# Antibody repertoire sequencing reveals systemic and mucosal immunosenescence in the short-lived turquoise killifish

William J Bradshaw[1,2], Michael Poeschla[1], Aleksandra Placzek[1], Dario Riccardo Valenzano[1,2*]

[1] Max Planck Institute for Biology of Aging, Joseph-Stelzmann-Str. 296, 50937 Cologne, Germany

[2] CECAD Research Center, University of Cologne, Joseph-Stelzmann-Straße 26, 50931 Cologne, Germany

* Correspondence to: dvalenzano@age.mpg.de

**Keywords:** Adaptive immunity, B cells, Immunoglobulin repertoire, aging, killifish, immunosenescence

---

## Abstract

Aging individuals exhibit a pervasive decline in adaptive immune function, with important implications for health and lifespan. Previous studies have found a pervasive loss of immune-repertoire diversity in human peripheral blood; however, little is known about repertoire aging in other immune compartments, or in species other than humans. Here, we perform the first study of immune-repertoire aging in an emerging model of vertebrate aging, the African turquoise killifish (*Nothobranchius furzeri*). Despite their extremely short lifespans, these killifish exhibit complex and individualised heavy-chain repertoires, with a generative process capable of producing millions of productive receptor sequences. Whole-body killifish repertoires decline rapidly in within-individual diversity with age, while between-individual variability increases. Large, expanded B-cell clones exhibit far greater diversity loss with age than small clones, suggesting an important difference in the age-sensitivity of different B-cell populations. Compared to the whole body, the immune repertoires of isolated intestinal samples exhibit much more dramatic age-related phenotypes, apparently due to an elevated prevalence of age-sensitive expanded clones. Our results highlight the importance of organ-specific dynamics in adaptive immunosenescence.

---

## Introduction

The adaptive immune system undergoes a severe and systemic decline in functionality with age, resulting in higher susceptibility to a wide range of infections and decreased efficacy of vaccination in elderly individuals (1–3). In the humoral immune system, aging is accompanied by a decline in naïve B-cell output from the

primary lymphoid organs; impaired production of specific antibodies in response to antigenic challenge; and a decline in antibody quality (1, 2, 4, 5), as well as impairments in the establishment of novel immune memory (5). These changes are major contributors to a generalized immunosenescent phenotype that significantly impairs health and quality of life in the elderly.

The efficacy of the humoral immune system rests on its ability to generate a vast array of different antibody sequences, with a correspondingly vast range of antigen specificities, and to progressively adjust the composition of this antibody population in response to antigen exposure (6–9). Sampling the resulting repertoire of antibody sequences in an individual using high-throughput sequencing can yield important insights into the diversity, clonal composition, and history of antibody-mediated immunity in that organism - as well as the effect of age, antigen exposure and other factors on the diversity and functionality of the adaptive immune system (10–12).

In humans, antibody-repertoire sequencing has uncovered a number of important age-related changes, including reduced numbers of clones and unique sequences, increased baseline mutation, increased clonal expansions, impaired B-cell selection, and a shift towards the memory compartment (11, 13, 14). The responsiveness of the peripheral repertoire to vaccination is also impaired during aging (11, 14). While within-individual repertoire diversity declines with age, between-individual variability increases, with repertoires from older individuals differing more from one another than those from young individuals (11).

Previous work in humans, however, has been limited by small sample sizes, a lack of temporal resolution, or a restriction to peripheral blood samples, which are known to systematically underrepresent the majority of B cells resident in other organs and tissues (15, 16). Very little is known about how repertoire aging differs between distinct organs; in particular, almost nothing is known about how aging affects antibody repertoires at mucosal surfaces, which represent a crucial interface between the body and its microbial environment (17, 18). Even less is known about how aging might affect the antibody repertoires of vertebrates other than mice and humans.

In this study, we introduce the African turquoise killifish (*Nothobranchius furzeri*) (19, 20) as a model for adaptive immunosenescence. As the shortest-lived vertebrate that can be bred in captivity, the turquoise killifish is an emerging model of vertebrate aging (19, 21, 22). Here, we perform the first immune-repertoire sequencing experiments in this species, demonstrating that adult killifish express diverse and individualized heavy-chain repertoires that undergo rapid loss of diversity with age. This loss primarily affects the composition of expanded clones, with small naïve clones exhibiting much smaller age-related changes. By sequencing the repertoires of isolated intestinal samples, we further find that the killifish intestinal antibody repertoire exhibits much more dramatic age-dependent diversity loss than the body as a whole, possibly due to a much higher prevalence of age-sensitive expanded clones. Taken together, our results highlight the importance of organ-specific dynamics

when studying immune repertoire aging, and establish the turquoise killifish as a powerful model for studying adaptive immune senescence.

# Results

## Establishing immunoglobulin sequencing in the turquoise killifish

To investigate the effect of age on the B-cell receptor repertoire diversity and composition in turquoise killifish, we implemented an RNA-based repertoire-sequencing protocol based on the published protocol of Turchaninova et al. (23), using template switching (24) to add unique molecular identifiers (UMIs) to each RNA transcript of the immunoglobulin heavy chain (Fig. 1) (20) to correct for errors and biases in abundance arising during PCR and Illumina sequencing (25). To test the validity and replicability of results obtained using this protocol, we performed three replicate library preps on whole-body total RNA samples from four adult (eight-week-old) adult male turquoise killifish from the short-lived GRZ strain (Fig. S1). Independent repertoires from the same individual show a high degree of similarity in their clonal composition, with an average inter-replicate correlation in clone size of $r = 0.89$ (Fig. S2). Inter-repertoire divergences computed with the published repertoire dissimilarity index (RDI) (26) metric consistently identify replicates from the same individual as much more similar than repertoires from different individuals (Fig. S3), demonstrating that this protocol is capable of accurately and reproducibly reconstructing the expressed heavy-chain repertoires of individual killifish.

## Aging in whole-body killifish repertoires

To investigate the effect of age on the structure and diversity of killifish antibody repertoires, we performed whole-body immunoglobulin sequencing on 32 adult male turquoise killifish from the short-lived GRZ strain (19) at four different ages from early adulthood to late life (Fig. 2A and Table S2 and Table S4). We then computed Hill diversity spectra, which provide a holistic overview of the diversity structure of a population (12, 27–29). Each curve in a Hill spectrum gives the diversity of a population as measured across a range of diversity orders; at low orders, species of different sizes are weighted more equally in the diversity calculation, while at higher orders less-abundant species are progressively downweighted relative to more-abundant species (SI Note 1). We separately computed spectra for the clonal repertoire (in which repertoire sequences are partitioned by clonal identity) and VJ repertoire (in which sequences are partitioned by V/J identity) of each individual (Fig. S4) and used these individual spectra to compute averaged alpha-diversity spectra for each age group (SI Note

2). At each diversity order, we tested for a significant age effect on repertoire diversity using a permutation test on the Kruskal-Wallis *H* statistic (Materials and Methods, SI Note 7).

The clonal diversity of the whole-body killifish antibody repertoire exhibits a significant decline with age ($p < 0.05$) at high diversity orders (Fig. 2B, Fig. S6, Fig. S7), indicating a significant and extremely rapid age-related decline in the diversity of the largest B-cell clones. In contrast, lower-order clonal diversity exhibits no significant change with age, suggesting that the *overall* composition of the whole-body repertoire remains relatively unchanged. Since the B-cell clonal repertoire is overwhelmingly dominated by small, predominantly naïve clones (Fig. S5A), low-order clonal diversity measurements are primarily driven by changes in the diversity of small clones. As such, these results indicate that the composition of small clones in the killifish antibody repertoire is much less sensitive to the effects of aging than that of large, expanded clones.

In contrast with the rapid age-related declines observed in high-order clonal diversity, the VJ diversity of the killifish antibody repertoire exhibits no significant age-related change at any diversity order (Fig. 2C). Examining the clone-size distribution of each V/J combination (Fig. S5B) reveals that even the largest V/J combinations in each age group are overwhelmingly dominated by small clones, suggesting that the observed lack of an age effect on VJ diversity is due to the observed age-insensitivity of small clones (Fig. 2B). To test this hypothesis, we filtered the repertoire dataset to separate sequences from small and large clones and computed VJ diversity repertoires for each subset (Fig. 2D-E). While both small and large clones in isolation show a significant age effect on VJ diversity, the relative reduction in VJ alpha-diversity with age is dramatically stronger for large clones, an effect observed across all diversity orders (Fig. 2F). As suggested by the clonal-diversity results, therefore, the repertoire diversity of large (expanded) clones in the killifish whole-body repertoire appears to be far more age-sensitive than that of small (predominantly naïve) clones.

In addition to the average within-individual diversity of a population (alpha diversity), the between-individual variation in composition (beta diversity) is also of interest. Previous studies of human peripheral-blood repertoires have suggested a decrease in alpha-diversity but an *increase* in beta diversity with age (11, 30). In our dataset, VJ beta-diversity spectra (SI Note 2 & SI Note 3) indicated a large age-related increase in beta diversity across a wide range of diversity orders (Fig. 2G), indicating a similar pattern of progressive individualization in repertoire composition with age. Concordantly, older killifish also exhibit significantly greater pairwise Repertoire Dissimilarity Index (RDI) distances (26), indicating progressive divergence in repertoire composition (Fig. 2H-I). As in humans, therefore, younger killifish exhibit antibody repertoires that are significantly more similar to one another, which then become increasingly distinct and individualized as the cohort increases in age.

## The killifish generative repertoire

The naïve sequence diversity of the antibody heavy-chain repertoire depends on the molecular processes underlying the generation of novel sequences in developing B cells: random selection of V, D and J segments during VDJ recombination; deletions and palindromic (P-) insertions at the ends of conjoined segments; and nonpalindromic (N-) insertions between segments (6, 31). Each of these contributes diversity to the overall generative process, increasing the variety of sequences that can be generated. Excluding nonfunctional sequences, the human generative process has an estimated Shannon entropy of roughly 70 bits, corresponding to a first-order Hill diversity of roughly $10^{21}$ possible unique sequences (9). However, little is known about how this generative diversity varies across species, or how it changes during aging.

To gain insight into these generative processes in the turquoise killifish, we used IGoR (32) to infer models of sequence generation from killifish repertoire data. In training these models, we restricted the dataset to nonfunctional naïve sequences, in order to avoid distortions introduced by positive and negative selection in the primary lymphoid organs (9, 32). As is often the case with RNA data (33), the number of naïve nonfunctional sequences available per individual was frequently low; hence, to better capture low-probability events in the generative process, we inferred models from pooled data from multiple individuals in the same age group. As the parameters of the generative model are typically very similar across conspecific individuals (Fig. S8 & Fig. S9), especially in an inbred line, pooling data like this is a useful way to infer more robust generative models using IGoR (33).

To model the generative process in its baseline state, we first inferred a model of the killifish generative repertoire from the four 8-week-old adult male individuals used in the pilot study (Fig. S1). Using this model, we estimated a total raw entropy for the killifish generative repertoire of roughly 33 bits (Fig. 3A). Of these 33 bits, roughly 8 arise from variability in VDJ segment choice, 12 from variability in the number and composition of junctional N-insertions, and 11 from P-insertions and deletions. Accounting for convergent production of identical sequences via different recombination events, and for events that give rise to nonfunctional nucleotide sequences (e.g. due to frame shift) reduces this initial raw estimate by 10 bits.

Before initial selection in the primary lymphoid organs, therefore, the killifish generative process has an estimated Shannon entropy of roughly 23 bits (Fig. 3A), corresponding to a first-order Hill diversity of roughly $10^{7}$ possible unique sequences. While, as in humans, only a small fraction of potential diversity will actually be

generated in any single individual, this nevertheless represents a highly complex and sophisticated system capable of generating highly individualised *IGH* repertoires.

While impressive, the potential generative diversity of the killifish repertoire is nevertheless vastly lower than in humans, with a difference in productive generative entropy of almost 50 bits (9). While all components of the generative process exhibit lower entropies in killifish than in humans, by far the greatest difference lies in the junctional N-insertions, which contribute almost 40 bits more to the generative entropy of the human repertoire than that of killifish. The difference in the productive generative entropy between killifish and human arises from the distributions of N-insertions inferred from killifish and human data: in humans, these distributions peak at around 5 nt per junction and often yield insertions of 10-20 nt (9), while in killifish the insertion distribution peaks at 0 nt per junction, and sequences with more than 5 nt of insertions at either junction are very rare (Fig. 3B and Fig. S8). Since N-insertions are the dominant source of sequence diversity in human repertoires, the large reduction in N-insertions in killifish relative to humans unsurprisingly results in a much lower overall generative diversity for the killifish adaptive immune system.

The relative lack of change in the small-clone antibody repertoire in older turquoise killifish (Fig. 2B-F) suggested to us that the diversity of the generative process in the primary lymphoid organs might remain relatively intact throughout the killifish lifespan. To test this hypothesis, we trained separate IGoR models for each individual in the 32-fish aging cohort (Fig. 2A, Fig. S10 and Fig. S11) and tested for an effect of age on the generative diversity inferred for each individual. As expected, no age effect was found in either total generative diversity (Fig. 3D) or the contributions of different diversification processes (Fig. S12). It therefore appears that, while some aspects of the killifish antibody repertoire certainly decline with age, the entropy of the generative process is not among them.

## Effect of age and microbiota transfer on killifish intestinal repertoires

The populations of B-lymphocytes associated with mucosal epithelia play a crucial role in defending the body from pathogenic threats (18), as well as in regulating the composition of resident microbial populations (17). Despite the importance of these distinctive B-cell compartments, relatively little is known about the structure of their antibody repertoires (18), and still less about how these repertoires change with age.

As the site of the greatest microbiota diversity, the intestine is of particular interest as an important and distinctive immune environment. Previous work on the killifish gut microbiota (34) has shown that it declines in alpha diversity and increases in beta diversity with age, patterns that mirror the changes seen in the whole-body

composition of the killifish repertoire (Fig. 2). Transfer of intestinal material from young to middle-aged fish has also been shown to extend lifespan (34). Given these findings, and the intimate relationship between intestinal lymphocytes and gut bacteria (17), we decided to investigate the effect of aging and microbiota transfer on the immune repertoires of gut-resident B-cell populations.

Using intestinal total RNA isolated by Smith *et al.* (34), we sequenced the intestinal *IGH* repertoires of eighteen male GRZ-strain individuals, including four untreated 6-week-old individuals and fourteen 16-week-old individuals from various microbiota-transfer treatment groups (Fig. 4, Table S3 & Table S5), and investigated the effect of age and treatment condition on repertoire diversity in the killifish intestine.

Contrary to our expectations, neither alpha nor beta diversity are significantly affected by microbiota transfer, with no significant difference in clonal diversity, VJ diversity, or RDI distance measures (Fig. S13, Fig. S15). In sharp contrast to the whole-body data, however, there is a strong and significant decline in both clonal and VJ alpha diversity with age across all diversity orders (Fig. 4B-C, Fig. S14, Fig. S15), without partitioning by clone size. This age-related decline in alpha diversity is consistently far more dramatic than that observed in the whole-body samples at any diversity order. The B cells of the killifish intestine, therefore, seem to exhibit a much stronger age-dependent decline in repertoire diversity than is seen in the killifish body overall.

While results from beta-diversity spectra show large increases in beta diversity with age at some diversity orders but not at others (Fig. 4D), the median pairwise RDI distance between individual gut repertoires increased substantially and significantly with age (Fig. 4E and Fig. S16), suggesting that, as in the whole body, killifish intestinal repertoires become increasingly distinct and individualized as they age.

One potential explanation for the apparent difference in how age affects whole-body and intestinal repertoires is that the rapid changes observed in the gut are a consequence of the constant strong antigen exposure experienced by intestinal B cells as a result of their interaction with the gut microbiota. This exposure could drive high levels of antigen-dependent clonal expansion, resulting in a loss in repertoire diversity (35). Another explanation, not mutually exclusive with the first, is that the difference is a consequence of the different clone-size distribution in the gut relative to the whole body. Unlike the whole-body repertoire, the gut does not include the primary lymphoid organs, and so would be expected to be far less dominated by small, naïve clones. Since the population of large clones appears to be more prone to reductions in diversity with age than that of small clones (Fig. 2), the stronger overall age-related diversity loss found in the gut repertoire could be a consequence of this greater relative prevalence of age-sensitive large clones.

Rarefaction analysis of clonal counts in whole-body and intestinal repertoires showed that the latter indeed contain far fewer small clones, resulting in a much higher proportion of large clones (Fig. 5A). If this difference in clonal composition, rather than some functional difference between intestinal and other B cells, is primarily responsible for the apparent difference in aging phenotypes between whole-body and intestinal repertoires, we would expect to find a faster rate of clonal diversity loss during aging in intestinal repertoires at low diversity orders (which are dominated by small clones in whole-body samples), but not at high orders (which are dominated by large clones in both sample types). Similarly, we would expect to find faster loss in intestinal samples of V/J diversity considered over all clones, but not when the V/J diversity calculation is restricted to large clones alone.

To test these hypotheses, we normalized the diversity measurements from each dataset by the mean diversity of the youngest group in that dataset, then fit generalized linear models for different diversity orders and methods of measuring diversity (Fig. 5B and Fig. S17), testing for a significant interaction between sample type (i.e. gut vs whole body) and the effect of age on repertoire diversity. Gut samples exhibit significantly higher rates of age-dependent diversity loss under low-order clonal-diversity or total VJ-diversity measures, i.e. those metrics for which clones of all sizes were included in the diversity calculation. Conversely, there is no significant difference in rate of diversity loss between sample types for higher-order clonal-diversity measures, nor for V/J-diversity measures restricted to only large clones, indicating that large clones undergo similar rates of age-dependent diversity loss in both sample types. These results closely match the predictions of the clonal-composition model: large clones in both gut and whole-body samples exhibit similarly strong aging phenotypes, but the higher proportions of large clones in gut samples result in these strong phenotypes manifesting more strongly in the behavior of the repertoire as a whole. It therefore appears that, as in whole-body samples, age-dependent diversity loss in killifish intestinal repertoires is primarily a phenomenon of mature, expanded clones.

# Discussion

The turquoise killifish is the shortest-lived vertebrate that can be bred in captivity (36, 37), with a median lifespan in the short-lived GRZ strain of just three to four months. Despite this, our findings show that the life of a turquoise killifish provides ample time both to develop a complex, diverse and individualized IgM heavy-chain repertoire (Fig. 3A, Fig. S2, Fig. S3), and for that repertoire to decline significantly in diversity with age. This decline appears to be primarily caused by a loss of clonal diversity in the minority of highly expanded clones in

the repertoire. At the same time, the composition of the repertoire becomes progressively more different in older killifish, resulting in a rapid age-dependent increase in beta diversity.

The relative lack of age-related change seen in small clones does not necessarily show that the process of generating new naïve B cells is unchanged in older killifish. An important alternative explanation, which we were not able to test effectively with our dataset, relates to body size: unlike humans and mice, turquoise killifish grow continuously with age, and it is possible that an age-related decline in the rate of B-cell output by the primary lymphoid organs is offset by a continuous increase in the size of these organs with age. Nevertheless, our results suggest that the bulk of the change in the IgM repertoire of aging killifish arises from changes in the size and composition of expanded antigen-experienced clones.

While the rate of naïve B-cell generation from the primary lymphoid organs may or may not change with age, the entropy of the process underlying heavy-chain sequence generation appears to stay relatively unchanged. This process, while capable of producing a highly complex repertoire, is nevertheless vastly less diverse than the equivalent process in humans (9), a difference arising primarily from a dramatic difference in the two species' respective N-insertion distributions. The cause of this difference is not known, nor whether it is specific to turquoise killifish or a more general feature of teleost antibody repertoires; however, it seems likely to involve the structure, expression or regulation of the terminal deoxynucleotidyltransferase enzyme (TdT) responsible for the N-insertion process (6, 31). More detailed investigation of the structure and function of this enzyme in turquoise killifish and related species may reveal important insights into the mechanism and regulation of the N-insertion process during primary antibody diversification.

As early as 2009, Caruso et al. (35) commented that the mucosal adaptive immune system might exhibit particularly strong loss of diversity with age, as a result of especially frequent antigenic challenges. In what is, to our knowledge, the first published test of this hypothesis, we sequenced the heavy-chain repertoires of isolated gut samples from young and old adult turquoise killifish, discovering that the intestinal repertoire does indeed undergo particularly dramatic loss of diversity with age. However, this difference between the gut and whole-body repertoires appears to result from a difference in clonal composition, rather than in the behaviour of any particular clonal subset: those expanded clones that are present appear to lose diversity at similar rates in both sample types, further supporting the conclusion that changes in repertoire diversity with age are primarily driven by expanded, antigen-experienced B-cell clones rather than the naïve compartment. Nevertheless, it would be important to investigate other mucosal immune organs – such as the gills, nasal passages, and skin (18, 38) – to undestand if these aging phenotypes represent a general pattern or are specific to the intestinal mucosa.

Apart from the nervous system itself, no other system in the vertebrate body exhibits such complex learning and memory behaviour as the adaptive immune system. The age-related decline in the functionality of this system is a major cause of mortality and morbidity in the elderly. Our results firmly establish the value of the turquoise killifish as a model for investigating this important and complex process, and demonstrate the importance of studying immune aging in compartments other than peripheral blood. Future experiments in this system have the potential to greatly expand our knowledge of the mechanisms, spatial distribution and temporal progression of immune repertoire aging, with potentially vital implications for the future treatment of immunosenescent phenotypes in aging populations.

# Materials and Methods

## Fish husbandry and sample preparation

Male turquoise killifish (*Nothobranchius furzeri*, GRZ-AD strain) from a single hatching cohort were raised under standard husbandry conditions (39) and housed from four weeks post-hatching in individual 2.8L tanks connected to a water-recirculation system. Fish received 12 hours of light per day on a regular light/dark cycle, and were fed bloodworm larvae and brine shrimp nauplii twice a day during the week and once a day during the weekend (34, 39).

After being sacrificed in 1.5 g/L Tricaine solution at room-temperature tank water (40), fish (Table S4) were flash-frozen in liquid nitrogen and ground to a homogenous powder with a pestle in a liquid-nitrogen-filled mortar. The powder was mixed thoroughly and stored at -80˚C prior to RNA isolation. Intestinal total RNA for the gut experiments was provided by Smith *et al.* (34).

## Immunoglobulin sequencing

Total RNA from whole-body killifish samples was isolated using QIAzol lysis reagent (QIAGEN, 1 ml of reagent per 0.1 g of homogenised tissue) and isopropanol precipitation; gut RNA from microbiota-transfer experiments (34) was already prepared and available. Quantification of RNA samples was performed with the Qubit 2.0 fluorometer (Thermo Fisher), while quality control and integrity measurement was performed using the TapeStation 4200 (Agilent).

Reverse transcription and template switching for library preparation was performed on total RNA samples using SMARTScribe Reverse Transcriptase, in line with the protocol specified in Turchaninova *et al.* (23) (SI Note 4). The reaction product was purified using SeraSure SPRI beads (SI Note 5), then underwent three successive rounds of PCR, each of which was followed by a further round of bead purification. The first of these PCR reactions added a second strand to the reverse-transcribed cDNA and amplified the resulting DNA molecules; the second added partial Illumina sequencing adapters and further amplified the library, and the third added complete Illumina adapters, including i5 and i7 indices.

The concentration of each library was then quantified and the libraries were pooled in equimolar ratio, concentrated using SeraSure beads, and size-selected with the BluePippin (Sage Science) to obtain a purified amplicon band. Finally, following a final round of quality control, the pooled and size-selected libraries were sequenced on an Illumina MiSeq System (MiSeq Reagent Kit v3, 2×300 bp reads, 30% PhiX spike-in), either at the Cologne Center for Genomics (whole-body libraries) or with Admera Health (intestinal libraries).

## Date processing and analysis

Pre-processing of raw sequencing data (including quality filtering, consensus-read generation, and clonotyping) was performed using the pRESTO (41) and Change-O (42) suites of command-line tools (SI Note 6, Fig. S18). Downstream analysis of processed data, including diversity-spectrum inference (SI Note 7), RDI computation, GLM fitting and rarefaction, was performed in R, as was figure generation and all statistical tests. Generative model inference was performed using IGoR (32). Snakemake (43) was used to design and run data-processing pipelines.

## Data and code availability

Raw data used in these analyses will be made available on request and in a future version of this preprint.

Processed data and code used in these analysis will be made available at https://github.com/willbradshaw/killifish-igseq

# Acknowledgments

We would like to thank Alexander Dilthey for detailed input on statistical approaches; Aleksandra Walczak, Thierry Mora, John Beausang and Susana Magadan for their generous advice on designing the experimental protocols and analysis; and Jason Vander Heiden and Quentin Marcou for their technical assistance with the

pRESTO / Change-O and IGoR software. We would further like to thank all the members of the Valenzano lab for continuing support and feedback on the project. This project was funded by the Max Planck Society, the Max Planck Institute for Biology of Aging, the CECAD Research Center in Cologne and the DFG Collaborative Research Center 1310 grant.

# References

1. A. Ademokun, Y.-C. Wu, D. Dunn-Walters, The ageing B cell population: Composition and function. *Biogerontology* **11**, 125–137 (2010).

2. I. Kogut, J. L. Scholz, M. P. Cancro, J. C. Cambier, B cell maintenance and function in aging. *Seminars in Immunology* **24**, 342–349 (2012).

3. D. K. Dunn-Walters, A. A. Ademokun, B cell repertoire and ageing. *Curr. Opin. Immunol.* **22**, 514–520 (2010).

4. S. Sasaki, *et al.*, Limited efficacy of inactivated influenza vaccine in elderly individuals is associated with decreased production of vaccine-specific antibodies. *J. Clin. Invest.* **121**, 3109–3119 (2011).

5. J. H. Aberle, K. Stiasny, M. Kundi, F. X. Heinz, Mechanistic insights into the impairment of memory B cells and antibody production in the elderly. *Age* **35**, 371–381 (2013).

6. D. G. Schatz, P. C. Swanson, V(D)J Recombination: Mechanisms of Initiation. *Annual Review of Genetics* **45**, 167–202 (2011).

7. J. M. D. Noia, J. M. Di Noia, M. S. Neuberger, Molecular Mechanisms of Antibody Somatic Hypermutation. *Annual Review of Biochemistry* **76**, 1–22 (2007).

8. B. G. Magor, Antibody Affinity Maturation in Fishes—Our Current Understanding. *Biology* **4**, 512–524 (2015).

9. Y. Elhanati, *et al.*, Inferring processes underlying B-cell repertoire diversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370** (2015).

10. J. A. Weinstein, N. Jiang, R. A. White 3rd, D. S. Fisher, S. R. Quake, High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324**, 807–810 (2009).

11. C. F. A. de Bourcy, *et al.*, Phylogenetic analysis of the human antibody repertoire reveals quantitative signatures of immune senescence and aging. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1105–1110 (2017).

12. E. Miho, *et al.*, Computational Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. *Front. Immunol.* **9**, 224 (2018).

13. N. Jiang, *et al.*, Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci. Transl. Med.* **5**, 171ra19 (2013).

14. C. Wang, *et al.*, Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires. *J. Immunol.* **192**, 603–611 (2014).

15. C.-A. Siegrist, R. Aspinall, B-cell responses to vaccination at the extremes of age. *Nat. Rev. Immunol.* **9**, 185–194 (2009).

16. H. Tabibian-Keissar, *et al.*, Aging affects B-cell antigen receptor repertoire diversity in primary and secondary lymphoid tissues. *European Journal of Immunology* **46**, 480–492 (2016).

17. Y. Belkaid, T. W. Hand, Role of the microbiota in immunity and inflammation. *Cell* **157**, 121–141 (2014).

18. S. Magadan, L. Jouneau, P. Boudinot, I. Salinas, Nasal Vaccination Drives Modifications of Nasal and Systemic Antibody Repertoires in Rainbow Trout. *J. Immunol.* **203**, 1480–1492 (2019).

19. D. R. Valenzano, *et al.*, The African Turquoise Killifish Genome Provides Insights into Evolution and Genetic Architecture of Lifespan. *Cell* **163**, 1539–1554 (2015).

20. W. J. Bradshaw, D. R. Valenzano, Extreme genomic volatility characterizes the evolution of the immunoglobulin heavy chain locus in cyprinodontiform fishes. *Proc. Biol. Sci.* **287**, 20200489 (2020).

21. M. Poeschla, D. R. Valenzano, The turquoise killifish: a genetically tractable model for the study of aging. *J. Exp. Biol.* **223** (2020).

22. K. Reichwald, *et al.*, Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. *Cell* **163**, 1527–1538 (2015).

23. M. A. Turchaninova, *et al.*, High-quality full-length immunoglobulin profiling with unique molecular barcoding. *Nat. Protoc.* **11**, 1599–1616 (2016).

24. P. Zajac, S. Islam, H. Hochgerner, P. Lönnerberg, S. Linnarsson, Base preferences in non-templated nucleotide incorporation by MMLV-derived reverse transcriptases. *PLoS One* **8**, e85270 (2013).

25. C. Vollmers, R. V. Sit, J. A. Weinstein, C. L. Dekker, S. R. Quake, Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13463–13468 (2013).

26. C. R. Bolen, F. Rubelt, J. A. Vander Heiden, M. M. Davis, The Repertoire Dissimilarity Index as a method to compare lymphocyte receptor repertoires. *BMC Bioinformatics* **18**, 155 (2017).

27. M. O. Hill, Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **54**, 427–432 (1973).

28. L. Jost, Entropy and diversity. *Oikos* **113**, 363–375 (2006).

29. L. Jost, Partitioning diversity into independent alpha and beta components. *Ecology* **88**, 2427–2439 (2007).

30. K. L. Gibson, *et al.*, B-cell diversity decreases in old age and is correlated with poor health status. *Aging Cell* **8**, 18–25 (2009).

31. H. W. Schroeder Jr, L. Cavacini, Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41–52 (2010).

32. Q. Marcou, T. Mora, A. M. Walczak, High-throughput immune repertoire analysis with IGoR. *Nat. Commun.* **9**, 561 (2018).

33. Q. Marcou, Personal communication (2019).

34. P. Smith, *et al.*, Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. *Elife* **6** (2017).

35. C. Caruso, *et al.*, Mechanisms of immunosenescence. *Immun. Ageing* **6**, 10 (2009).

36. A. Cellerino, D. R. Valenzano, M. Reichard, From the bush to the bench: the annual Nothobranchius fishes as a new model system in biology. *Biol. Rev. Camb. Philos. Soc.* **91**, 511–533 (2016).

37. I. Harel, A. Brunet, The African Turquoise Killifish: A Model for Exploring Vertebrate Aging and Diseases in the Fast Lane. *Cold Spring Harb. Symp. Quant. Biol.* **80**, 275–279 (2015).

38. Z. Xu, *et al.*, Teleost skin, an ancient mucosal surface that elicits gut-like immune responses. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13097–13102 (2013).

39. J. Dodzian, S. Kean, J. Seidel, D. R. Valenzano, A Protocol for Laboratory Housing of Turquoise Killifish (Nothobranchius furzeri). *J. Vis. Exp.* (2018) https:/doi.org/10.3791/57073.

40. K. M. Carter, C. M. Woodley, R. S. Brown, A review of tricaine methanesulfonate for anesthesia of fish. *Rev. Fish Biol. Fish.* **21**, 51–59 (2011).

41. J. A. V. Heiden, *et al.*, pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).

42. N. T. Gupta, *et al.*, Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

43. J. Köster, S. Rahmann, Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).

44. S. Fisher, *et al.*, A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1 (2011).

45. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–40 (2013).

46. N. Nouri, S. H. Kleinstein, Optimized Threshold Inference for Partitioning of Clones From High-Throughput B Cell Repertoire Sequencing Data. *Front. Immunol.* **9**, 1687 (2018).

47. J. N. H. Stern, *et al.*, B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci. Transl. Med.* **6**, 248ra107 (2014).

48. A. Untergasser, *et al.*, Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115–e115 (2012).

# Main text figures

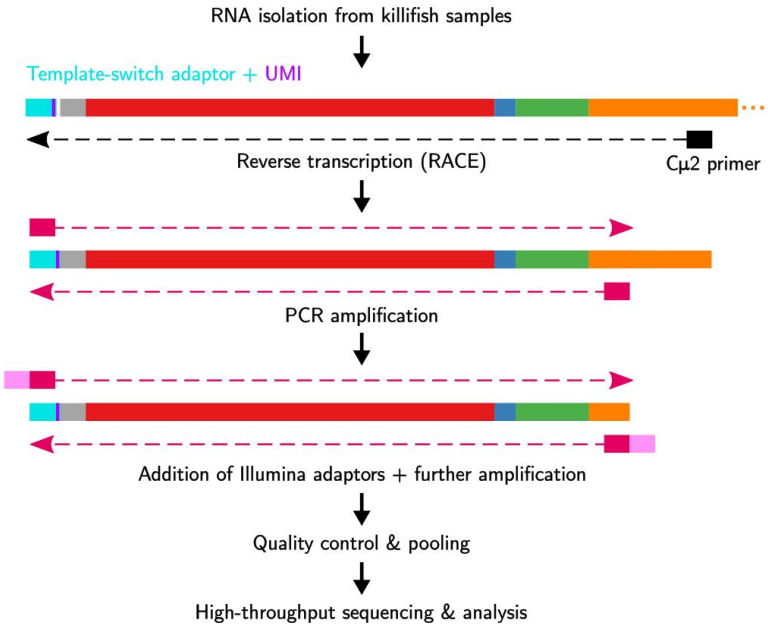


**Fig. 1**. Immunoglobulin sequencing from turquoise-killifish total-RNA samples. Each sample undergoes reverse-transcription with template switching to attach a 5' adaptor sequence and unique molecular identifier (UMI), followed by multiple rounds of PCR amplification and addition of Illumina sequencing adaptors. Libraries are then pooled, undergo size selection, and are sequenced on an Illumina MiSeq sequencing machine

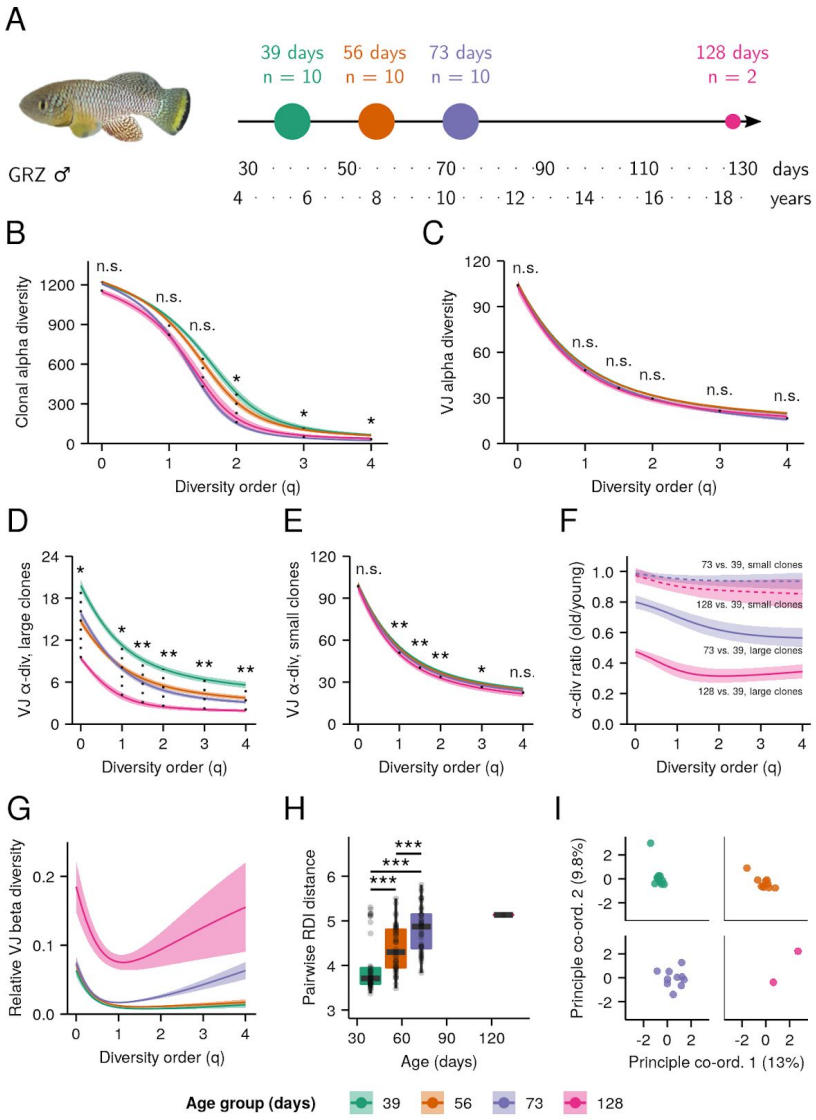


**Fig. 2.** Aging in whole-body killifish *IGH* repertoires. *(A)* Experimental design. Adult male GRZ-strain turquoise killifish were sacrificed at 39, 56, 73 and 128 days post-hatching, flash-frozen and homogenized. *(B-E)* Alpha-diversity spectra, indicating average within-individual repertoire diversity for each age group and diversity order (*: $0.05 \leq 0.01$, **: $0.01 \leq p \leq 0.001$, Kruskal-Wallis permutation test, SI Note 7). *(B)* Clonal alpha-diversity spectra. *(C)* VJ alpha-diversity spectra, all clones. *(D)* VJ alpha-diversity spectra, large clones (> 4 unique sequences) only. *(E)* VJ alpha-diversity spectra, small clones (< 5 unique sequences) only. *(F)* VJ alpha-diversity ratios for old vs young killifish at each diversity order, for small (dashed lines) or large (solid lines) clones. Colour indicates the older age group being compared to young (39 days) fish. *(G)* Normalised VJ beta-diversity spectra, indicating between-individual variability in repertoire composition for each age group and diversity order. *(H)* Distributions of pairwise Repertoire Dissimilarity Index (RDI) distances between individuals in each age group (***: $p \leq 0.001$, Mann-Whitney U tests for pairwise age differences), based on the VJ composition of each individual's repertoire. *(I)* Principal Co-ordinate Analysis (PCoA) of pairwise RDI distances for each age group, visualizing the progressively greater dispersion seen at later ages. Each curve in (B-G) represents the mean across 2000 bootstrap replicates (SI Note 7); shaded regions indicate 95% confidence intervals over the same.

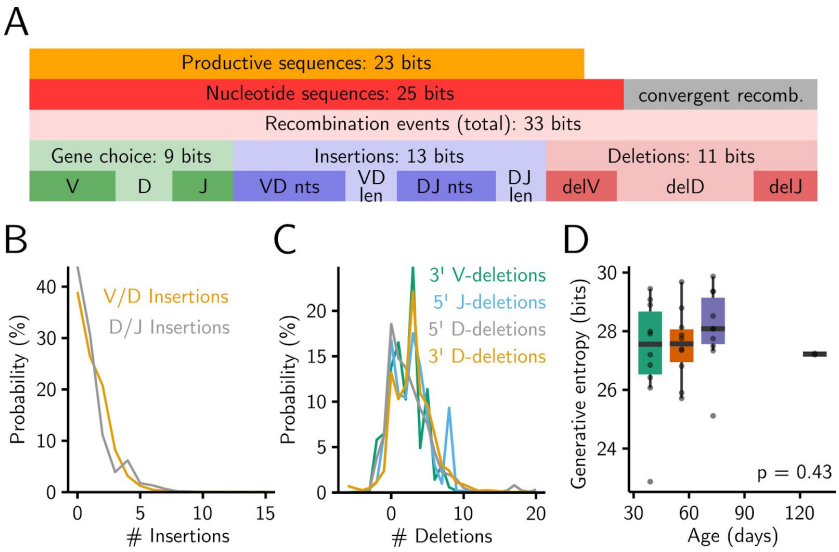


**Fig. 3.** The killifish generative repertoire. *(A)* Entropy composition of the generative process from four eight-week-old GRZ-strain adult male turquoise killifish. *(B)* Probability distributions of junctional N-insertions in the same dataset. *(C)* P-insertions and deletion distributions inferred from the same dataset, with P-insertions modelled as negative deletions. *(D)* Boxplots of total recombination entropy values for models inferred separately for each individual in the 32-individual aging cohort (p = 0.43, Kruskal-Wallis one-way ANOVA for an age effect).

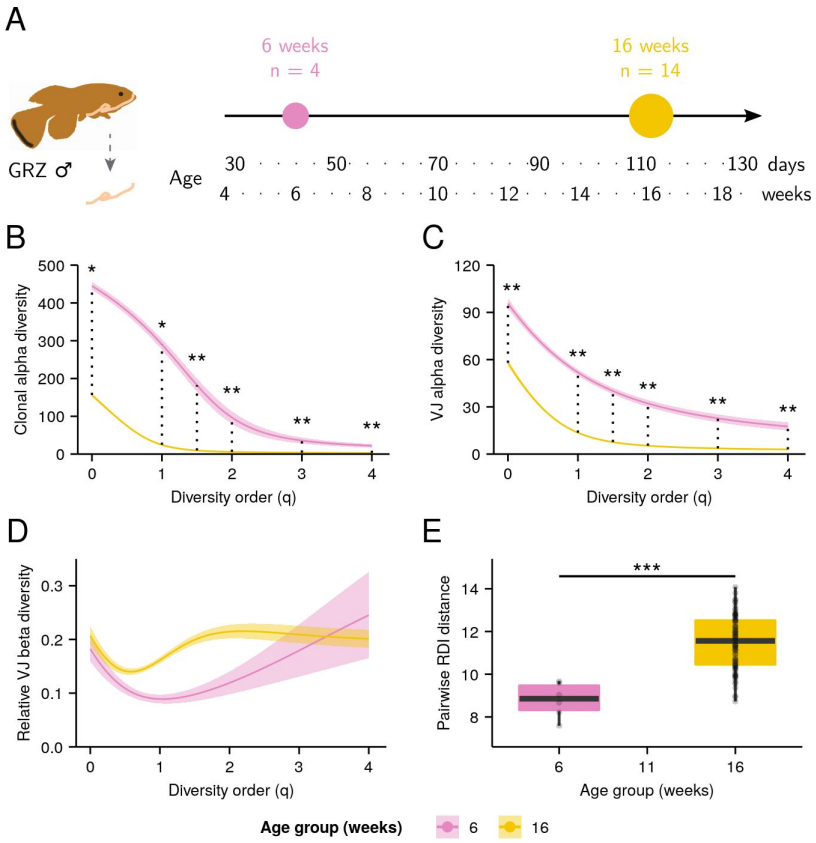


**Fig. 4.** Aging in killifish intestinal repertoires. *(A)* Experimental design. Adult male GRZ-strain turquoise killifish were sacrificed at 6 and 16 weeks post-hatching, and total RNA was extracted from the dissected intestine. *(B-C)* Alpha-diversity spectra, indicating average within-individual repertoire diversity for each age group and diversity order (*: $0.05 \leq 0.01$, **: $0.01 \leq p \leq 0.001$, Kruskal-Wallis permutation test, SI Note 7). *(A)* Clonal alpha-diversity spectra. *(B)* VJ alpha-diversity spectra, all clones. *(D)* Normalised VJ beta-diversity spectra, indicating between-individual variability in repertoire composition for each age group and diversity order. *(E)* Distribution of pairwise RDI distances between killifish intestinal repertoires at different ages (***: $p \leq 0.001$, Mann-Whitney U tests for pairwise age differences). Each curve in (A-C) represents the mean across 2000 bootstrap replicates (SI Note 7); shaded regions indicate 95% confidence intervals over the same.

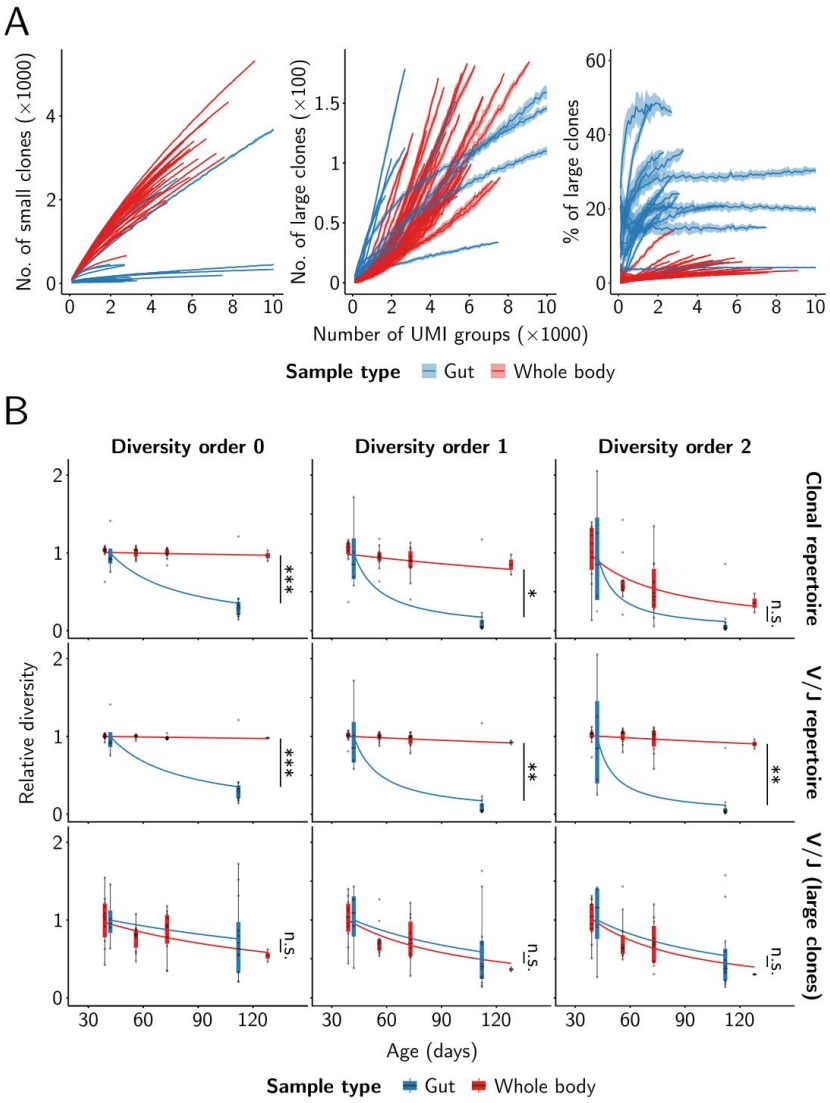


**Fig. 5.** Relative clonal composition and aging phenotypes of whole-body and intestinal repertoires. *(A)* Rarefaction analysis of clonal composition of antibody repertoires from whole-body and intestinal samples, showing the average number of small (left, < 5 unique sequences) and large (middle, 5 unique sequences) clones for each individual across 20 independent replicates at each sample size, as well as the average proportion of all clones in each repertoire which are large (right). Shaded regions around each line show the region within one standard deviation of the mean value. *(B)* Boxplots of individual diversity measurements of repertoires from each age group in the whole-body and intestinal datasets, divided by the mean diversity of the youngest age group in each dataset. Fitted curves show the maximum-likelihood prediction of a Gamma-distributed generalised linear model of diversity vs. age and sample type for the whole body and intestinal dataset, relative to the average diversity of the youngest age group in each experiment, testing for a significant effect of sample type on the rate of diversity change with age (Student's t-test ,*: 0.01 < p  0.05; ***: p  0.001).

# Supporting Information

## Supporting Notes

### **Note 1**: Hill diversity spectra.

The Hill number or "true diversity" $D$ of a population $X$ describes the *effective species richness* of that population: the number of equally-common species that would be required to produce a population with the same overall diversity as $X$ (28, 29). This value will increase with the number of species in the population, as well as with the evenness with which these species are distributed.

The term "species" here can refer to any mutually-exclusive and exhaustive set of classes into which elements in the population can be assigned; the nature of these classes will depend on the nature of the populations and elements being investigated. In the context of immune repertoire sequencing, a "population" is an individual repertoire, an "element" is a unique sequence, and a "species" will typically refer to either clones (sequences descended from a single naive ancestor B cell) or V(D)J gene-segment combinations. In this paper, Hill diversities computed using clones as species are referred to as *clonal diversities*, while those computed using VJ segment combinations as species are referred to as *VJ diversities*.

In addition to the species composition of $X$, the Hill number depends on the diversity order $q$, which denotes the degree to which less-abundant species are downweighted relative to more-abundant species when calculating diversity. At a given value of $q$, the Hill number of $X$ is given by:

$$^{q}D(X) = \left( \sum_{s \in S} p_s^q \right)^{\frac{1}{1-q}}$$

[Equation S1]

where $S$ is the set of species in $X$ and $p_s$ is the relative frequency of a species $s$ (i.e. the fraction of all elements that belong to $s$). Note that, as higher values of $q$ progressively downweight the contributions of less-abundant species, $^{q}D(X)$ is monotonically decreasing with increasing $q$. As each diversity order captures a different aspect of the diversity structure of the population, plotting the value of $^{q}D(X)$ against $q$ – a diversity *spectrum* – gives a more informative view of the diversity of $X$ than any single diversity order alone.

Importantly, many other common diversity metrics can be converted into Hill numbers through simple transformations: ${}^{0}D(X)$ gives species richness, ${}^{2}D(X)$ gives the reciprocal of Simpson's index, $\lim_{q\to\infty} {}^{q}D(X)$ gives the reciprocal of the Berger-Parker index, and $\lim_{q\to 1} {}^{q}D(X)$ gives the exponential of the Shannon entropy (12, 28). Hill diversity measurements of the appropriate order can therefore be easily converted to and from these other diversity metrics if needed. In practice, the exponential shannon entropy is typically substituted for ${}^{1}D(X)$ (which is undefined) to produce a continuous function when plotting Hill diversity spectra:

$$ {}^{q}D(X) = \begin{cases} \left(\sum_{s\in S} p_s^q\right)^{\frac{1}{1-q}} & q \neq 1 \\ \exp\left(-\sum_{s\in S} p_s \cdot \ln p_s\right) & q = 1 \end{cases} $$

[Equation S2]

**Note 2**: Alpha and beta diversity.

In some cases, a population $C$ can be partitioned into some number of disjoint *subpopulations* $X_1, X_2, \ldots, X_M$, each with its own set of species $S_X$ and species frequencies $p_{X,s}$. The diversity of $C$ can then be conceptualised in several ways (29):

- The **gamma diversity** of $C$ is the diversity across the whole population, ignoring its subpopulation structure.
- The **alpha diversity** of $C$ is the diversity arising from differences in species identity among individuals *within* each subpopulation, and is given by a weighted generalised average of the diversities of those subpopulations; it can be thought of as the expected diversity of a single subpopulation drawn at random from $C$.
- The **beta diversity** of $C$ is the diversity arising from variability in species composition *between* subpopulations; it is lowest when all subpopulations have identical species composition and highest when they have no species in common.

In terms of Hill diversity, ${}^{q}D_{\gamma}(C)$ gives the effective richness of the whole population, ignoring its subpopulation structure; ${}^{q}D_{\alpha}(C)$ gives the effective richness of an ``average" subpopulation from $C$, and ${}^{q}D_{\beta}(C)$ gives the effective number of disjoint subpopulations in $C$ (i.e. the number of subpopulations with no species in common that would give rise to the same gamma diversity, given an alpha-diversity value).

The alpha and beta diversities of a population are independent; two different structured populations can have identical alpha and very different beta diversities, or vice versa, depending on the exact species compositions of their subpopulations. The alpha and beta diversity of a structured population also completely determine its gamma diversity (29):

$$D_\gamma(C) = D_\alpha(C) \times D_\beta(C)$$ [Equation S3]

and therefore

$$D_\beta(C) = \frac{D_\gamma(C)}{D_\alpha(C)}$$ [Equation S4]

This (Equation S4) is typically the easiest way of computing the beta diversity of a given structured population. Note that, as the quotient of two monotonically-decreasing functions, $D_\beta(C)$ is itself not necessarily monotonically decreasing.

Under this framework, the diversities of order $q$ for a structured population $C$ are given (29) by

$$^{q}D_\gamma(C) = \left[ \frac{\sum_{s \in S} \left( \sum_{X \in C} w_X p_{X,s} \right)^q}{\left( \sum_{X \in C} w_X \right)^q} \right]^{\frac{1}{1-q}}$$ [Equation S5]

$$^{q}D_\alpha(C) = \left[ \frac{\sum_{X \in C} \sum_{s \in S_X} (w_X p_{X,s})^q}{\sum_{X \in C} w_X^q} \right]^{\frac{1}{1-q}}$$ [Equation S6]

$$^{q}D_\beta(C) = \frac{^{q}D_\gamma(C)}{^{q}D_\alpha(C)}$$ [Equation S7]

where $w_X$ is the relative weighting of a subpopulation $X$. When all the subpopulations have equal sizes $N_X = \frac{N}{M}$ and equal weights $w = \frac{1}{M}$, as is the case with our repertoire data, these formulae simplify to:

$$ {}^{q}D_{\gamma}(C) = \left[ \frac{\sum_{s \in S} \left( \sum_{X \in C} p_{X,s} \right)^{q}}{M^{q}} \right]^{\frac{1}{1-q}} $$

[Equation S8]

$$ {}^{q}D_{\alpha}(C) = \left[ \frac{\sum_{X \in C} \sum_{s \in S_X} (p_{X,s})^{q}}{M} \right]^{\frac{1}{1-q}} $$

[Equation S9]

These equations are valid for all values of $q \in \mathbb{R}$ except 1, providing a spectrum of alpha-, beta- or gamma-diversity measures analogous to the diversity spectra provided for simple populations in SI Note 1. As in that case, a special case needs to be made for $q = 1$ in order to make these functions continuous, substituting the limit of the diversity as $q \to 1$ for the undefined value at $q = 1$:

$$ {}^{1}D_{\gamma}(C) = \exp\left( -\sum_{s \in S} p_s \cdot \ln p_s \right) = {}^{1}D(C) $$

[Equation S10]

$$ {}^{1}D_{\alpha}(C) = \exp\left[ \frac{1}{M} \sum_{X \in C} \left( -\sum_{s \in S_X} (p_{X,s} \cdot \ln p_{X,s}) \right) \right] = \exp\left[ \frac{1}{M} \sum_{X \in C} \ln {}^{1}D(X) \right] $$

[Equation S11]

More detailed derivations of Equation S8 to Equation S11 are available upon request.

## **Note 3**: Rescaling beta diversity

As discussed in SI Note 2, while alpha and gamma diversity are expressed in terms of an effective number of species, beta diversity is expressed in terms of an effective number of subpopulations. Since the effective number of subpopulations is determined in part by the actual number of subpopulations, this means that the beta diversity, unlike alpha and gamma diversity, is directly dependent on the number of subpopulations $M$. If two different structured populations contain different numbers of subpopulations, it is therefore not possible to compare their beta diversity values directly; first, the beta diversity spectra of the populations must be *rescaled* to a common range.

The *minimum* beta diversity of a structured population obtains when all subpopulations have identical species composition. In this case, the beta diversity for the structured population is equal to $1$. The *maximum* beta diversity, meanwhile, obtains when there is no overlap in species between any pair of subpopulations in the structured population. In this case beta diversity for the structured population is equal to the number of subpopulations $M$ (derivations available upon request).

The beta diversity for a structured population with $M$ subpopulations therefore ranges between $1$ (identical composition) and $M$ (maximally-divergent composition). The beta diversities of such a population can thus be transformed onto a new scale from $0$ (minimum beta diversity) to $1$ (maximum beta diversity) as follows:

$$^{q}D_{\beta\,\mathrm{rescaled}}(C) = \frac{^{q}D_{\beta}(C) - {}^{q}D_{\beta\,\min}(C)}{^{q}D_{\beta\,\max}(C) - {}^{q}D_{\beta\,\min}(C)} = \frac{^{q}D_{\beta}(C) - 1}{M - 1}$$

By transforming the beta-diversity spectra of different structured populations onto this common scale, the inter-subpopulation variability of those populations can be meaningfully compared, even if they differ in the number of subpopulations they contain.

**Note 4**: IgSeq library preparation.

To reverse-transcribe antibody heavy-chain transcripts from total RNA samples, and to attach template-switch adapter oligos and unique molecular identifiers (UMIs), 750 ng of total RNA from a killifish sample was combined with 2 µl of 10µmol gene-specific primer (GSP, Table S11), homologous with the second constant-region exon of *N. furzeri* IgM ($C_{\mu}2$) (20). The reaction volume was brought to a total of 8 µl with nuclease-free water, and the resulting mixture was incubated for 2 minutes at 70℃ to denature the RNA, then cooled to 42℃ to anneal the GSP (23). Following annealing, the RNA-primer mixture was combined with 12 µl of reverse-transcription master-mix (Table S6), including the reverse-transcriptase enzyme and template-switch adapter (SmartNNNa, Table S11 and Fig. S19, sequence provided in (23)). The complete reaction mixture was incubated for 1 hour at 42℃ for the reverse-transcription reaction, then mixed with 1µl of uracil DNA glycosylase (QIAGEN, 100 mg $ml^{-1}$) and incubated for a further 40 minutes at 37℃ to digest the template-switch adapter.

Following reverse-transcription, the reaction product was purified using SeraSure SPRI (solid-phase reversible immobilisation) bead preparation (44) (SI Note 5, Table S10). The reaction mixture then underwent three successive rounds of PCR with 2× Kapa HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems), using the general reaction protocol described in Table S7 and reaction-specific volumes and parameters from Table S8. Each

round of PCR was followed by a further round of bead cleanups (SI Note 5, Table S10) to purify the reaction product prior to the next reaction.

Following the final PCR and cleanup, the total concentration of each library was assayed with the Qubit 2.0, while the size distribution of each library was obtained using the TapeStation 4200. To obtain the concentration of complete library molecules in each case (as opposed to primer dimers or other off-target bands), the ratio between the concentration of the desired library band (c. 620-680 bp) and the total concentration of the sample was calculated for each TapeStation lane, and the total concentration of each library as measured by the Qubit was multiplied by this ratio to obtain an estimate of the desired quantity. All the libraries for a given experiment were then pooled, such that the estimated concentration of each library in the final pooled sample was equal and the total mass of nucleic acid in the pooled sample was at least 240 ng. The pooled libraries underwent a final bead cleanup (Table S10) to concentrate the samples, and the beads were separated and removed from the sample.

The pooled and concentrated samples underwent size selection with the BluePippin (Sage Science, 1.5 % DF Marker R2, broad 400-800 bp target band) according to the manufacturer's instructions. The size-selected libraries then underwent a final round of quality control with the Qubit and TapeStation (as above) to confirm their collective concentration (at least 1.5 nmol) and size distribution. Finally, the pooled and size-selected libraries were sequenced on an Illumina MiSeq System (MiSeq Reagent Kit v3, 2×300bp reads, 30% PhiX spike-in) as described in the main text.

### **Note 5**: Nucleic-acid cleanup with SPRI beads

To prepare 50 ml of SeraSure bead suspension, a stock of SeraMag beads (GE Healthcare, 50 mg $ml^{-1}$) was vortexed thoroughly, and 1 ml was transferred to a new tube. This tube was then transferred to a magnetic rack and incubated at room temperature for 1 min, then the supernatant was removed and replaced with 1 ml TET buffer (10 mM Tris base, 1 mM $Na_2$-EDTA, 0.05 % (v/v) Tween 20, pH 8.0) and the tube was removed from the rack and vortexed thoroughly. This washing process was repeated twice more, for a total of three washes in TET. A fourth cycle was used to replace the TET with incomplete SeraBind buffer (4.2 M NaCl, 16.8 mM Tris base, 1.68 mM $Na_2$-EDTA, pH 8.0). The vortexed 1 ml aliquot of beads was then transferred to a conical tube containing 28 ml incomplete SeraBind buffer and mixed by inversion. 20 ml 50 % (w/v) PEG 8000 solution was dispensed slowly down the side of the conical tube, bringing the total volume to 49 ml. Finally, this was brought to 50 ml by adding 250µl 10 % (w/v) Tween 20 solution and 750µl autoclaved water to complete the SeraSure bead suspension.

To perform a bead cleanup, an aliquot of prepared SeraSure suspension was vortexed thoroughly to completely resuspend the beads, then the appropriate relative volume of SeraSure suspension was added to a sample, mixing thoroughly by gentle pipetting. The sample was incubated at room temperature for 5 minutes to allow the beads to bind the DNA, then transferred to a magnetic rack and incubated for a further 5 min to draw as many beads as possible out of suspension. The supernatant was removed and discarded and replaced with 80% ethanol, to a volume sufficient to completely submerge the bead pellet. The sample was incubated for 0.5-1 minutes, then the ethanol was replaced and incubated for a further 0.5-1 minutes. The second ethanol wash was removed, and the tube left on the rack until the bead pellet was almost, but not completely, dry, after which it was removed from the rack. The bead pellet was resuspended in a suitable volume of elution buffer (10 mM Tris-HCl, pH 8.5) then incubated at room temperature for at least 5 minutes to allow the nucleic-acid molecules to elute from the beads.

Unless otherwise specified, the beads from a cleanup were left in a sample during subsequent applications. To remove beads from a sample, the sample was mixed gently but thoroughly to resuspend the beads, incubated for an extended time period (at least 10 minutes) to maximise nucleic-acid elution, then transferred to a magnetic rack and incubated for 2-5 minutes to remove the beads from suspension. The supernatant (containing the eluted nucleic-acid molecules) was then transferred to a new tube, and the beads discarded.

**Note 6**: Pre-processing repertoire-sequencing data

Unless otherwise specified, utilities used in the pre-processing pipelines were provided by the pRESTO (41) and Change-O (42) suites of command-line tools. All code used as part of this pipeline, as well as for downstream analysis and visualisation, will be made available at https://github.com/willbradshaw/killifish-igseq/.

Demultiplexed, adapter-trimmed MiSeq sequencing data were uploaded by the sequencing provider to Illumina BaseSpace and accessed via the Illumina utility program BaseMount. Library annotation information (fish ID, sex, strain, age at death, death weight, etc.) was added to the read headers of each library FASTQ file, and library was assigned a replicate and individual identity. Reads from different sequencing libraries were pooled together, then split by replicate identity; this pooling and re-splitting process enables all reads considered to be a single replicate to be processed together even if sequenced separately, maximising the effectiveness of UMI-based pre-processing, while also allowing all replicates to be processed in parallel.

After pooling and re-splitting, the raw read set underwent quality control, discarding any read with an average Phred score of less than 20. Invariant primer sequences were removed and UMI sequences identified and extracted.

In order to reduce the level of barcode errors in each dataset, primer-masked reads then underwent barcode clustering, in which reads with the same replicate identity and highly similar UMI sequences were grouped together into the same molecular identifier group (MIG). To do this, 5'-reads were first clustered by UMI sequence using a 90 % sequence identity cutoff. In order to split any genuinely distinct MIGs accidentally united by this process, as well as to reduce the level of barcode collisions, the reads within each MIG were clustered again based on their insert sequences; the cluster dendrogram was cut at 75 % total sequence identity, and each subcluster was separated into its own distinct MIG. These clustering thresholds (90 % for barcode clustering, 75 % for barcode splitting) were identified empirically as the values that maximise the number of reads passing downstream quality checks in turquoise-killifish data.

The cluster annotations from these two clustering steps were combined into a single annotation, uniquely identifying each MIG in each replicate. These annotations were further modified to designate the replicate identity of each read, giving each MIG a unique annotation across the entire dataset. These annotations were copied to the reverse reads, such that each read pair had a matching MIG annotation, and reads without a mate (due to differential processing of the two reads files) were discarded. The 5'-reads were then grouped based on MIG identity, and the reads in each cluster grouping were aligned and collapsed into a consensus read sequence; an identical consensus-read-generation step was performed on the reverse reads.

After consensus-read generation had been performed for both 5' and 3' reads, the annotations attached to each read were again unified across read pairs with matching cluster identities, and consensus reads without a mate of the same cluster identity were dropped. Consensus-read pairs with matching cluster annotations were then aligned and merged into a single contiguous sequence. To convert this dataset of MIGs (representing distinct RNA molecules) into one of unique sequences (representing distinct B cells), merged consensus sequences with identical insert sequences but distinct cluster identities were collapsed together into a single FASTQ entry, recording the number, size and UMI makeup of contributing MIGs in each case. Sequences represented by only a single read across all MIGs in the dataset were discarded as unreliable, and the remaining non-singleton sequences from all replicates were combined into a single FASTA file.

V/D/J identities were assigned to each sequence using IgBLAST (45), and the sequences imported into a tab-delimited Change-O database file for downstream processing. Clonotype identities inferred using Change-O's nearest-neighbour distance-threshold technique (42, 46), with replicates from the same individual pooled before undergoing clonotyping to enable comparison of clonotype composition across replicates. A full-length germline sequence was constructed for each sequence entry, and improved V/D/J assignments were made based on the assignments of other sequences in the same clone. Finally, each sequence in the dataset was

annotated according to whether or not it possessed V/D/J assignments and whether these assignments were ambiguous (i.e. whether multiple possible assignments were given rather than just one), and combined VJ and VDJ assignments were obtained by concatenating the individual segment assignments as appropriate. The processed sequence databases were then passed on to downstream analysis pipelines.

## **Note 7**: Diversity-spectrum inference and statistical comparison

Methods for computing diversity spectra were adapted from those presented in the Alakazam R package (42, 47), following completion of the preprocessing pipeline described in SI Note 6.

To ensure comparability between individuals, and obtain estimates of confidence intervals for diversity estimates, the repertoire of each individual underwent multinomial bootstrap resampling: for each of 2000 bootstrap replicates, the same number of unique sequences was sampled, with replacement, from each individual. Following bootstrapping, Hill diversity estimates for each bootstrap replicate and diversity order were calculated as appropriate for the type of diversity being estimated:

- **Individual diversity** spectra (i.e. for individual fish, not aggregated by age group) were computed as specified in Equation S2.
- **Alpha-diversity** spectra for each age group were computed across all individuals in that age group as specified in Equation S9 and Equation S11.
- **Gamma-diversity** spectra for each age group were computed across all individuals in each age group as specified in Equation S8 and Equation S10.
- **Beta-diversity** spectra for each age group were computed as specified in Equation S7, and rescaled as described in SI Note 3.

Having computed diversity values for each bootstrap replicate, diversity spectra were visualised as the mean across all replicates, while 95% confidence intervals were computed empirically as the 0.025- and 0.975-quantiles of diversity across replicates.

To estimate the statistical significance of an age effect on individual repertoire diversity, we adopted a permutation-based approach. For each of 3000 permutations, the age-group assignments of the individuals in the dataset were randomly reshuffled, and a Kruskal-Wallis one-way analysis-of-variance test for an age effect was performed on the individual diversity estimates for each permutation, bootstrap replicate, and diversity order. For each permutation and diversity order, the largest Kruskal-Wallis $H$ statistic across all bootstrap replicates was identified. The same process was repeated on the (unpermuted) real data for each bootstrap replicate and

diversity order, again taking the largest $H$ statistic across all bootstrap replicates for each diversity order. The $p$-value of the age effect at a given diversity order was then computed as the proportion of permutations with a higher maximum $H$ value than that obtained from the real data.

# Supplementary Figures

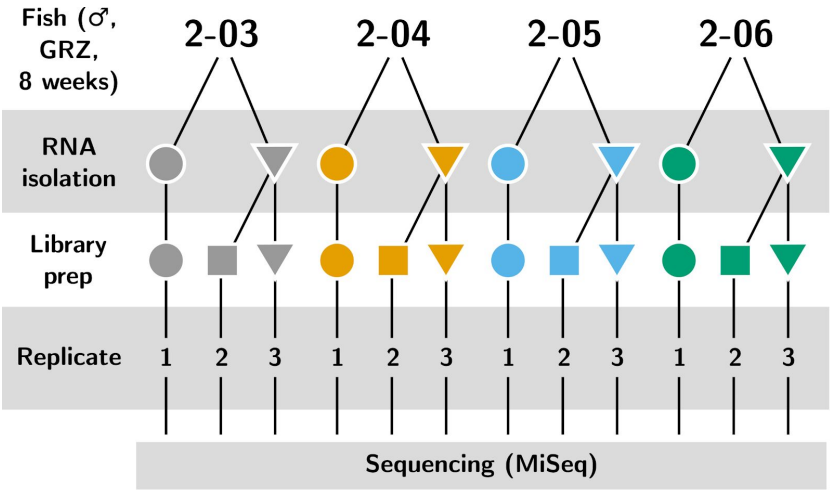


**Fig. S1.** Experimental design of the IgSeq pilot experiment in the turquoise killifish. Four individuals from the 8-week-old group in Table S4 were each independently sequenced three times: once from a separate whole-body total RNA sample (replicate 1) and twice independently from a second sample (replicates 1 and 2).

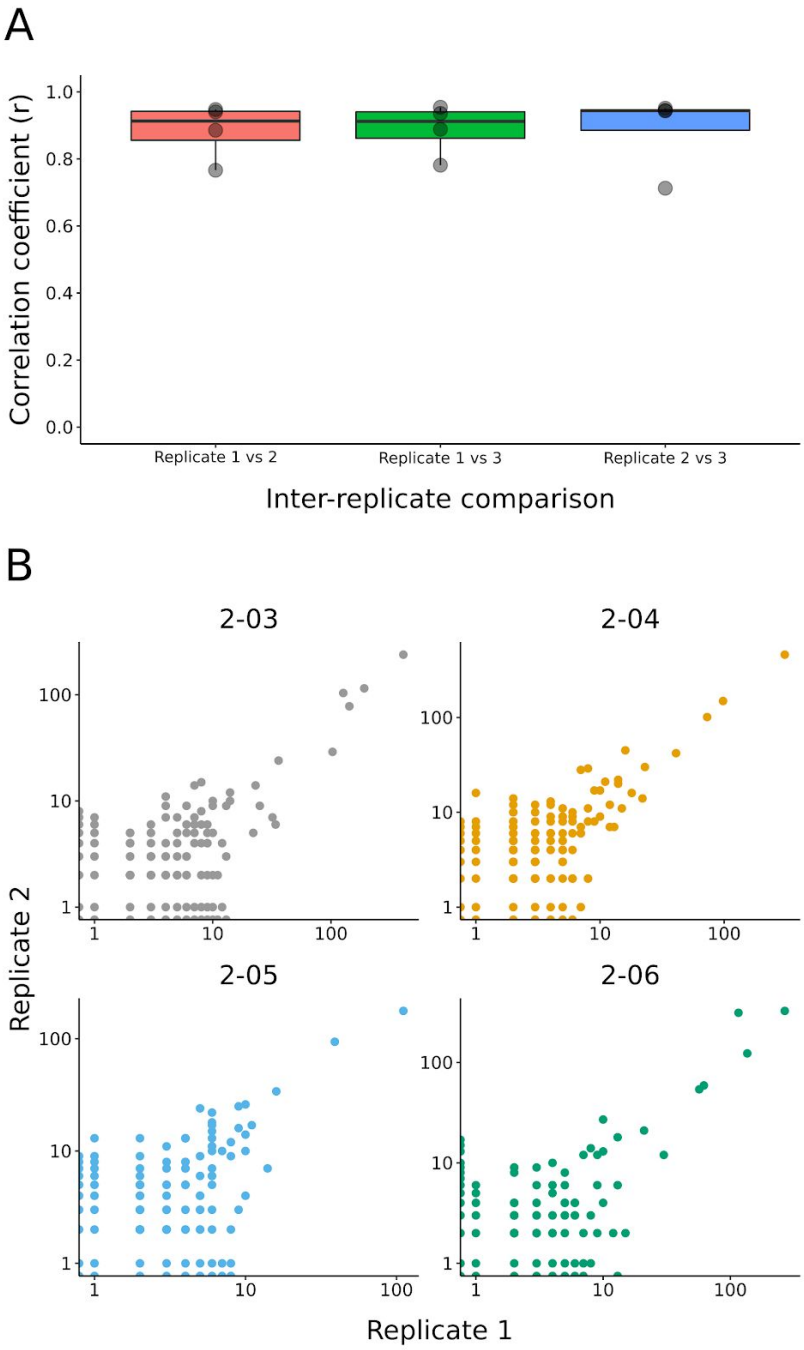


**Fig. S2.** (A) Boxplots of inter-replicate correlation in clone sizes across the four individuals in the pilot experiment (Fig. S1). (B) Example scatter plot for each individual, showing the relationship between the size of a clone in replicate 1 and the same clone in replicate 2.

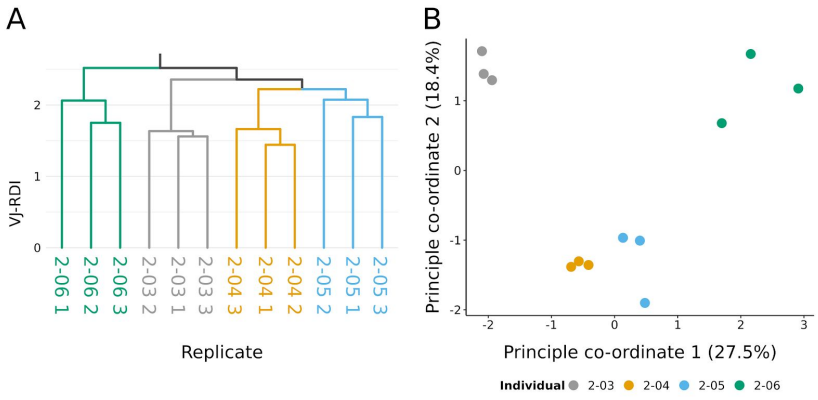


**Fig. S3.** (A) UPGMA clustering dendrogram on repertoire pilot replicates (Fig. S1), based on pairwise RDI distances computed on the V/J composition of each replicate and coloured according to individual origin. (B) Principal Coordinate Analysis (PCoA) of pairwise VJ-RDI distances between all replicates, coloured by individual origin.

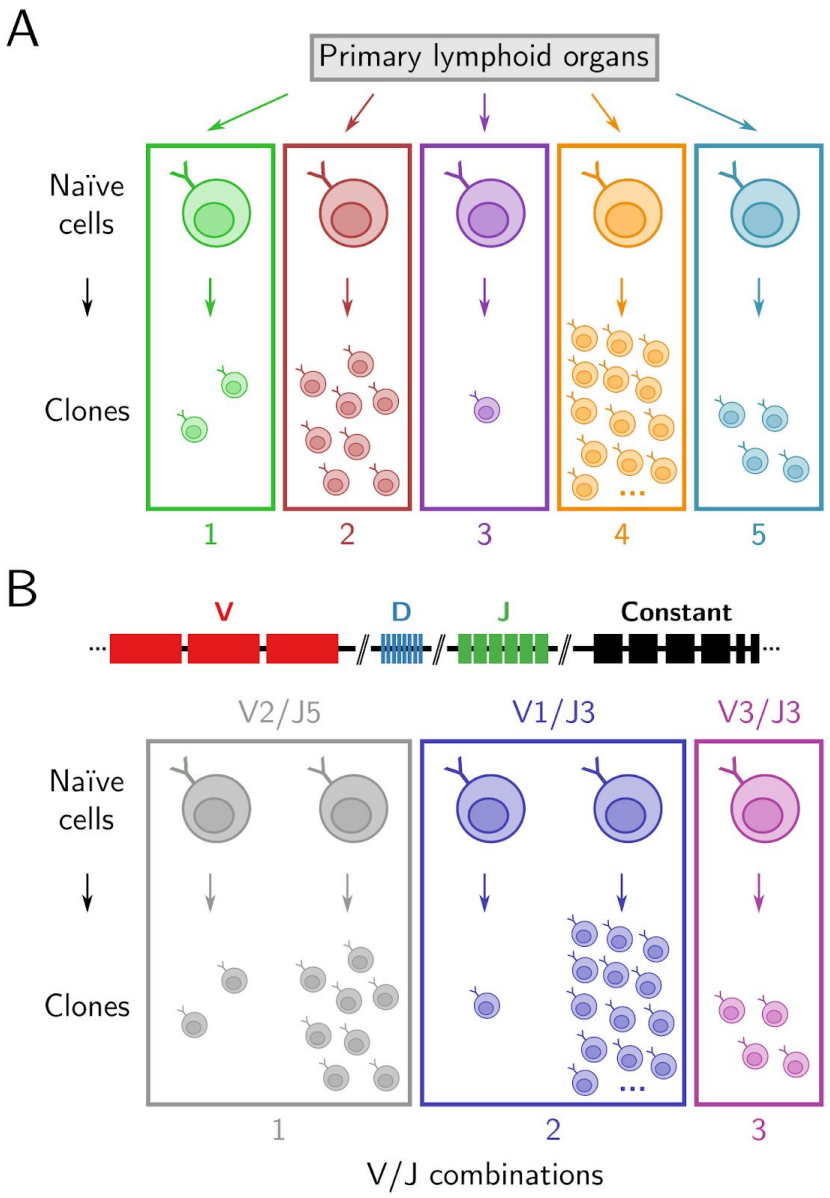


**Fig. S4.** Clonal and V/J diversity in antibody repertoires. (A) A *clone* is a group of B cells descended from the same naïve ancestor cell. In this figure, a group of 28 unique sequences in an antibody repertoire, each representing a single cell, is partitioned into five groups based on their inferred clonal identity. (B) Each developing B cell selects a single V, D, and J gene segment from those present in the native immunoglobulin heavy-chain (*IGH*) locus and recombines them together to produce its antigen-binding sequence. In this figure, the same 28 sequences are instead grouped by the V and J segments their naïve ancestors selected during development. As each developing B cell selects from the same range of possible gene segments, this results in multiple clones being grouped together in a single V/J category, resulting in only 3 groups.

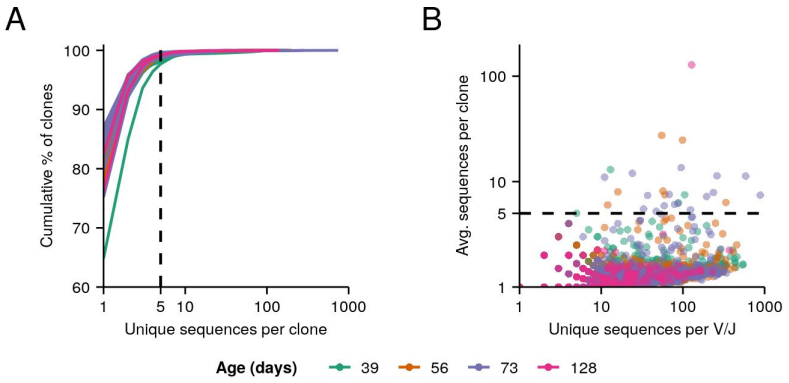


**Fig. S5.** Clone size distributions in killifish whole-body repertoires. (A) Cumulative clone-size distributions for each individual in the aging dataset. (B) V/J combinations present in each individual whole-body repertoire in the dataset, with each point representing a single V/J combination in a single individual, its x-coordinate indicating the number of unique IGH sequences assigned to that V/J combination in that individual, and its y-coordinate indicating the average clone size among those sequences.

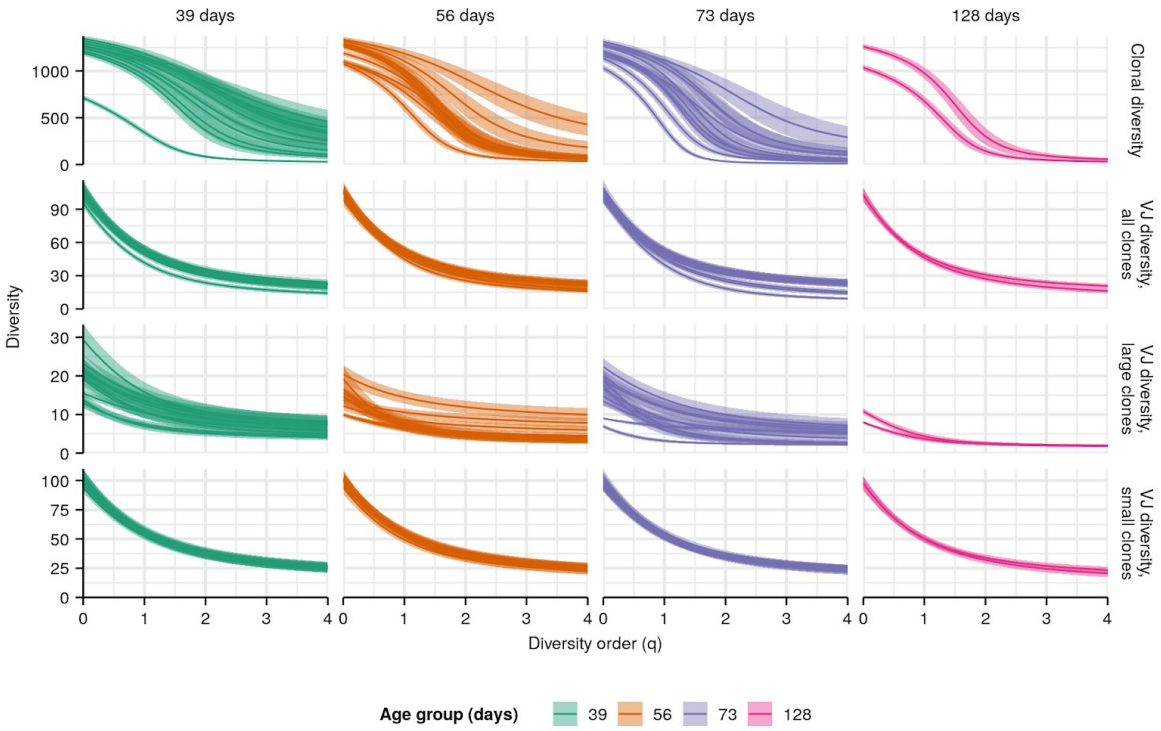


**Fig. S6.** Individual diversity spectra in killifish whole-body repertoires. Clonal (top row) or VJ (other rows) diversity spectra of individual turquoise killifish of different ages. Each curve represents the mean Hill diversity for that individual across 2000 bootstrap replicates (SI Note 7); shaded regions indicate 95% confidence intervals over the same.

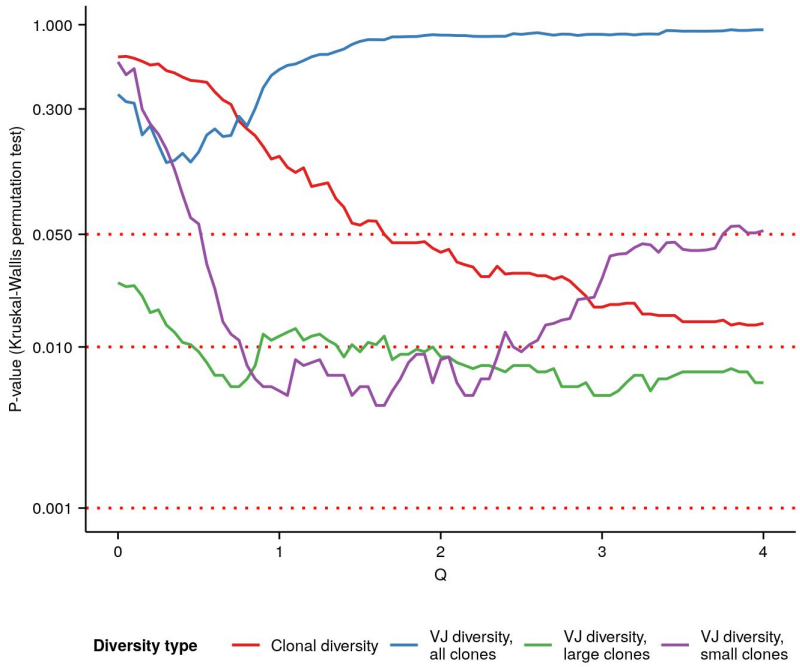


**Fig. S7.** P-values of Kruskal-Wallis permutation tests for an age effect (SI Note 7) in whole-body killifish *IGH* repertoires at different Hill diversity orders. Colour indicates type of diversity (clonal vs VJ etc.) being tested; see main text and SI Note 1 for more details.

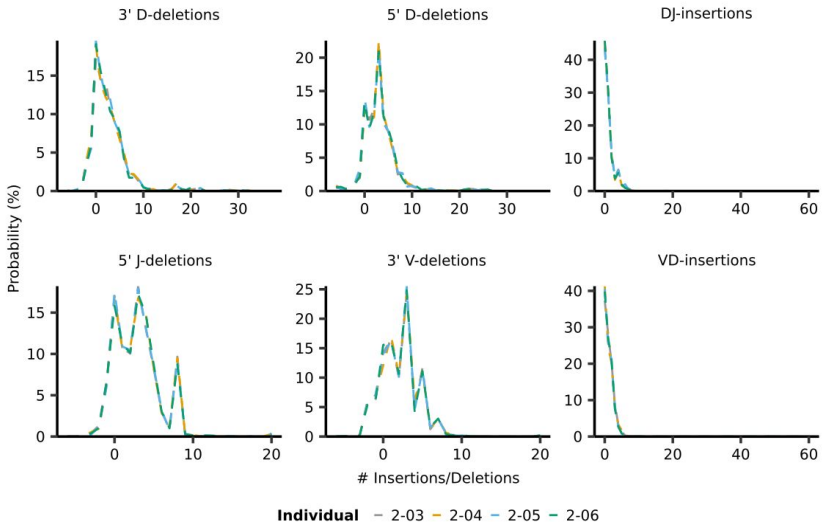


**Fig. S8.** Individual IGoR-inferred insertion/deletion distributions (Fig. 3A) for individuals in the pilot dataset (Fig. S1)

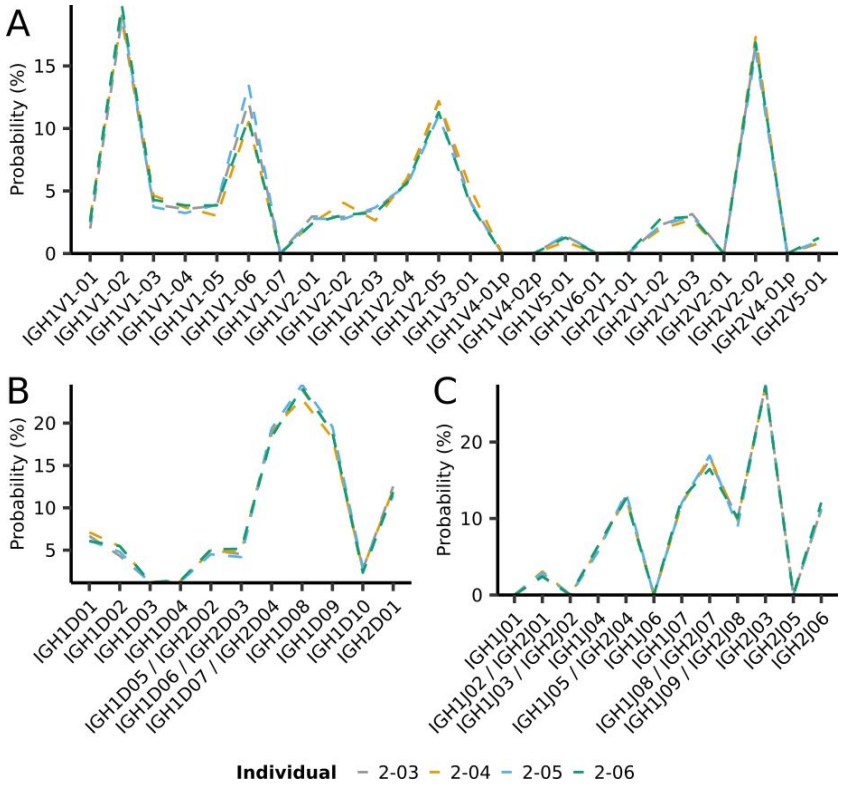


**Fig. S9.** Individual IGoR-inferred segment-usage distributions (Fig. 3B) for individuals in the pilot dataset (Fig. S1)

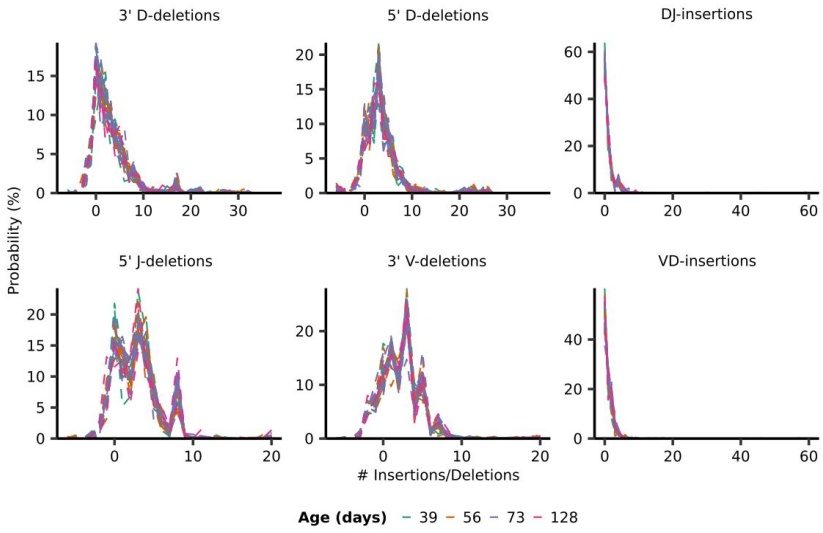


**Fig. S10.** Individual IGoR-inferred insertion/deletion distributions (Fig. 3A) for all individuals in the whole-body killifish dataset (Fig. 2A & Table S4)

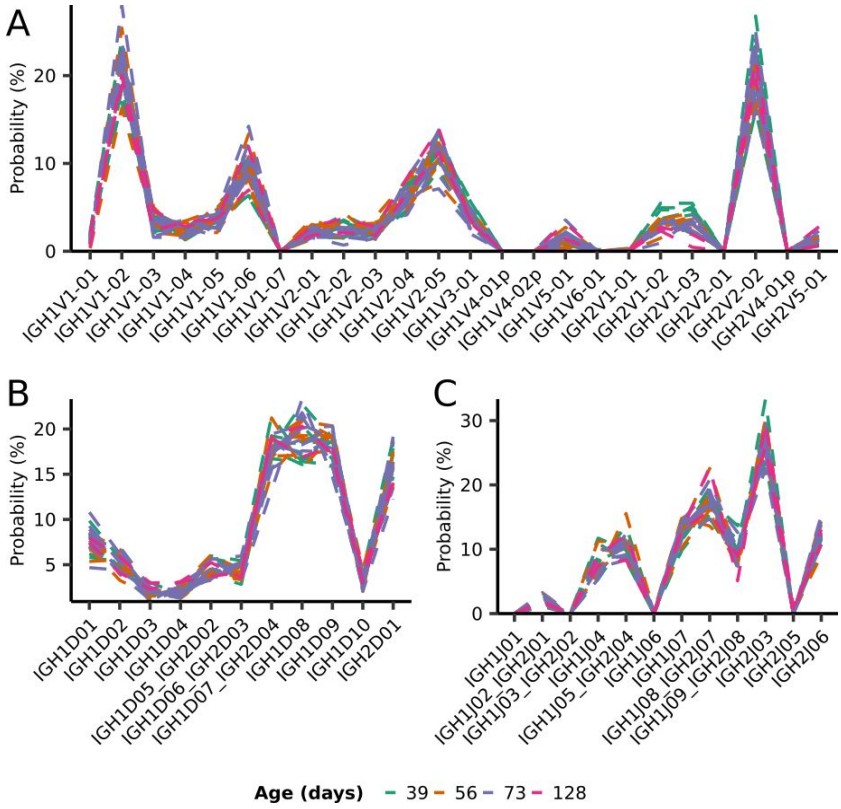


**Fig. S11.** Individual IGoR-inferred segment-usage distributions (Fig. 3B) for all individuals in the whole-body killifish dataset (Fig. 2A & Table S4)

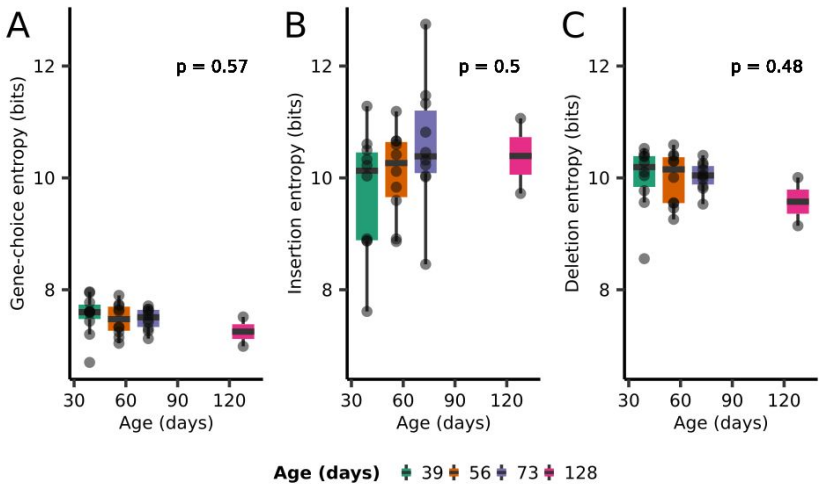


**Fig. S12.** Boxplots of the entropy contributions of *(A)* gene choice, *(B)* N-insertions, and *(C)* P-insertions and deletions to the generative repertoires of individual turquoise killifish at different ages (Fig. 2A & Table S4). The stated *p*-values are from Kruskal-Wallis one-way analysis-of-variance tests for an age effect.

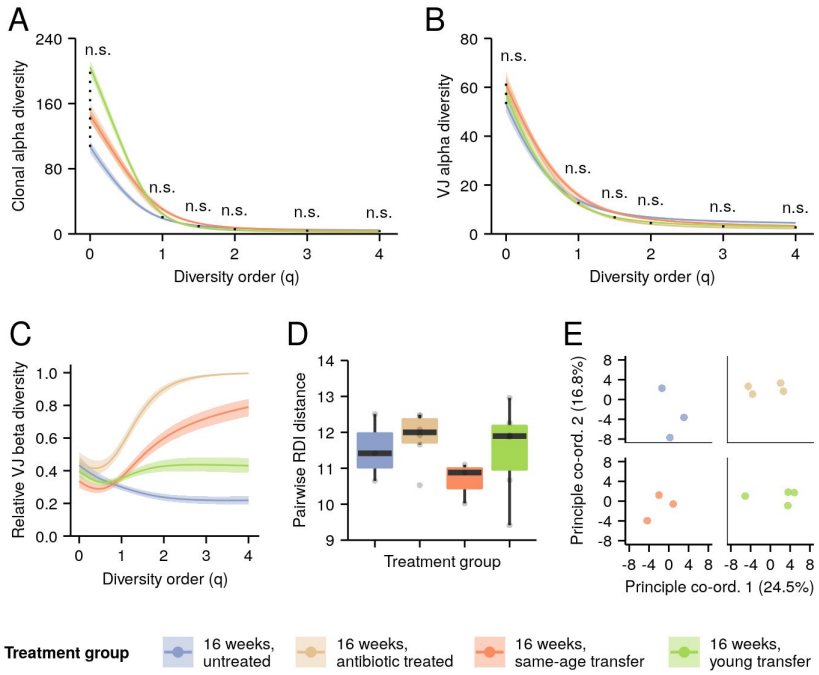


**Fig. S13.** Effect of microbiota transfer on the intestinal repertoires of 16-week-old turquoise killifish. *(A-B)* Alpha-diversity spectra, indicating average within-individual repertoire diversity for each treatment group and diversity order (n.s.: $p > 0.05$ Kruskal-Wallis permutation test, SI Note 7). *(A)* Clonal alpha-diversity spectra. *(B)* VJ alpha-diversity spectra, all clones. *(D)* Normalised VJ beta-diversity spectra, indicating between-individual variability in repertoire composition for each treatment group and diversity order. *(E)* Distribution of pairwise RDI distances between killifish intestinal repertoires of different treatment groups (no significant difference between any two groups, Mann-Whitney U tests for pairwise differences). Each curve in (A-C) represents the mean across 2000 bootstrap replicates (SI Note 7); shaded regions indicate 95% confidence intervals over the same.

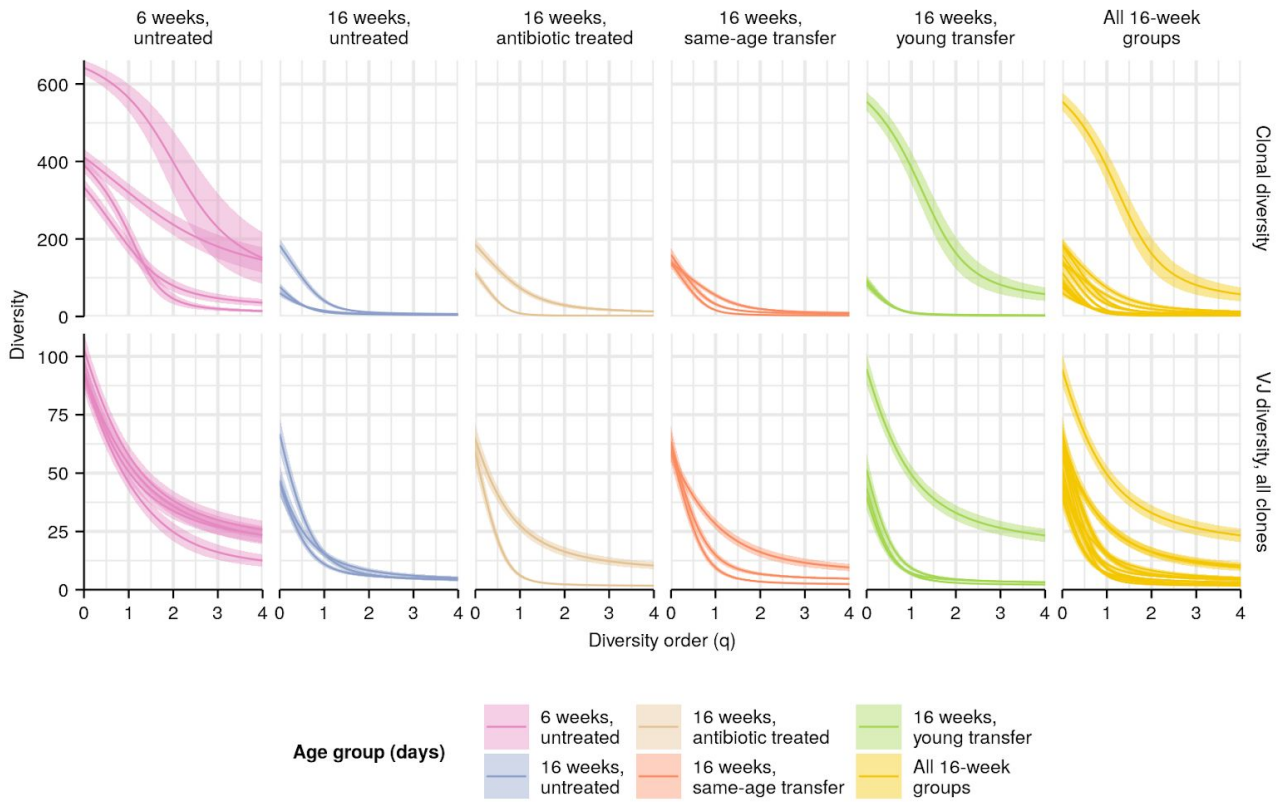


**Fig. S14.** Individual diversity spectra in killifish intestinal repertoires. Clonal (top row) or VJ (bottom row) diversity spectra of the intestines of individual turquoise killifish of different ages. Each curve represents the mean Hill diversity for that individual across 2000 bootstrap replicates (SI Note 7); shaded regions indicate 95% confidence intervals over the same.

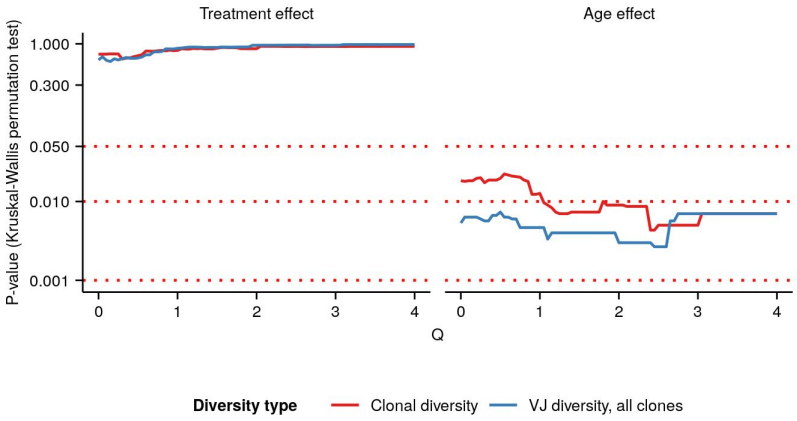


**Fig. S15.** P-values of Kruskal-Wallis permutation tests (SI Note 7) for *(left)* an effect of microbiota transfer treatment group or *(B)* an age effect in intestinal killifish *IGH* repertoires at different Hill diversity orders. Colour indicates type of diversity (clonal vs VJ) being tested; see main text and SI Note 1 for more details.

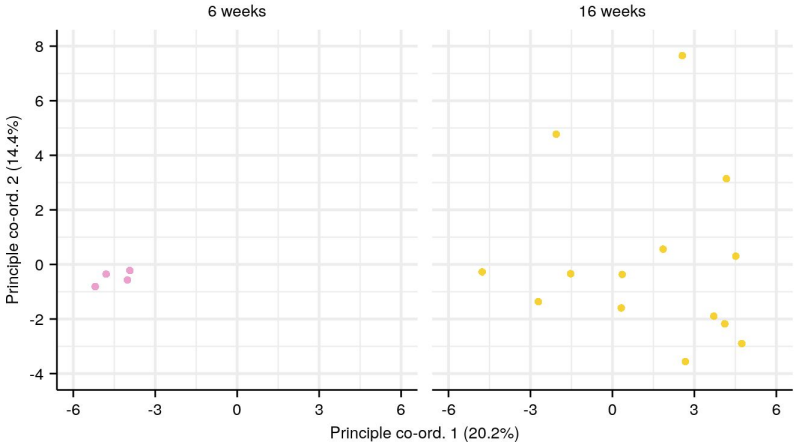


**Fig. S16.** Principal Co-ordinate Analysis (PCoA) of pairwise RDI distances for each age group in the killifish intestinal dataset (Fig. 4E, Table S3).

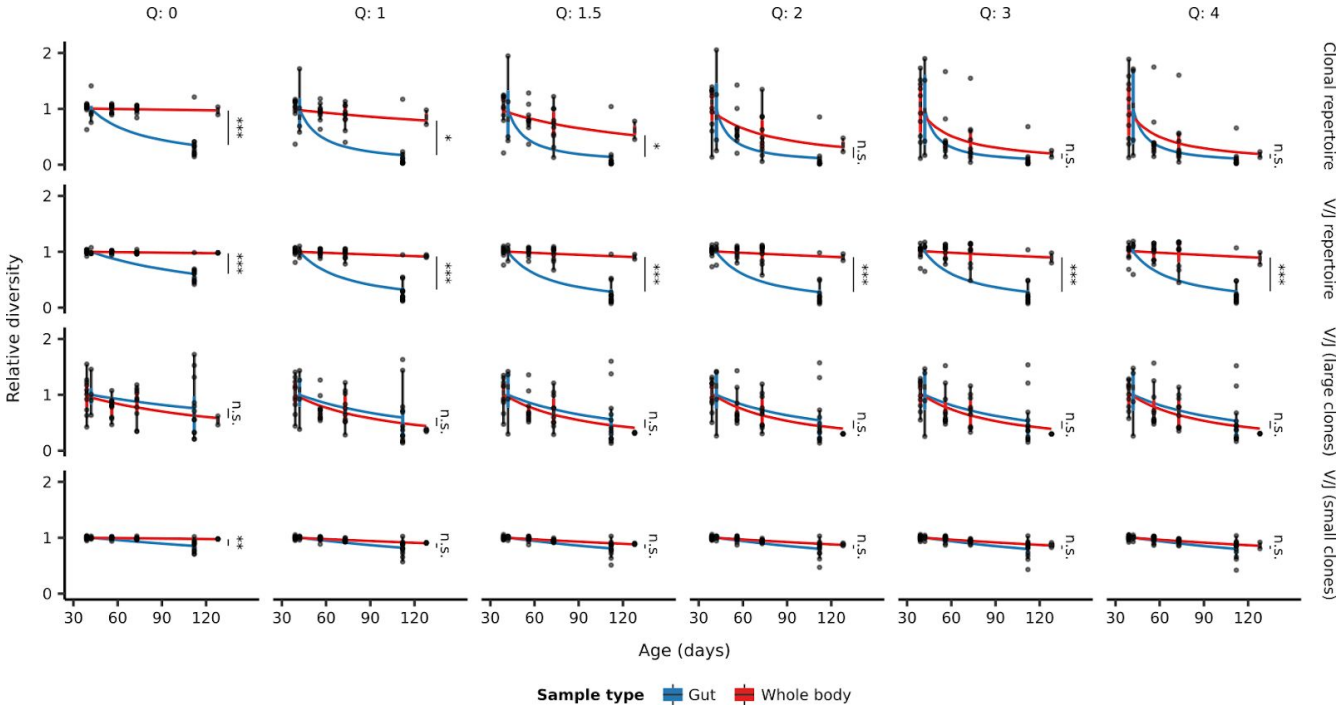


**Fig. S17.** Extended boxplots of individual diversity measurements of repertoires from each age group in the whole-body and intestinal datasets, divided in each case by the mean diversity of the youngest age group in that dataset. Fitted curves show the maximum-likelihood prediction of a Gamma-distributed generalised linear model of diversity vs. age and sample type, testing for a significant effect of sample type on the rate of diversity change with age (Student's t-test ,*: $0.01 < p \leq 0.05$; **: $0.01 < p \leq 0.001$, ***: $p < 0.001$).

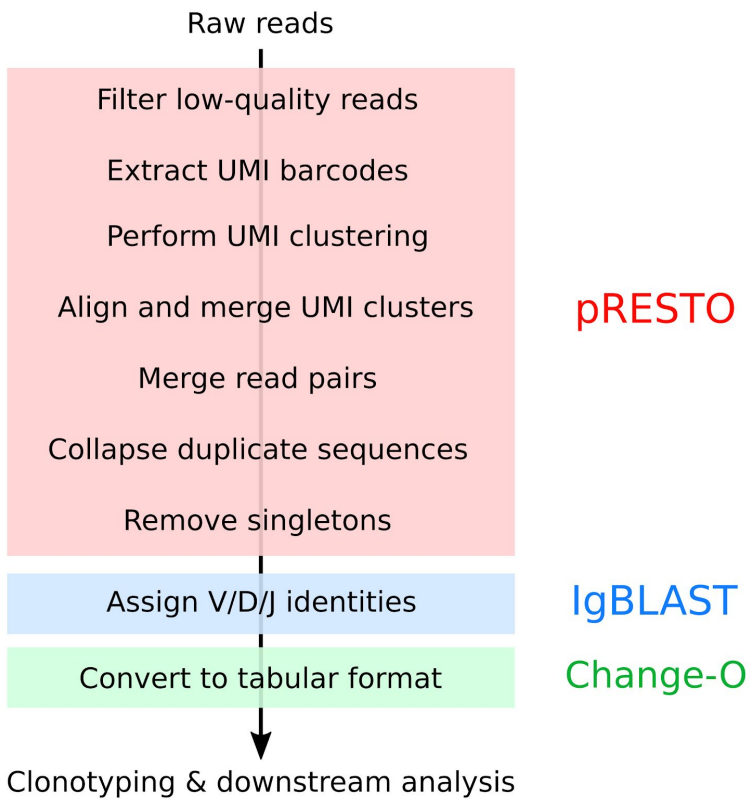


**Fig. S18.** Summary of pre-processing pipeline applied to killifish repertoire sequencing data (SI Note 6).

AAGCAGUGGTAUCAACGCAGAGUNNNN—

—UNNNNUNNNNUCTTrGrGrGrG

**Fig. S19**. The SmartNNNa template-switch adapter. Annotated sequence of the SmartNNNa barcoded template-switch adapter (TSA) used in template-switch reverse transcription for immunoglobulin-sequencing library preparation (23). The 5'-terminal pink characters represent an invariant sequence used for primer-binding in downstream PCR steps, while the grey N characters represent the random nucleotides constituting the unique molecular identifier (UMI), each of which could take any value from A, C, G or T. The blue U residues represent deoxyuridine, which is specifically digested after reverse transcription to remove residual TSA oligos from the reaction mixture. The orange, 3'-terminal rG characters indicate riboguanosine residues, which pair with terminal-transferase added cytidine residues to enable template switching.

# Supplementary tables

**Table S1**. Software versions used in computational analyses

| Program | Version |
|---|---|
| Basemount | 0.15.96.2154 |
| BLAST | 2.7.1 |
| CD-HIT-EST | 4.6.8 |
| Change-O | 0.4.5 |
| EMBOSS (FUZZNUC) | 6.6.0 |
| IgBLAST | 1.7.0 |
| IGoR | 1.3.0 |
| IMGT/DomainGapAlign | 4.9.2 |
| PRANK | v.170427 |
| pRESTO | 0.5.10 |
| Primer3 | 2.3.6 |
| Python 2 | 2.7.14 |
| Python 3 | 3.6.4 |
| R | 3.4.1 / 3.5.2 |
| SAMtools | 1.9 |
| sed | 4.2.2 |
| seqtk | 1.3 |
| Snakemake | 5.3.0 |
| VSEARCH | 2.8.0 |

**Table S2**. Summary of killifish used in IgSeq pilot and aging experiments all fish are GRZ-AD strain and male, and hatched on 9th May 2016.

| Group | # Fish | Death date | Age (d) | Age (wk) | Mean weight (g) |
|---|---|---|---|---|---|
| 1 | 10 | 2016-06-17 | 39 | 5.57 | 1.30 |
| 2 | 10 | 2016-07-04 | 56 | 8.00 | 1.37 |
| 3 | 10 | 2016-07-21 | 73 | 10.4 | 1.76 |
| 4 | 2 | 2016-09-14 | 128 | 18.3 | 2.30 |

**Table S3**. Summary of killifish used in IgSeq gut-microbiota transfer experiment. All fish are GRZ-Bellemans strain and male.

| Group | Age (wk) | # Fish (sequenced / total) | Antibiotic treatment? | Microbiota transfer? |
|---|---|---|---|---|
| YI_6 | 6 | 4 / 4 | No | No |
| WT_16 | 16 | 3 / 4 | No | No |
| ABX_16 | 16 | 4 / 4 | Yes | No |
| SMT_16 | 16 | 3 / 4 | Yes | Yes (9.5wk donor, 9.5wk recipient) |
| YMT_16 | 16 | 4 / 4 | Yes | Yes (6wk donor, 9.5wk recipient) |

**Table S4**. IDs and death weights of all individual turquoise killifish used in IgSeq aging experiment (Table S2).

| Group | # | ID | Death weight (g) |
|---|---|---|---|
| 1 | 1 | grz-AD_4194_E | 1.24 |
| 1 | 2 | grz-AD_4107_E | 1.39 |
| 1 | 3 | grz-AD_4127_E | 1.29 |
| 1 | 4 | grz-AD_4204_E | 1.35 |
| 1 | 5 | grz-AD_4189_E | 1.43 |
| 1 | 6 | grz-AD_4160_E | 0.68 |
| 1 | 7 | grz-AD_4164_E | 1.57 |
| 1 | 8 | grz-AD_4171_E | 1.4 |
| 1 | 9 | grz-AD_4200_E | 1.42 |
| 1 | 10 | grz-AD_4131_E | 1.27 |
| 2 | 1 | grz-AD_4159_E | 1.37 |
| 2 | 2 | grz-AD_4179_E | 1.47 |
| 2 | 3 | grz-AD_4152_E | 1.33 |
| 2 | 4 | grz-AD_4132_E | 1.35 |
| 2 | 5 | grz-AD_4177_E | 1.22 |
| 2 | 6 | grz-AD_4158_E | 1.51 |
| 2 | 7 | grz-AD_4182_E | 1.12 |
| 2 | 8 | grz-AD_4202_E | 1.54 |
| 2 | 9 | grz-AD_4143_E | 1.28 |
| 2 | 10 | grz-AD_4201_E | 1.55 |
| 3 | 1 | grz-AD_4155_E | 2.06 |
| 3 | 2 | grz-AD_4193_E | 1.92 |
| 3 | 3 | grz-AD_4170_E | 1.8 |
| 3 | 4 | grz-AD_4135_E | 1.65 |
| 3 | 5 | grz-AD_4190_E | 1.87 |
| 3 | 6 | grz-AD_4099_E | 1.94 |
| 3 | 7 | grz-AD_4198_E | 1.49 |
| 3 | 8 | grz-AD_4024_E | 1.73 |
| 3 | 9 | grz-AD_4044_E | 1.53 |
| 3 | 10 | grz-AD_4117_E | 1.57 |
| 4 | 1 | grz-AD_4173_E | 2.2 |
| 4 | 2 | grz-AD_4197_E | 2.4 |

**Table S5**. Individual killifish samples used in IgSeq gut-microbiota transfer (Table S3).

| Fish ID | Group | RNA Integrity Number | Date of library prep | Sequenced? |
|---|---|---|---|---|
| 1271 | ABX_16 | 7.7 | 2018-11-20 | Yes |
| 1274 | ABX_16 | 7.3 | 2018-12-01 | Yes |
| 1309 | ABX_16 | 6.9 | 2018-12-01 | Yes |
| dash | ABX_16 | 6.8 | 2018-12-01 | Yes |
| 1015 | SMT_16 | 7 | 2018-11-20 | Yes |
| 1298 | SMT_16 | 5.5 | 2018-11-20 | Yes |
| 1301 | SMT_16 | 5.9 | 2018-11-20 | Yes |
| 402 | WT_16 | 7 | 2018-11-20 | Yes |
| 938 | WT_16 | 7.8 | 2018-11-20 | Yes |
| 940 | WT_16 | 7.2 | 2018-11-20 | Yes |
| 1403 | YI_6 | 5.5 | 2018-11-20 | Yes |
| 1409 | YI_6 | 6.4 | 2018-12-01 | Yes |
| 1412 | YI_6 | 6.7 | 2018-11-20 | Yes |
| 1414 | YI_6 | 5.5 | 2018-11-20 | Yes |
| 1009 | YMT_16 | 7 | 2018-11-20 | Yes |
| 1026 | YMT_16 | 6.3 | 2018-11-20 | Yes |
| 1305 | YMT_16 | 7.3 | 2018-11-20 | Yes |
| 999 | YMT_16 | 7.1 | 2018-12-01 | Yes |
| 400 | WT_16 | 7.1 | — | No (not enough RNA) |
| 1005 | SMT_16 | 4.4 | — | No (RNA integrity too low) |

**Table S6**. Master-mix components for reverse transcription reaction used in library preparation protocol (per sample).

| Volume (µl) | Component | Initial concentration |
|---|---|---|
| 2 | SMARTScribe reverse transcriptase | 100 U $\mu l^{-1}$ |
| 3 | SMARTScribe first-strand buffer | 5× |
| 2 | SmartNNNa barcoded TSA | 10 µM |
| 2 | DTT | 20 mM |
| 2 | dNTP mix | 10 µM per nucleotide |
| 0.5 | RNasin RNase inhibitor | 40 U $\mu l^{-1}$ |

**Table S7**. Overall PCR thermocycler protocol for library preparation protocol (see Table S8 for missing parameter values).

| Step | Temperature (℃) | Duration (s) | # Cycles |
|---|---|---|---|
| Initial denaturation | 95 | 180 | 1 |
| Denaturation | 98 | 20 | $n_c$ (see Table S8) |
| Annealing | $T_a$ (see Table S8) | 15 | |
| Extension | 72 | $t_{ext}$ (see Table S8) | |
| Final extension | 72 | $t_{ext}$ × 4 (see Table S8) | 1 |

**Table S8**. Specific PCR protocols used for different stages of *N. furzeri* immunoglobulin sequencing library preparation protocol (see Table S7 for overall thermocycler protocol and Table S9 for cycle numbers).

| PCR Step | Protocol details | | | Primers[a] | | Volumes (µl)[b] | | | |
|---|---|---|---|---|---|---|---|---|---|
| | # cycles | $T_a$ (℃) | $t_{ext}$ (s) | F | R | Template | Primers[c] | Kapa | $H_2O$ |
| 1 | 18 | 65 | 15 | IGHC-B | M1SS | 10.5 | 1 (2×) | 12.5 | 0 |
| 2 | 13 | 65 | 15 | M1S+P2 | IGHC-C+P1 | 1 | 0.5 (2×) | 12.5 | 10.5 |
| 3 | 8 to 12[d] | 68 | 15 | P2 | P1 | 2 | 0.75 (2×) | 12.5 | 9 |

[a] See Table S11 for sequence information.
[b] If the number of samples to be sequenced was small, all volumes of PCR 3 were doubled for a 50µl total PCR volume.
[c] The stated volumes apply separately to both forward and reverse primers for each reaction. All primers were diluted to and stored at an initial concentration of 10 µM.
[d] See Table S9 for specific cycle numbers used in each experiment.

**Table S9**. PCR cycle numbers during N. furzeri IgSeq library preparation protocol (Table S7 and Table S8).

| Experiment | Number of PCR cycles | | |
|---|---|---|---|
| | PCR 1 | PCR 2 | PCR 3 |
| Pilot experiment (Fig. S1) | 18 | 13 | 9 |
| 32-individual aging cohort | 18 | 13 | 8 |
| Intestinal RNA samples | 18 | 13 | 12 |

**Table S10**. Bead cleanups during *N. furzeri* IgSeq library preparation protocol.

| Stage[a] | Sample volume (µl) | Beads volume (µl) | | Elution volume (µl)[c] |
|---|---|---|---|---|
| | | µl | ×[b] | |
| RT | 21 | 14.7 | 0.7 | 16.5 |
| PCR 1 | 25 | 17.5 | 0.7 | 25 |
| PCR 2 | 25 | 17.5 | 0.7 | 15 |
| PCR 3 | 25[d] | 20[d] | 0.8 | 15[d] |
| Pooling | Varies[e] | Varies[e] | 2.5 | 35 |

[a] Each bead cleanup takes place immediately *after* its corresponding stage.
[b] Bead volumes are usually given as multiples of the sample volume.
[c] All elutions performed in the specified volume of elution buffer (10 mM Tris-HCl, pH 8.5).
[d] If the PCR 3 reaction volume differed from 25 µl, bead and elution volumes were rescaled proportionally to sample volume as appropriate.
[e] In each experiment, samples were pooled in equimolar ratio, so the input volume depended on the number of samples and the concentration of the libraries.

**Table S11**. Primers and oligonucleotides used in N. furzeri IgSeq library preparation protocol.

| Name | Sequence | Purpose |
|---|---|---|
| SmartNNNa[a] | AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNU NNNNUCTTrGrGrGrG | Template-switch adapter |
| GSP | TGGTCTTGCCAGCTGGTGATTTCCGCC | Cμ2-specific reverse-transcription primer |
| M1SS | AAGCAGTGGTATCAACGCA | PCR1 forward primer |
| IGH-B | CCACATGGCACCAGAGGAAAC | PCR1 reverse primer |
| M1S+P2 | GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TCAGTGGTATCAACGCAGAG | PCR2 forward primer |
| IGH-C+P1 | ACACTCTTTCCCTACACGACGCTCTTCCGATCT ATGGCACCAGAGGAAACACAAC | PCR2 reverse primer |
| P2 | ACAAGCAGAAGACGGCATACGAGATNNNNNNNN GTGACTGGAGTTCAGACGTGTGC | PCR3 forward primer |
| P1 | AATGATACGGCGACCACCGAGATCTACACNNNN NNNNACACTCTTTCCCTACACGACGC | PCR3 reverse primer |

[a] See Fig. S19 for more information about the structure of the template-switch adapter.
[b] For each P2 oligo (D701-D712), the N characters are replaced by the appropriate Illumina i7 index sequence from Table S13.
[c] For each P1 oligo (D501-D508), the highlighted N characters are replaced by the appropriate Illumina i5 index sequence from Table S12.
[d] The template-switch adapter and primer sequences homologous to it (M1SS and M1S) were provided by Turchaninova *et al.* (23), while those homologous to constant-region exons (GSP, IGH-B and IGH-C) were designed using Primer3 (48).

**Table S12**. i5 index sequences

| Name | Index sequence |
|---|---|
| D501 | AGGCTATA |
| D502 | GCCTCTAT |
| D503 | AGGATAGG |
| D504 | TCAGAGCC |
| D505 | CTTCGCCT |
| D506 | TAAGATTA |
| D507 | ACGTCCTG |
| D508 | GTCAGTAC |

**Table S13**. i7 index sequences

| Name | Index sequence |
|---|---|
| D701 | CGAGTAAT |
| D702 | TCTCCGGA |
| D703 | AATGAGCG |
| D704 | GGAATCTC |
| D705 | TTCTGAAT |
| D706 | ACGAATTC |
| D707 | AGCTTCAG |
| D708 | GCGCATTA |
| D709 | CATAGCCG |
| D710 | TTCGCGGA |
| D711 | GCGCGAGA |