%) (Fig. 6A). As expected from the highly connected network structure of igm CDR3 clonotypes, the maximum degree of connectivity was reduced by about 55% (Fig. 6B). By contrast, randomly removing the same number of private clones had

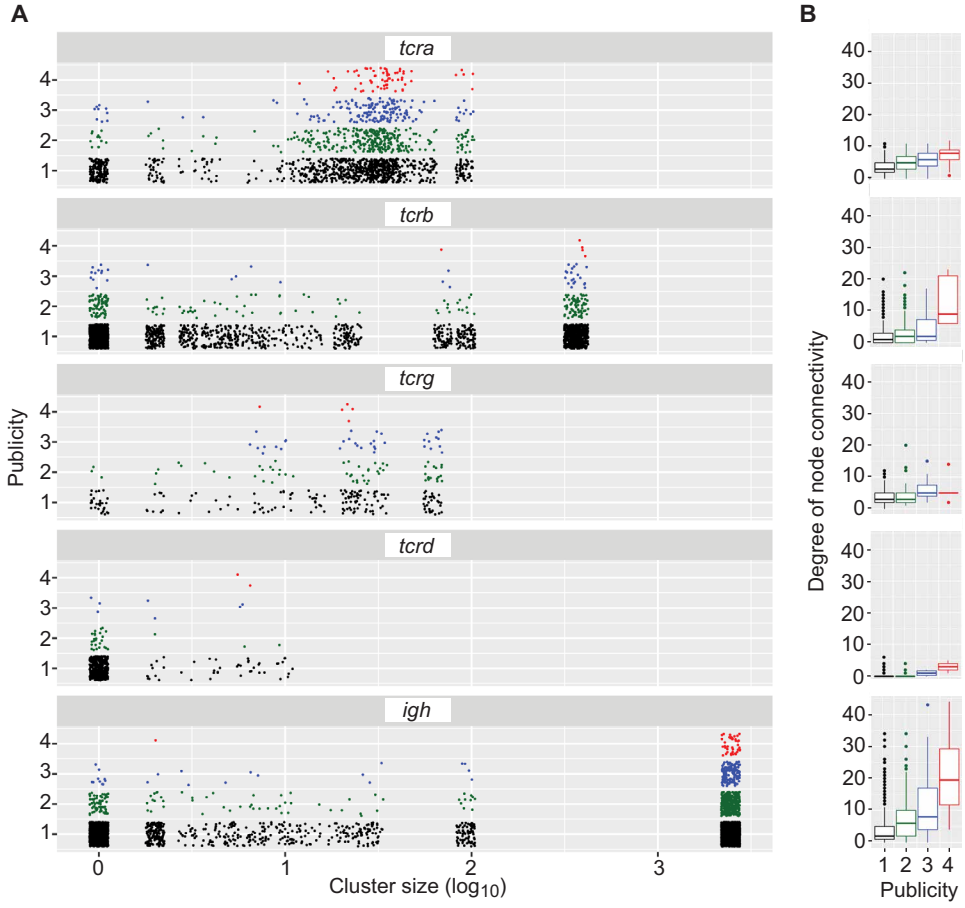


Fig. 5. Public clones are associated with larger clusters. (**A**) Distribution of individual CDR3 sequences of the five antigen receptors according to cluster size and degree of publicity (color-coded). (**B**) Summary statistic of the degree of connectivity of CDR3 clonotypes according to their publicity.

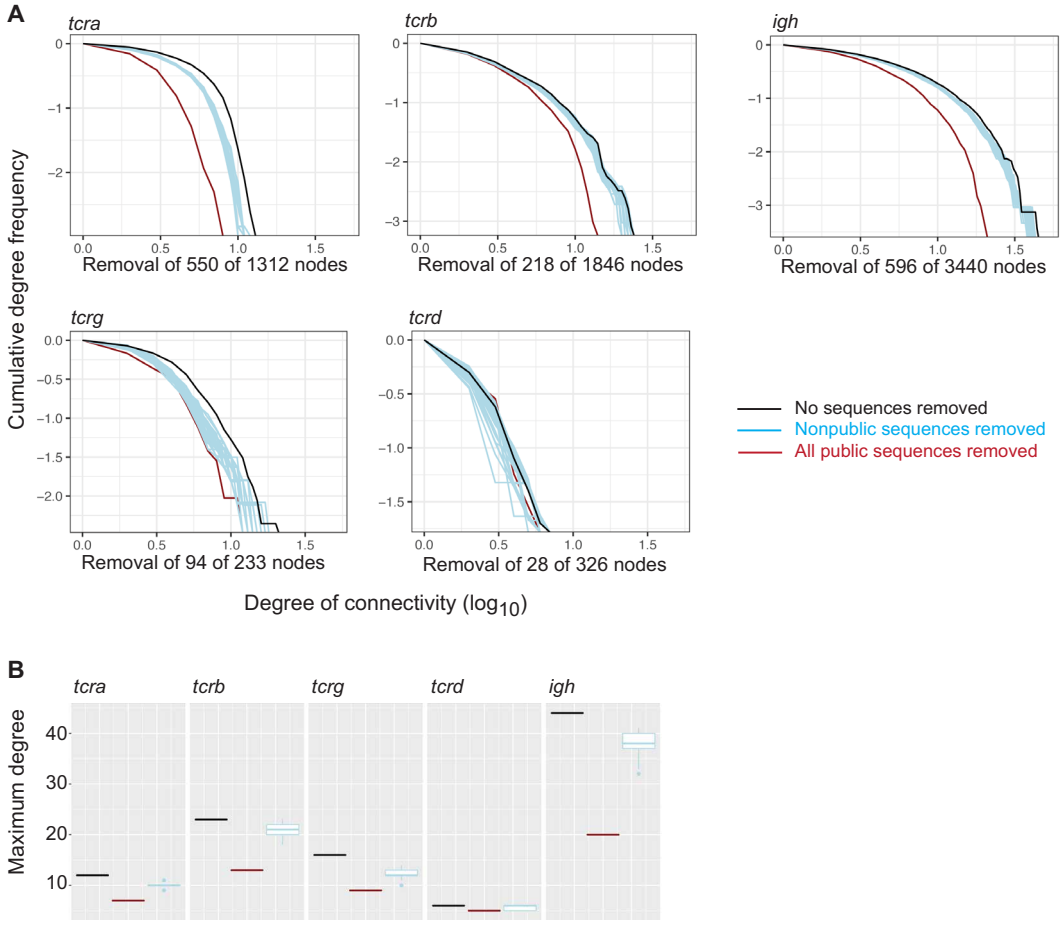


Fig. 6. Stability of antigen receptor networks. (**A**) The degree of network connectivity is a measure of network structure. The cumulative frequency distribution is shifted to the left, if removal of nodes reduces connectivity. In all but one network, removal of all public CDR3 clonotypes reduces the maximum degree of connectivity. Removal of the same numbers of nonpublic sequences (40 iterations of randomly chosen sequences are shown in the blue lines) has a less marked effect. (**B**) Summary statistic of the maximum degree of connectivity in antigen receptor networks after removal of public and nonpublic CDR3 clonotypes as shown in (A).

correspondingly little impact (~14%; Fig. 6, A and B). This notably different outcome is reproducible across different fish and highly significant (fig. S7), even when only two-thirds of public clones (and a similar number of nonpublic clones) are removed. These results echo the observations of Miho *et al.* (*7*) in Igh networks of mouse and human after removal of public clones.

A similarly marked reconfiguration of the connectivity of network structures of tcra, tcrb and, to a lesser extent, of tcrg was observed after the removal of all public CDR3 clonotypes (Fig. 6, A and B), whereas the network connectivity remained largely unchanged after removal of an equivalent number of nonpublic CDR3 clonotypes. By contrast, removal of public and nonpublic CDR3 clonotypes had an equally minor effect in the tcrd network (Fig. 6, A and B), as expected from its largely unconnected configuration (Fig. 4, E and F). Collectively, our studies reaffirm the centrality of public CDR3 clonotypes in the networks of igm and tcra, tcrb, and tcrg clonotypes as a general design principle.

Impact of J element numbers on network structure

As shown in Fig. 1B, the antigen receptor loci of minifish are characterized by a much reduced number of J elements when compared to zebrafish, yet the CDR3 clonotype clusters in the networks are often dominated by one or few J elements (Fig. 4). State-of-the-art prediction algorithms of antigen specificity (*43*–*45*) have assigned a prominent role to CDR3 regions of TCR β chains, although the TCRdist algorithm (*45*) also takes CDR1 and CDR2 regions into account. Since clusters of related tcrb sequences typically contain only one or two J elements, loss or gain of J elements can have a substantial effect on the functional capacity and structure of the antigen receptor repertoire. Whereas a reduction of J β elements would result in greater connectivity (perhaps more akin the igh network) and, hence, a much more focused repertoire, a larger number of J elements would lead to a more fragmented structure, with much reduced cluster sizes, approaching the structure of the tcra network. We therefore propose that the number and kind of genetic elements available for tcra and tcrb assemblies are linked to the number of T cells in a species, to optimally achieve antigen discrimination and hence recognition in the context of MHC peptide presentation. Interspecific comparisons will be required to determine the scaling factors underlying this relationship.

DISCUSSION

Our study uncovers a number of unexpected features of the immune system of one of the smallest known vertebrates. Despite its miniature body size and the correspondingly small numbers of

lymphocytes, the compact minifish genome encodes the key elements of a canonical vertebrate adaptive immune system. In line with the small size of the lymphocyte compartment, the numbers of individual elements in the antigen receptor loci are substantially reduced. The two T cell lineages of minifish both express receptors composed of one chain with limited diversity (tcra and tcrg) and a second chain with considerably greater diversity (tcrb and tcrd). However, the reduction of diversity is achieved differently, by reciprocal contributions of combinatorial and junctional diversity. For *tcra*, few instances of random nucleotide addition are found in CDR3 regions, and diversity is mainly driven by combinatorial pairing of V-J elements; by contrast, few combinatorial choices exist for *tcrg* sequences that, instead, rely primarily on junctional diversity. Given that large fractions of the repertoires of the less diverse antigen receptor chains are expressed in each fish, it is likely that minifish antigen receptors have evolved to function as dual-purpose devices, with the structurally limited receptor functioning as a kind of pattern recognition receptor tuned to a restricted set of antigens and with the more diverse receptor chains modulating the strengths of these interactions. This recognition mode may mitigate the risk of self-reactivity in the repertoire, especially when operating in concert with additional adaptations, such as the reduced signaling capacity of the CD3 complex.

The unique design of the antigen receptor repertoire described here appears to be an efficient solution for a miniaturized immune system, which must ensure immune protection with a relatively small number of lymphocytes. This strategy appears to have been deployed also in other situations; the dominant role of (semi-) invariant lymphocytes in minifish is reminiscent of the structure of the T cell compartment in amphibian tadpoles that confers strong antiviral immunity despite its small size (*46*). Therefore, the immunogenetic features described here may also be found in the microhylid frog *Paedophryne amauensis*, which rivals minifish in miniature body size (*47*). However, even in humans, a certain degree of publicity of TCR clonotypes is not uncommon (*7*, *36*, *48*), with T cells expressing semi-invariant receptors, such as iNKT (*37*) and mucosa associated invariant T (MAIT) (*49*) cells representing extreme examples of evolutionary exploitation of the same general principle.

Last, our analysis of the minifish immune system lends strong support to the notion that the self-similar (fractal) nature of the antigen receptor repertoire is a general property of the vertebrate immune system. This finding has important functional implications. First, a fractal design provides the flexibility required to accommodate immune systems with small or large number of lymphocytes (*50*); the body mass (and likely also lymphocyte numbers) of human and minifish differs by seven orders of magnitude. Second, a fractal organization confers robustness to the system when cells are lost (*51*) or clonally fluctuate, which are the functional equivalents of transient relative expansion and corresponding diminutions of individual clones that are typically associated with immune responses. No information is available about the characteristics of the immune response in minifish; however, our simulations suggest that network structures are stable even when a considerable fraction of clonotypes is lost. These unique features ensure the maintenance of overall immune reactivity even in the face of significant perturbations (physiological transient expansion/contraction of clonotypes; loss of lymphocytes as a consequence of cytopathic insults), helping to explain the remarkable success of adaptive immunity in vertebrates of vastly different body sizes.

MATERIALS AND METHODS

Animals

All fish work followed the Guidelines on the Care and Use of Animals for Scientific Purposes of the National Advisory Committee for Laboratory Animal Research in Singapore. Work with minifish was approved under permit number 065/06. Minifish specimens were collected from the Sumatran Singkep Island, supplied by a local dealer, and kept in the laboratory (*23*) before they were flash-frozen in liquid nitrogen and used for DNA and RNA extraction.

Genome sequencing

Genomic DNA was extracted from a male and a female individual each using the phenol/chloroform extraction method, followed by ethanol precipitation. One microgram of DNA was used to prepare a polymerase chain reaction (PCR)–free library using the KAPA HyperPrep Kit (Kapa Biosystems, Wilmington, MA). The DNA was sheared using M220 Focused-ultrasonicator (Covaris, Woburn, MA), followed by double-sided size selection with Agencourt AMPure XP (Beckman Coulter, Brea, CA) to obtain 500–base pair (bp) inserts. For the male minifish, a total of 140 million paired-end reads of 250-bp length were generated on an Illumina HiSeq 2500 platform. These reads were quality filtered for low-complexity sequences resulting in ~138 million paired-end reads. For the female minifish, the same pipeline resulted in ~175 million good-quality paired-end reads. Details on the estimation of genome size and heterozygosity levels, genome assembly (scaffolding, gap-filling, evaluation of completeness, repeat content prediction, and annotation), and RNA sequencing (RNA-seq) are reported in Supplementary Materials and Methods.

cDNA synthesis and barcoding of RNA samples

Total RNA from four minifish individuals were extracted separately using the TRIzol reagent (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. The RNA was treated with deoxyribonuclease I (New England Biolabs, Ipswich, MA, USA) before cDNA synthesis. Fifty percent of the total RNA extracted from each fish (~3 to 8 μg) was used for cDNA synthesis in reactions containing no more than 0.5 μg of total RNA. cDNA synthesis was performed using the SMARTScribe Reverse Transcriptase (Clontech, Mountain View, CA, USA) with an oligo-dT primer (5′-AAGCAGTGGTATCAACGCAGAGTTTTTTTTTTTTT TTTTTTTTTTVN) and SMARTer_Oligo_UMI primer (5′-AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUNNNNUCTT[rGrGrGrG]) according to the SMARTer RACE 5'RACE protocol (Clontech, Mountain View, CA, USA). The SMARTer_Oligo_UMI is a hybrid primer with riboguanosines representing the last five bases and the remainder representing deoxyribonucleotides, including the U (deoxyuracil); the Ns represent the bar code. The cDNA synthesized was treated with uracil-DNA glycosylase before all reactions from the same individual were combined together. The combined cDNA was purified using the QIAquick PCR Purification Kit (QIAGEN, Hilden, Germany), eluted with 70 μl of diethyl pyrocarbonate (DEPC)–treated water, and vacuum-dried.

Library preparation

The cDNA samples were reconstituted in 100 μl of DEPC water, and 64 μl each was used in 32 amplifications of antigen receptor gene transcripts, essentially according to the protocol of Turchaninova *et al.* (*52*), except that Illumina multiplexing primer sequences p5

(5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT) and p7 (5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT) were appended to the 5′ ends of their second reaction primers. In this way, a total of approximately one-third of the original total RNA material per fish was subjected to analysis. The first round of PCR amplification was carried out in multiplex manner: 1× Q5 buffer, 0.5 mM deoxynucleoside triphosphate (dNTP), 0.2 μM UPM_S primer (5′-CTAATACGACTCACTATAGGGC), 0.04 μM UPM_L primer (5′-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT), and 0.2 μM of each gene-specific primer (GSP), 2 μl of cDNA, water to 49.5 μl, 0.5 μl of Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs); 98°C for 90 s followed by 23 cycles of 98°C for 10 s, 65°C for 20 s, and 72°C for 45 s, followed by 8-min final extension at 72°C. GSPs used in the first round were Mf_a_R_1 (*tcra*, 5′-CCAAAAAGCCGCCGTGCTGCTTAACGC), Mf_b_R_1 {*tcrb1* [in cyprinids, few transcripts contain *Cb2* sequences (*32*)], 5′-CTGAAGCCACACATGTGAGTGTCCGGTG}, Mf_g_R_1 (*tcrg*, 5′-CCAGCTGCATCTTTCCATTCTCCCGTGCTG), Mf_d_R_1 (*tcrd*, 5′-CAGTTCTCAATGGGGGAATCGTTGAAGCCAGC), and OBG_101(*igm*, 5′-CTCAGTGAGCTGATTCTGGTG). Amplicons were size-separated on agarose gels, the region between 500- and 1000-bp excised, and the DNA was extracted using the QIAquick Gel Extraction Kit (QIAGEN) following the protocol provided by the manufacturer (with two PE washes) and lastly eluted in 50 μl of water. For the second round of PCR amplification, each target locus was amplified separately. For each locus, 2% of the first-round amplicon material (1 μl) was used for 50 μl of reactions, using 0.2 μM (combined final concentration) of an equimolar mix of P7 + UPM_S_4N (5′-gtgactggagttcagacgtgtgctcttccgatctNNNNCTAATACGACTCACTATAGGGC), P7, UPM_S_5N (5′-gtgactggagttcagacgtgtgctcttccgatctNNNNNCTAATACGACTCACTATAGGGC), and P7 + UPM_S_6N (5′-gtgactggagttcagacgtgtgctcttccgatctNNNNNNCTAATACGACTCACTATAGGGC) primers together with 0.2 μM GSPs; other conditions were as for the first round except that amplification was performed for only 20 cycles at an annealing temperature of 55°C. GSPs used in the second round were Mf_a_R_2 + P5 + 4 N (*tcra*, 5′-acactctttccctacacgacgctcttccgatctNNNNCCATTGTCAACCTTGTAAATAGC), Mf_b_R_2 + P5 + 4 N (*tcrb*, 5′-acactctttccctacacgacgctcttccgatctNNNNNTCTTACAACTCTCCTTAACATGGG), Mf_g_R_2 + P5 + 4 N (*tcrg*, 5′-acactctttccctacacgacgctcttccgatctNNNNNNCTTGTCTTCTGACTGGTACACCGAC), Mf_d_R_2 + P5 + 4 N (*tcrd*, 5′-acactctttccctacacgacgctcttccgatctNNNNCTTGGCAAGACTGACAGAACAGG), and OBG100 + P5 + 6 N (*igm*, 5′-acactctttccctacacgacgctcttccgatctNNNNNNGACGATGGTCCAGATGGTG). The resulting material was purified with AMPure XP beads (0.65×) and barcoded with NEBNext multiplex oligonucleotides for Illumina. Last, gel purification was used to avoid sequencing fragments shorter than 500 bp in the sequencer. Paired-end sequencing was performed in an Illumina MiSeq instrument at a read length of 300 bp.

Minifish MHC sequences were amplified from cDNA and sequenced on an Illumina MiSeq platform, after barcoding using the NEBNext multiplex oligos for Illumina (New England Biolabs).

Analysis of antigen receptor assemblies

For the extraction of the sequences, an R pipeline was developed that is available at GitHub (https://github.com/obgiorgetti/minifish). Briefly, unique molecular identifier (UMI) barcodes were used to account for the numbers of cDNA molecules by matching the sequences of UMI, CDR3 region (including the entire J sequence), and a V gene sequence identified from the dictionary search. Each unique combination of UMI, V, and CDR3 (including the J) was considered to represent a single cDNA molecule but was kept for analysis only if read at least three times and was otherwise discarded. Sequences with UMIs at a distance of one nucleotide and CDR3 sequences at a distance of two nucleotides or less were considered errors; in these instances, only the variant with highest numbers of reads was retained (note, however, that reads not considered after this cutoff are nonetheless contained in the deposited sequence collections to be found at www.ncbi.nlm.nih.gov/sra/PRJNA612865).

For repertoire analysis, the paired 5′- and 3′- ends of the molecules were not joined but mapped to the V segments separately. The CDR3 region of igm sequences was operationally defined as the sequences occurring between and including the characteristic C-terminal cysteine of V elements and the characteristic tryptophan residue in J region sequences; for tcr sequences, the CDR3 region was operationally defined as the sequences occurring between and including the characteristic C-terminal cysteine of V elements and the characteristic phenylalanine residue in J region sequences.

Entropy analysis for *ig* and *tcr* assemblies

Given the random variables [S, complete Ig or TCR sequence; CDR3, defined as a sequence from and including cysteine to tryptophan (Ig) or phenylalanine (TCR) residues; V, V gene; J, J gene;

L = CDR3 length (where sequence elements and their lengths are either amino acid or nucleotide residues)], we estimate the entropy H that a given Ig or TCR system S can generate as follows

$$\begin{aligned} H(S) &= H(CDR3, V, J) \\ &= H(CDR3 \mid V, J) + H(V, J) \end{aligned}$$

For each l in L

$$H(S \mid L = l) = H(CDR3 \mid V, J, L = l) + H(V, J \mid L = l) = H(CDR3 \mid L = l) - I(CDR3; V, J \mid L = l) + H(V, J \mid L = l)$$

however, instead of calculating

$$I(CDR3; V, J \mid L = l)$$

which would require a large number of clones for each V-J pair, we take the maximum value of

$$I(CDR3_n; V \mid L = l) \text{ or } I(CDR3_n; J \mid L = l)$$

for each position n of CDR3. This substitution is justified, because V and J have low mutual information content, as observed in our data (fig. S5).

The sum over all l in L gives

$$\sum_{l \in L} p(l)\, H(S \mid L = l)$$

and lastly

$$H(S) = H(L) + H(S \mid L) - H(L \mid S)$$

where the H(L|S) is 0, because if the sequence is known, then its length is also known.

Network analysis

To generate a network of sequence similarity, clones where collapsed into one node when their amino acid CDR3 sequences were identical

(irrespective of the particular V or J segments used in the assembly). Nodes representing CDR3 sequences at a Levenshtein distance of 1 were connected by edges, resulting in an undirected graph. Sequences containing stop codons and out-of-frame rearrangements were excluded from the analysis. For the network construction and analysis, the igraph package (*53*) was used. The code for the analysis is available at https://github.com/obgiorgetti/minifish.

Estimation of the number of lymphocytes in minifish

Because no live specimens of this species were available for cytological and histological analyses, we instead measured the number of T lymphocytes in zebrafish of about 3 weeks of age, which are similar in size and body weight to minifish, assuming that the cyprinid body plan and the general structure of the hematopoietic tissues are conserved between these two species. In *lck:CFP* zebrafish [Tg(lck:CFP)/fr104Tg], the fluorescent reporter marks T lineage cells; on average, 36,885 ± 14,794 (means ± SD; $n = 7$) cells were found, providing a numerical benchmark for the present analysis of the antigen receptor repertoires. The lck:CFP transgene was constructed by cloning a 5.8-kb fragment (*54*) upstream of the ATG initiation codon situated in exon 2 of the zebrafish *lck* gene into the pCS2:CFP vector (*55*).

Genetic relatedness of specimens

To avoid erroneous conclusions when analyzing a possible overlap of clonotypes between individuals, we assessed the degree of genetic relatedness by determining partial sequences of their *mhc* genes using the primers listed below. Reverse transcription PCR reactions were carried out under the following conditions: 1× Q5 buffer, 0.2 mM dNTP, 0.25 μM of each GSP, 0.2 μl of cDNA (equivalent to 1/1500 of total RNA), water to 49.5 μl, 0.5 μl of Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs); 98°C for 90 s followed by 32 cycles of 98°C for 10 s, 55°C for 20 s, 72°C for 40 s, followed by 8-min final extension at 72°C. *mhc1* sequences were amplified using primers MHC1a.5_F (5′-CACGGCCTCGTCAGGAATC) and OBG 28 (5′-CAAGAGACACGTCCTCGTGAAC); *mhc2a* sequences were amplified using primers OBG33 (5′-GTTACTCTGCCTGACTTCTCAG) and OBG38 (5′-GTCGGTACTGACTCAGACTG); *mhc2b* sequences were amplified using primers OBG40 (5′-TAGATGCCTCCACAGCGCTC) and OBG42 (5′-GATTGTTGACGCTGGCGTGTTC), OBG40 and OBG43 (5′-GAGTGGATCTGATAGTACCAGTC), OBG41 (5′-CGATCTGAGTGACATGGTGTTC) and OBG42, and OBG41 and OBG43, respectively. Although the primers do not capture all *mhc*-related sequences, the results indicated the presence of distinct sets of partially overlapping sequences (fig. S6), suggesting that the four fish included in the present analysis are outbred individuals, rather than clonally related.

Statistical analysis

The sample size for animal experiments was limited by the availability of wild-caught specimens of this uncommon species. The code underlying the antigen receptor analyses is available at https://github.com/obgiorgetti/minifish.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/1/eabd8180/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. T. Boehm, M. Hirano, S. J. Holland, S. Das, M. Schorpp, M. D. Cooper, Evolution of alternative adaptive immune systems in vertebrates. *Annu. Rev. Immunol.* **36**, 19–42 (2018).
2. G. W. Litman, M. K. Anderson, J. P. Rast, Evolution of antigen binding receptors. *Annu. Rev. Immunol.* **17**, 109–147 (1999).
3. M. D. Cooper, M. N. Alder, The evolution of adaptive immune systems. *Cell* **124**, 815–822 (2006).
4. J. L. Xu, M. M. Davis, Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* **13**, 37–45 (2000).
5. I. Engel, S. M. Hedrick, Site-directed mutations in the VDJ junctional region of a T cell receptor β chain cause changes in antigenic peptide recognition. *Cell* **54**, 473–484 (1988).
6. P. Bradley, P. G. Thomas, Using T cell receptor repertoires to understand the principles of adaptive immune recognition. *Annu. Rev. Immunol.* **37**, 547–570 (2019).
7. E. Miho, R. Roškar, V. Greiff, S. T. Reddy, Large-scale network analysis reveals the sequence space architecture of antibody repertoires. *Nat. Commun.* **10**, 1321 (2019).
8. A. Fischer, A. Rausell, Primary immunodeficiencies suggest redundancy within the human immune system. *Sci. Immunol.* **1**, eaah5861 (2016).
9. W. S. DeWitt III, A. Smith, G. Schoch, J. A. Hansen, F. A. T. Matsen, P. Bradley, Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *Elife* **7**, e38358 (2018).
10. G. Chen, X. Yang, A. Ko, X. Sun, M. Gao, Y. Zhang, A. Shi, R. A. Mariuzza, N.-P. Weng, Sequence and structural analyses reveal distinct and highly diverse human $CD8^+$ TCR repertoires to immunodominant viral antigens. *Cell Rep.* **19**, 569–583 (2017).
11. W. S. DeWitt, P. Lindau, T. M. Snyder, A. M. Sherwood, M. Vignali, C. S. Carlson, P. D. Greenberg, N. Duerkopp, R. O. Emerson, H. S. Robins, A public database of memory and naive B-cell receptor sequences. *PLOS ONE* **11**, e0160853 (2016).
12. O. V. Britanova, M. Shugay, E. M. Merzlyak, D. B. Staroverov, E. V. Putintseva, M. A. Turchaninova, I. Z. Mamedov, M. V. Pogorelyy, D. A. Bolotin, M. Izraelson, A. N. Davydov, E. S. Egorov, S. A. Kasatskaya, D. V. Rebrikov, S. Lukyanov, D. M. Chudakov, Dynamics of individual T cell repertoires: From cord blood to centenarians. *J. Immunol.* **196**, 5005–5013 (2016).
13. R. O. Emerson, W. S. DeWitt, M. Vignali, J. Gravley, J. K. Hu, E. J. Osborne, C. Desmarais, M. Klinger, C. S. Carlson, J. A. Hansen, M. Rieder, H. S. Robins, Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* **49**, 659–665 (2017).
14. B. Briney, A. Inderbitzin, C. Joyce, D. R. Burton, Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* **566**, 393–397 (2019).
15. C. Soto, R. G. Bombardi, A. Branchizio, N. Kose, P. Matta, A. M. Sevy, R. S. Sinkovits, P. Gilchuk, J. A. Finn, J. E. Crowe Jr., High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398–402 (2019).
16. M. M. Davis, P. J. Bjorkman, T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
17. R. Covacu, H. Philip, M. Jaronen, J. Almeida, J. E. Kenison, S. Darko, C. C. Chao, G. Yaari, Y. Louzoun, L. Carmel, D. C. Douek, S. Efroni, F. J. Quintana, System-wide analysis of the T cell response. *Cell Rep.* **14**, 2733–2744 (2016).
18. J. A. Weinstein, N. Jiang, R. A. White III, D. S. Fisher, S. R. Quake, High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324**, 807–810 (2009).
19. N. Jiang, J. A. Weinstein, L. Penland, R. A. White III, D. S. Fisher, S. R. Quake, Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5348–5353 (2011).
20. F. Rubelt, C. R. Bolen, H. M. McGuire, J. A. Vander Heiden, D. Gadala-Maria, M. Levin, G. M. Euskirchen, M. R. Mamedov, G. E. Swan, C. L. Dekker, L. G. Cowell, S. H. Kleinstein, M. M. Davis, Individual heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-experiemced cells. *Nat. Commun.* **7**, 11112 (2016).
21. M. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (Dover, 2009).
22. M. Kottelat, R. Britz, T. H. Hui, K.-E. Witte, Paedocypris, a new genus of Southeast Asian cyprinid fish with a remarkable sexual dimorphism, comprises the world's smallest vertebrate. *Proc. Biol. Sci.* **273**, 895–899 (2006).
23. S. Liu, T. H. Hui, S. L. Tan, Y. Hong, Chromosome evolution and genome miniaturization in minifish. *PLOS ONE* **7**, e37305 (2012).
24. M. Malstrom, R. Britz, M. Matschiner, O. K. Torresen, R. K. Hadiaty, N. Yaakob, H. H. Tan, K. S. Jakobsen, W. Salzburger, L. Rüber, The most developmentally truncated fishes show extensive hox gene loss and minaturized genomes. *Genome Biol. Evol.* **10**, 1088–1103 (2018).
25. J. S. Swann, A. Weyn, D. Nagakubo, C. C. Bleul, A. Toyoda, C. Happe, N. Netuschil, I. Hess, A. Haas-Assenbaum, Y. Taniguchi, M. Schorpp, T. Boehm, Conversion of the thymus into a bipotent lymphoid organ by replacement of FOXN1 with its paralog, FOXN4. *Cell Rep.* **8**, 1184–1197 (2014).
26. T. Boehm, I. Hess, J. B. Swann, Evolution of lymphoid tissues. *Trends Immunol.* **33**, 315–321 (2012).

27. A. Brendolan, M. M. Rosado, R. Carsetti, L. Selleri, T. N. Dear, Development and function of the mammalian spleen. *Bioessays* **29**, 166–177 (2007).
28. L. Xie, Y. Tao, R. Wu, Q. Ye, H. Xu, Y. Li, Congenital asplenia due to a *tlx1* mutation reduces resistance to *Aeromonas hydrophila* infection in zebrafish. *Fish Shellfis Immunol.* **95**, 538–545 (2019).
29. Y. Chi, Z. Huang, Q. Chen, X. Xiong, K. Chen, J. Xu, Y. Zhang, W. Zhang, Loss of *runx1* function results in B cell immunodeficiency but not T cell in adult zebrafish. *Open Biol.* **8**, 180043 (2018).
30. N. Danilova, J. Bussmann, K. Jekosch, L. A. Steiner, The immunoglobulin heavy-chain locus in zebrafish: Identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat. Immunol.* **6**, 295–302 (2005).
31. S. L. Seelye, P. L. Chen, T. C. Deiss, M. F. Criscitiello, Genomic organization of the zebrafish (*Danio rerio*) T cell receptor alpha/delta locus and analysis of expressed products. *Immunogenetics* **68**, 365–379 (2016).
32. N. D. Meeker, A. C. Smith, J. K. Frazer, D. F. Bradley, L. A. Rudner, C. Love, N. S. Trede, Characterization of the zebrafish T cell receptor β locus. *Immunogenetics* **62**, 23–29 (2010).
33. F. Wan, C.-B. Hu, J. X. Ma, K. Gao, L.-X. Xiang, J.-Z. Shao, Characterization of γδ T cells from zebrafish provides insights into their important role in adaptive humoral immunity. *Front. Immunol.* **7**, 675 (2016).
34. H. Li, C. Ye, G. Ji, X. Wu, Z. Xiang, Y. Li, Y. Cao, X. Liu, D. C. Douek, D. A. Price, J. Han, Recombinatorial biases and convergent recombination determine interindividual TCRβ sharing in murine thymocytes. *J. Immunol.* **189**, 2404–2413 (2012).
35. N. L. La Gruta, S. Gras, S. R. Daley, P. G. Thomas, J. Rossjohn, Understanding the drivers of MHC restriction of T cell receptors. *Nat. Rev. Immunol.* **18**, 467–478 (2018).
36. H. Tanno, T. M. Gould, J. R. McDaniel, W. Cao, Y. Tanno, R. E. Durrett, D. Park, S. J. Cate, W. H. Hildebrand, C. L. Dekker, L. Tian, C. M. Weyand, G. Georgiou, J. J. Goronzy, Determinants governing T cell receptor α/β-chain pairing in repertoire formation of identical twins. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 532–540 (2020).
37. P. J. Brennan, M. Brigl, M. B. Brenner, Invariant natural killer T cells: An innate activation scheme linked to diverse effector functions. *Nat. Rev. Immunol.* **13**, 101–117 (2013).
38. D. Dong, L. Zheng, J. Lin, B. Zhang, Y. Zhu, N. Li, S. Xie, Y. Wang, N. Gao, Z. Huang, Structural basis of assembly of the human T cell receptor–CD3 complex. *Nature*, 546–552 (2019).
39. J. A. Yoder, T. M. Orcutt, D. Traver, G. W. Litman, Structural characteristics of zebrafish orthologs of adaptor molecules that associate with transmembrane immune receptors. *Gene* **401**, 154–164 (2007).
40. G. Gaud, R. Lesourne, P. E. Love, Regulatory mechanisms in T cell receptor signalling. *Nat. Rev. Immunol.* **18**, 485–497 (2018).
41. V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* **10**, 707–710 (1966).
42. A. Madi, A. Poran, E. Shifrut, S. Reich-Zeliger, E. Greenstein, I. Zaretsky, T. Arnon, F. Van Laethem, A. Singer, J. Lu, P. D. Sun, I. R. Cohen, N. Friedman, T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *eLife* **6**, e22057 (2017).
43. J. Glanville, H. Huang, A. Nau, O. Hatton, L. E. Wagar, F. Rubelt, X. Ji, A. Han, S. M. Krams, C. Pettus, N. Haas, C. S. Lindestam Arlehamn, A. Sette, S. D. Boyd, T. J. Scriba, O. M. Martinez, M. M. Davis, Identifying specificity groups in the T cell repertoire. *Nature* **547**, 94–98 (2017).
44. H. Huang, C. Wang, F. Rubelt, T. J. Scriba, M. M. Davis, Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide screening. *Nat. Biotechnol.* **38**, 1194–1202 (2020).
45. P. Dash, A. J. Fiore-Gartland, T. Hertz, G. C. Wang, S. Sharma, A. Souquette, J. C. Crawford, E. B. Clemens, T. H. O. Nguyen, K. Kedzierska, N. L. La Gruta, P. Bradley, P. G. Thomas, Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **647**, 89–93 (2017).
46. J. Robert, E.-S. Edholm, A prominent role for invariant T cells in the amphibian *Xenopus laevis* tadpoles. *Immunogenetics* **66**, 513–523 (2014).
47. E. N. Rittmeyer, A. Allison, M. C. Gründler, D. K. Thompson, C. C. Austin, Ecological guild evolution and the discovery of the world's smallest vertebrate. *PLOS ONE* **7**, e29797 (2012).
48. H. S. Robins, S. K. Srivastava, P. V. Campregher, C. J. Turtle, J. Andriesen, S. R. Riddell, C. S. Carlson, E. H. Warren, Overlap and effective size of the human $CD8^+$ T cell receptor repertoire. *Sci. Transl. Med.* **2**, 47ra64 (2010).
49. N. M. Provine, P. Klenerman, MAIT cells in health and disease. *Annu. Rev. Immunol.* **38**, 203–228 (2020).
50. C. Song, S. Havlin, H. A. Makse, Self-similarity of complex networks. *Nature* **433**, 392–395 (2005).
51. R. Albert, H. Jeong, A. L. Barabasi, Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
52. M. A. Turchaninova, O. V. Britanova, D. A. Bolotin, M. Shugay, E. V. Putintseva, D. B. Staroverov, G. Sharonov, D. Shcherbo, I. V. Zvyagin, I. Z. Mamedov, C. Linnemann, T. N. Schumacher, D. M. Chudakov, Pairing of T-cell receptor chains via emulsion PCR. *Eur. J. Immunol.* **43**, 2507–2515 (2013).
53. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal Complex Systems*, 1695 (2006).
54. D. M. Langenau, A. A. Ferrando, D. Traver, J. L. Kutok, J.-P. Hezel, J. P. Kanki, L. I. Zon, A. T. Look, N. S. Trede, In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7369–7374 (2004).
55. I. Hess, T. Boehm, Intravital imaging of thymopoiesis reveals dynamic lympho-epithelial interactions. *Immunity* **36**, 298–309 (2012).
56. G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
57. E. S. Lander, M. S. Waterman, Genomic mapping by fingerprinting random clones: A mathematical analysis. *Genomics* **2**, 231–239 (1988).
58. G. W. Vurture, F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, M. C. Schatz, GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
59. N. I. Weisenfeld, S. Yin, T. Sharpe, B. Lau, R. Hegarty, L. Holmes, B. Sogoloff, D. Tabbaa, L. Williams, C. Russ, C. Nusbaum, E. S. Lander, I. MacCallum, D. B. Jaffe, Comprehensive variation discovery in single human genomes. *Nat. Genet.* **46**, 1350–1355 (2014).
60. Z. Zhang, S. Schwartz, L. Wagner, W. Miller, A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
61. X. Huang, A. Madan, CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868–877 (1999).
62. N. H. Putnam, B. L. O'Connell, J. C. Stites, B. J. Rice, M. Blanchette, R. Calef, C. J. Troll, A. Fields, P. D. Hartley, C. W. Sugnet, D. Haussler, D. S. Rokhsar, R. E. Green, Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
63. A. V. Zimin, G. Marcais, D. Puiu, M. Roberts, S. L. Salzberg, J. A. Yorke, The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013).
64. A. C. English, S. Richards, Y. Han, M. Wang, V. Vee, J. Qu, X. Qin, D. M. Muzny, J. G. Reid, K. C. Worley, R. A. Gibbs, Mind the gap: Upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLOS ONE* **7**, e47768 (2012).
65. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
67. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
68. A. Yates, W. Akanni, M. R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C. Cummins, P. Clapham, S. Fitzgerald, L. Gil, C. G. Giron, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, S. Keenan, I. Lavidas, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, M. Nuhn, A. Parker, M. Patricio, M. Pignatelli, M. Rahtz, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, E. Birney, J. Harrow, M. Muffato, E. Perry, M. Ruffier, G. Spudich, S. J. Trevanion, F. Cunningham, B. L. Aken, D. R. Zerbino, P. Flicek, Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).
69. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
70. A. Smit, R. Hubley, *RepeatModeler Open-1.0* (2008–2015); www.repeatmasker.org.
71. G. Abrusán, N. Grundmann, L. DeMester, W. Makalowski, TEclass—A tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329–1330 (2009).
72. W. Bao, K. K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
73. A. Smit, R. Hubley, *RepeatMasker Open-4.0* (2013–2015); www.repeatmasker.org.
74. B. L. Cantarel, I. Korf, S. M. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sanchez Alvarado, M. Yandell, MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
75. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19** (suppl. 2), ii215–ii225 (2003).
76. U. Grimholt, K. Tsukamoto, T. Azuma, J. Leong, B. F. Koop, J. M. Dijkstra, A comprehensive analysis of teleost MHC class I sequences. *BMC Evol. Biol.* **15**, 32 (2015).

Acknowledgments: We thank J. Rauh for advice and H. Pircher for comments on the manuscript. **Funding:** Work in T.B.'s laboratory was funded by the Max Planck Society. Work in B.V.'s laboratory was supported by the Biomedical Research Council of A*STAR, Singapore.
Author contributions: T.B. and B.V. conceived and designed the research. H.H.T. provided fish samples. B.-H.T. prepared genomic DNA, RNA-seq, and cDNA libraries. P.S. and N.E.P.

assembled the genomes. N.E.P. and A.P. assembled RNA-seq transcripts. P.S. performed the genome annotation. V.R., P.S., and B.V. carried out the genome analysis. O.B.G. performed antigen receptor and immunogenome analyses. All authors analyzed and interpreted the data. T.B. wrote the paper with input from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The datasets generated in the current study were deposited in the following databases. The whole-genome sequences of the male and female minifish have been deposited at DDBJ/ENA/GenBank under accessions JABUMV000000000 and JABUMW000000000, respectively. RNA-seq reads for the male and female minifish have been deposited in the NCBI Sequence Read Archive under accessions SRR11930059 and SRR11930060, respectively. The sequencing reads for antigen receptor assemblies are available at www.ncbi.nlm.nih.gov/sra/PRJNA612865. The custom code underlying the analysis is available at GitHub (https://github.com/obgiorgetti/minifish). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 14 July 2020
Accepted 4 November 2020
Published 1 January 2021
10.1126/sciadv.abd8180

Citation: O. B. Giorgetti, P. Shingate, C. P. O'Meara, V. Ravi, N. E. Pillai, B.-H. Tay, A. Prasad, N. Iwanami, H. H. Tan, M. Schorpp, B. Venkatesh, T. Boehm, Antigen receptor repertoires of one of the smallest known vertebrates. *Sci. Adv.* **7**, eabd8180 (2021).

Science Advances

Antigen receptor repertoires of one of the smallest known vertebrates

Orlando B. Giorgetti, Prashant Shingate, Connor P. O'Meara, Vydianathan Ravi, Nisha E. Pillai, Boon-Hui Tay, Aravind Prasad, Norimasa Iwanami, Heok Hui Tan, Michael Schorpp, Byrappa Venkatesh and Thomas Boehm

Sci Adv **7** (1), eabd8180.
DOI: 10.1126/sciadv.abd8180