HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2016 December 03.

The African turquoise killifish genome provides insights into evolution and genetic architecture of lifespan

Dario Riccardo Valenzano[1,9,7,10], **Bérénice A. Benayoun**[1,10], **Param Priya Singh**[1,10], **Elisa Zhang**[1], **Paul D. Etter**[2], **Chi-Kuo Hu**[1], **Mathieu Clément-Ziza**[3], **David Willemsen**[4], **Rongfeng Cui**[4], **Itamar Harel**[1], **Ben E. Machado**[1], **Muh-Ching Yee**[1,8], **Sabrina C. Sharp**[1], **Carlos D. Bustamante**[1], **Andreas Beyer**[5], **Eric A. Johnson**[2], and **Anne Brunet**[1,6,9]

[1]Department of Genetics, Stanford University, California 94305, USA

[2]Institute of Molecular Biology, University of Oregon, Oregon 97403, USA

[3]CMMC, University of Cologne, Cologne, 50931, Germany

[4]Max Planck Institute for Biology of Ageing, Cologne, 50931, Germany

[5]Cellular Networks and Systems Biology, CECAD, University of Cologne, Cologne, 50931, Germany

[6]Glenn Laboratories for the Biology of Aging, Stanford University, California 94305, USA

Summary

Lifespan is a remarkably diverse trait ranging from a few days to several hundred years in nature, but the mechanisms underlying the evolution of lifespan differences remain elusive. Here we de novo assemble a reference genome for the naturally short-lived African turquoise killifish, providing a unique resource for comparative and experimental genomics. The identification of genes under positive selection in this fish reveals potential candidates to explain its compressed

[9]Corresponding authors: anne.brunet@stanford.edu; dario.valenzano@age.mpg.de.
[7]Present address: Max Planck Institute for Biology of Ageing, Cologne, 50931, Germany
[8]Present address: Carnegie Institution for Science, Stanford, California 94305, USA
[10]Equal contribution

lifespan. Several aging genes are under positive selection in this short-lived fish and long-lived species, raising the intriguing possibility that the same gene could underlie evolution of both compressed and extended lifespans. Comparative genomics and linkage analysis identify candidate genes associated with lifespan differences between various turquoise killifish strains. Remarkably, these genes are clustered on the sex chromosome, suggesting that short lifespan might have co-evolved with sex determination. Our study provides insights into the evolutionary forces that shape lifespan in nature.

Introduction

Nature offers an amazing diversity in the lifespan of species with a 150,000 fold difference between the shortest-lived and longest-lived species (Austad, 2010). Shortlived species have long been recognized as essential for experimental studies, and central pathways that regulate aging – notably the insulin-FOXO and mTOR pathways – have been discovered in short-lived models such as yeast, worms, and flies (Kaeberlein and Kennedy, 2011; Kapahi et al., 2010; Kenyon, 2010). These pathways have turned out to modulate aging all the way to humans (Flachsbart et al., 2009; Johnson et al., 2013). Many short-lived species tend to occupy unique ecological niches where the environment is harsh. Thus, these species can be used experimentally for the identification of conserved genes important for lifespan and can also provide insight into the evolutionary forces that shape lifespan strategies in nature.

At the other end of the lifespan spectrum, exceptionally long-lived species, such as the naked mole rat (~30 years), the Brandt's bat (~40 years), and the bowhead whale (~200 years) have recently been used in comparative genomic studies to identify genes and residues that are uniquely changed or under positive selection in these organisms (Keane et al., 2015; Kim et al., 2011; Seim et al., 2013). Interestingly, several genes involved in aging and metabolic pathways are uniquely changed in the Brandt's bat (e.g. insulin-like growth factor 1 receptor [*IGF1R*] and growth hormone receptor) (Seim et al., 2013) and in the naked mole rat (e.g. the uncoupling protein *UCP1*) (Kim et al., 2011). Despite these advances, much remains to be learned about the genes that drive natural differences in lifespan and their evolutionary selection in the wild.

Here we examine the evolution and genetic architecture of lifespan in a naturally short-lived vertebrate, the African turquoise killifish (*Nothobranchius furzeri*), which has been proposed as a vertebrate model for aging (Valdesalici and Cellerino, 2003). We report the de novo assembly and annotation of a reference genome for the turquoise killifish. Evolutionary genomics reveals genes that are positively selected in this shortlived fish and may underlie unique traits in this species, including its short life cycle. By performing a linkage analysis between two strains with experimental lifespan differences, we also identify candidate genes that are likely associated with lifespan differences between strains. Our study provides a resource for experimental and comparative genomics and gives insights into the evolution of naturally short lifespans in nature.

Results

The African turquoise killifish: a vertebrate with a naturally compressed lifespan

The African turquoise killifish provides a unique system to test the evolution and genetics of longevity (Cellerino et al., 2015). This fish species comprises distinct populations that populate ephemeral ponds in arid regions of Zimbabwe and Mozambique (Figure 1A, 1B). The ponds are only present for 4–6 months during the wet season, and the turquoise killifish has evolved a state of embryonic diapause – 'suspended animation' – to survive through the dry season (Figure 1C). In the laboratory, diapause can be skipped, and this fish has a captive lifespan of 4–6 months (Figure 1C). Thus far, the turquoise killifish is the shortest-lived vertebrate that can be bred in captivity (Figure 1D). Interestingly, strains derived from the various populations of turquoise killifish from different regions of Zimbabwe and Mozambique can exhibit different experimental lifespans under similar conditions (Terzibasi et al., 2008) (Figure 1D). The extreme diversity in lifespan between the turquoise killifish (4–6 months) and other species, as well as the presence of strains with different captive lifespans, offers an unprecedented paradigm to explore the evolution of lifespan.

De novo assembly and annotation of the turquoise killifish genome

To gain insight into the evolution and genetic architecture of lifespan in the turquoise killifish, we de novo sequenced and assembled its genome. The GRZ strain, which originates from Zimbabwe, was selected to build the reference genome because it is inbred and has a low percentage of heterozygosity (Kirschner et al., 2012; Valenzano et al., 2009). By sequencing 10 paired-end or mate-pair Illumina libraries with a range of insert sizes, we obtained a 1.02Gb genome assembly with a contig N50 of 9.3kb and a scaffold N50 of 118kb (Figure 2A). Using paired-end RNA-seq libraries and our high-density linkage map (see Figure 6C), we improved the contiguity of the genome resulting in a scaffold N50 of 247kb (Figure 2A). The computational genome size estimate ranges from 1.3Gb to 2.2Gb (Figure 2A). The assembly statistics for our turquoise killifish genome are in the range of Illumina-based genome assemblies, although the assembly is still fragmented likely due to the high number of repeats (over 45% (Reichwald et al., 2009)) (Figure S1A). We next assessed the completeness and quality of the turquoise killifish genome assembly. Computational assessment indicates that the assembly contains 96.8% of core eukaryotic genes (Figure 2A). Importantly, assembled transcripts, paired-end RNA-seq reads, and contigs obtained by Sanger shotgun sequencing (Reichwald et al., 2009) properly mapped to the genome (Figures 2A, S1B, S1C and S1D). Thus, the reference turquoise killifish genome, while fragmented, contains most genes and is well assembled.

We next annotated the reference turquoise killifish genome. Using de novo gene prediction and sequence homology to 19 animal species, we identified 28,494 protein-coding gene models (Figure 2B, S1G). There was maximal enrichment of RNA-seq reads over protein-coding gene bodies (Figure S1E) and maximal enrichment of trimethylated lysine 4 on histone H3 (H3K4me3), a chromatin mark associated with promoters, at predicted transcription start sites (Figure S1F). The majority of the predicted protein-coding genes in the turquoise killifish have an ortholog in other vertebrate genomes (Table S2A). There was a good paralog correspondence and large blocks of syntenic genes between turquoise

killifish and medaka, a fish with a high-quality Sanger-sequenced genome (Figure S2A–S2D). Together, these results indicate that protein-coding genes, including paralogs, are well annotated in the turquoise killifish assembly, although some split genes may be misannotated as paralogs.

We also identified 5,859 high-confidence long non-coding (lnc) RNA genes and predicted other classes of non-coding RNA genes, including miRNAs (Figures 2B, S2E and S2F, and Table S1A). Finally, we identified several families of transposable elements (Figure 2C and Table S1C), at least some of which were actively transcribed (Tables S1D,E). To facilitate the use of the genome by the community, we have deployed a fully functional genome browser website (http://africanturquoisekillifishbrowser.org) (Figure 2D).

Evolutionary analysis of the African turquoise killifish genome

To gain insight into unique features of the African turquoise killifish, we analyzed its genome from an evolutionary perspective. The reconstituted phylogeny based on high-confidence one-to-one orthologs in 20 species (Tables S2B,C, Figure 3A) is consistent with fossil records and previous estimates (Reichwald et al., 2009), confirming the quality of the turquoise killifish genome assembly and annotation. To understand the evolution of features unique to this short-lived fish, we identified sites under positive selection in protein-coding genes in the turquoise killifish compared to seven longer-lived fish species (Figure 3A). We used a maximum likelihood approach to determine positive selection at the codon level (Yang, 2007) (Figures 3B and S3A). After multiple hypothesis correction, we identified 497 genes with at least one site under positive selection in the turquoise killifish genome (Table S3A). Of these, 249 genes had one or more sites with very high probability of being under positive selection, and represent the highest-confidence list of genes under positive selection (Table S3B). In the remainder of the manuscript, we used this highest-confidence list, unless otherwise indicated.

A Gene Ontology enrichment analysis revealed that the genes under positive selection in the turquoise killifish are enriched for components of signal transduction pathways, metabolism, development, proteostasis, and immunity (Figure 3C and Table S3C). Interestingly, several of these processes are relevant to the modulation of aging (Kennedy et al., 2014; Lopez-Otin et al., 2013), although these categories are broad and could underlie other biological processes. Positively selected genes include the helicase *RTEL1*, which is involved in telomere elongation and has been associated with dyskeratosis congenita, a human syndrome with premature aging characteristics (Ballew et al., 2013). Another intriguing positively selected gene is *PSMD11*, which is involved in proteostasis and whose ortholog is important for lifespan in *C. elegans* (Vilchez et al., 2012). Using in silico prediction tools originally designed for human disease variants, we found that 340 of the 2,009 positively selected residues (corresponding to 121 proteins) had predicted functional effects on the protein by two independent methods (Figures 3D, S3B, and Table S3D). The genes under positive selection in the turquoise killifish tend to display gene expression changes throughout life ($p = 0.045$ in Fisher's exact test at FDR 5%; Figure S3D, S3E and Table S3E). Thus, a subset of the positively selected genes in the turquoise killifish could be important for the evolution of a naturally short life trajectory. However, these genes could

also underlie the evolution of other traits in this fish (e.g. diapause, resistance to high temperature, morphology, or sensitivity to microbial communities or pathogens).

Aging and longevity genes are under positive selection in the turquoise killifish

To focus on genes with potential roles in the turquoise killifish compressed lifespan trajectory, we intersected the set of 497 genes under positive selection in the turquoise killifish with known aging-related genes from GenAge (mouse and human combined) and LongevityMap (Table S4A) (Budovsky et al., 2013; de Magalhaes et al., 2009). While the overlap was not statistically significant, 22 previously known aging-related genes were under positive selection in the turquoise killifish (Tables S4B,S4C). These genes include the insulin receptor A (*INSRA*, Figure 4A) and *IGF1* receptor (*IGF1R(1of2*), also known as *IGF1RA*, Figure 4B). Reduced function in the insulin/IGF1 receptors extends lifespan from worms to mice (Bluher et al., 2003; Holzenberger et al., 2003; Kenyon et al., 1993), and variants in IGF1 receptor are associated with exceptional longevity in humans (Suh et al., 2008) (Figure 4C). Aging genes under positive selection also include *LMNA3* (*LMNA(3of3)*, Figure 4E), which encodes a nuclear Lamin-A/C. Mutations in human *LMNA* can give rise to Hutchinson-Gilford Progeria Syndrome (Eriksson et al., 2003), although other variants are associated with exceptional longevity in humans (Conneely et al., 2012) (Figure 4C). Finally, another gene under positive selection is the DNA repair gene *XRCC5* (also known as *KU80*, Figure S4A). *XRCC5* deficiency in mice leads to premature aging (Vogel et al., 1999), though specific variants are also associated with human longevity (Figure 4C) (Soerensen et al., 2012).

We next mapped the variants under positive selection on the protein structures or domains of several aging-related proteins, highlighting variants with predicted functional effects (Tables S4E–G). The majority of residues with functional effect under positive selection in the insulin/IGF1 receptors (INSRA and IGF1R1A) are located in the extracellular domains implicated in ligand binding (Figures 4A, 4B, and 4D). In LMNA3, the functional residues under positive selection are located in the filamentous domain involved in the interaction with chromatin (Figure 4E). Finally, in XRCC5, they are located in the DNA binding domain (Figure S4A). Thus, positively selected residues with predicted functional effects are located in important domains in aging-related proteins (see also Figure S4B–S4D), and might have consequences on the role of these proteins in lifespan regulation.

The ability to undergo diapause – a developmental stage associated to drought-resistance - is likely under intense selective pressure given the transient nature of the ponds in which these fish live. We asked if the 497 genes under positive selection in the turquoise killifish overlapped with genes involved in diapause in other species, such as the insulin/IGF1-FOXO and the TGFβ pathways (Gottlieb and Ruvkun, 1994; Patterson et al., 1997). Only the genes in the insulin/IGF1-FOXO pathway (*INSRA, IGF1RA, FOXO1B(2of2)*), which are also involved in lifespan, were positively selected in the turquoise killifish (Figure S4E and Table S4D). Thus, positively selected genes in the insulin/IGF1 pathway might play a role both in diapause and compressed life cycle in the turquoise killifish, perhaps depending on external conditions.

Comparison of aging genes in the turquoise killifish and other species or groups with exceptional longevity

Intriguingly, the genes under positive selection in the short-lived turquoise killifish are also under positive selection or uniquely changed in species with exceptional longevity (naked mole rat, Brandt's bat and bowhead whale). Indeed, *IGF1R(1of2)* was found to be uniquely changed in the long-lived Brandt's bat (Seim et al., 2013) (Figures 4C and 4D) and under positive selection in the short-lived marmoset (The Marmoset Genome Sequencing and Analysis Consortium, 2014). More generally, 11 other genes are under positive selection or uniquely changed in both turquoise killifish and "extreme longevity" species or groups of individuals (Figure 4C). These genes include a carboxyl ester lipase *CEL(7of7)*, which is involved in cholesterol metabolism and diabetes in humans (Raeder et al., 2006), and the complement system component *C3(3of3)*, which is implicated in age-related degenerative pathologies and Alzheimer's disease (Proitsi et al., 2012). These observations raise the intriguing possibility that the same genes can be under positive selection in both extremely short-lived and long-lived species.

Are the residues in proteins that are positively selected in short-lived and long-lived species similar or different? We mapped residues from the short-lived turquoise killifish, long-lived Brandt's bat, and humans onto the well-studied IGF1 receptor (Figure 4D) and LMNA (Figure 4E). Many of the residues under positive selection in the turquoise killifish and the Brandt's bat are in proximity on the IGF1 receptor sequence, but differ (Figure 4D). Furthermore, the residues under positive selection in the turquoise killifish and those associated with longevity in human are both located in the predicted IGF1R ligand-binding domains, but are different (Figure 4D). These residues also differ from *C. elegans* longevity mutations in the insulin/IGF1 receptor (DAF-2, Figure S4E). Similarly, the LMNA3 residues under positive selection in the turquoise killifish also differ from variants in human centenarians or Hutchinson Gilford Progeria Syndrome (Figure 4E). More generally, for the same protein, the residues under selection in the turquoise killifish differ from those uniquely changed in the long-lived bowhead whale (Table S4G, and mapping for *CEL(7of7)* in Figure S4F). Thus, proteins that act as central nodes could have been selected to underlie both compressed and extended life trajectories, depending on the residues. Alternatively, the same proteins could have been selected because both the turquoise killifish and long-lived species exhibit resistance to stress – during diapause for the turquoise killifish and throughout life for long-lived species.

Sequencing individuals from additional turquoise killifish strains reveals variants in aging-related genes

Within the turquoise killifish species, there exist several strains with reported differences in lifespan in specific laboratory environments (Kirschner et al., 2012; Terzibasi et al., 2008) (Figure 5A, S5A, and 6B), and these differences could be leveraged to understand the genetic architecture of lifespan. To assess the genetic differences among turquoise killifish strains, we sequenced at lower coverage individuals from two additional strains that were captured in Mozambique in 2004 and 2007 (MZM-0403 and MZM-0703, respectively) and from a control GRZ individual (Figure 5A). This analysis uncovered over three million single nucleotide polymorphisms (SNPs) between MZM-0403 or MZM-0703 and the

reference GRZ genome (Figures 5A, 5B, and S5B). As expected, there were fewer SNPs between the GRZ individual and the reference GRZ genome (Figures 5A and S5B).

To identify potential genetic differences associated with phenotypic diversity (e.g. lifespan, color, etc.) among strains, we focused on SNPs that are shared between the longer-lived red-tailed MZM strains, but not by the shorter-lived yellow-tailed GRZ reference strain (Figure 5C). We identified 22,389 non-synonymous SNPs in 10,638 genes, and 139 of these genes overlapped with known aging-related genes that encompassed all 9 "hallmarks of aging" (Lopez-Otin et al., 2013) (Figure 5D and Table S5A). A number of these SNPs are predicted to have functional impact on the protein (Figure 5D and Table S5B). Genes with functional variants between turquoise killifish strains include insulin/IGF signaling pathway genes (*GHR(1of4)* and *FOXO4* transcription factor) and inter-cellular communication genes, such as progranulin (*GRN*, also known as *PGRN*) (Figure 5D). While this analysis does not identify the specific variants that explain lifespan differences between strains, it provides a resource for the study of genetic variation in this species.

Identification of a genomic region significantly associated with differences in lifespan between turquoise killifish strains

To identify genomic regions that are important for lifespan differences between turquoise killifish strains, we performed a genetic linkage analysis in a cross between the shorter-lived and longer-lived strains (Figure 6A). We used the shorter-lived reference strain captured in Zimbabwe in 1968 (GRZ) and longer-lived strains derived from expeditions in Mozambique in 2007 and 2006 (MZM-0703 and Soveia) (Figure S5A). These strains exhibit differences in lifespan in the laboratory conditions used in this study (Figure 6B and Table S6A), although the number of animals was low for the Soveia strain and differences between strains could have been accentuated by the laboratory conditions used. We crossed a female from the GRZ strain to a male from the MZM-0703 strain ("cross GxM", Figure 6A), as well as a female from the Soveia strain with a male from the GRZ strain ("cross SxG", Table S6A). F1 fish were interbred and the lifespan of as many F2 fish as possible (430 for cross GxM, and 130 for cross SxG) was scored. F1 and F2 fish live longer than the short-lived GRZ parental strain in both crosses (Figure 6B, and Table S6A), suggesting that the long lifespan trait is dominant over the short lifespan trait. Similar lifespan results were obtained with cross GxM (in which the female is from the shorter-lived strain) and cross SxG (in which the male is from the shorter-lived strains), suggesting that there is no major maternal contribution to lifespan. Surprisingly, while males and females had a similar lifespan in the parental strains, males were significantly longer-lived than females in the F1 and F2 generation of cross GxM (Figure S6A–S6D), implying a possible interplay between lifespan and sex.

To identify the genomic regions significantly associated with lifespan differences, we conducted a genome-wide linkage analysis using cross GxM because of the higher number of F2 individuals in this cross. Using restriction-site associated DNA sequencing (RAD-seq) to genotype P0 grandparents (2), F1 fish (16), and a large subset of F2 fish (207), we identified 8,399 high-confidence genomic markers that displayed SNPs between the grandparents. These markers allowed us to build a high-resolution linkage map with 19

linkage groups (LGs), providing an anchor to our reference genome (Figure 6C). Importantly, the number of LGs is the same as the haploid number of chromosomes in this species (19) (Reichwald et al., 2009).

We then mapped individual lifespan as a phenotype using a method based on Random Forest. The most significant genetic region associated with lifespan is on linkage group 3 (LG-3) (Figure 6C). This region was found to be associated with lifespan differences by two independent statistical methods, Random Forest and Log-Rank test, and it reached genome-wide significance in the Random Forest QTL analysis (FDR < 5%). Importantly, the lifespan QTL on LG-3 was 'non-transgressive', i.e. the individuals with the longer-lived grandparents genotype at this locus (Figure 6D, red) were the longest-lived. The fact that individuals with a heterozygous genotype at this locus do not live longer than individuals with either grandparent genotype (Figure 6D) argues against hybrid vigor, consistent with the absence of global enrichment in heterozygous fish in longest-lived F2 individuals (Figure S6E). Individuals with one or both alleles from the longer-lived grandparent at the QTL peak marker exhibited a ~30% increase in lifespan (p=1.6x10^{-3} and p=1.8x10^{-3}, respectively, in Log-Rank tests) compared to individuals with both alleles from the shorter-lived grandparent (Figures 6E and S6F). Thus, this genetic region on LG-3 has a significant contribution to the lifespan difference in the F2 generation of this cross.

The genomic region associated with lifespan differences between turquoise killifish strains is on the sex chromosome

Interestingly, the genetic region associated with differences in lifespan among turquoise killifish strains is close to the sex-determination region (Figure 7A). A previous study identified a lifespan QTL in the linkage group linked to sex but at the time this QTL could not be resolved from the sex-determination region (Kirschner et al., 2012). Using our high-density linkage map, we found that the lifespan QTL and the sex-determining region are on the same linkage group, but 36–38cM apart (Figure 7B, upper panel), although distance estimates in a sex-linked region cannot be fully accurate due to suppressed recombination near the sex-determining region. Individuals with both alleles from the long-lived grandparent at the lifespan QTL peak marker exhibited significant increase in lifespan (Figure 7B, bottom panel). The lifespan QTL also remained significant even after regressing for sex (Figures 6D and 7C). Importantly, the marker with the highest significance for lifespan was associated with longevity in both males and females (Figure S6F), ruling out the possibility that significance is due to a potential bias coming from the males. Together, these results strongly support that the lifespan QTL and the sex-determining region are linked, but distinct.

Genes in the region associated with lifespan differences

We next used the turquoise killifish genome to identify the specific genes in the region associated with lifespan differences between strains. The lifespan QTL region captures genomic scaffolds encompassing 31 protein-coding genes, 6 long non-coding RNA genes, and 2 small nucleolar RNA genes (snoRNA) (Figure 7C and Tables S7A–E). Among the 31 protein-coding genes, 7 had already been linked to the regulation of aging or lifespan in humans or model organisms (Figure 7C and Table S7A). These include the gene encoding

progranulin (*GRN*), which has been implicated in neurodegenerative diseases (Wang et al., 2010) and lifespan regulation in mice (Ahmed et al., 2010) (Figure 7C). Another interesting candidate is *NUDT1*, which is involved in the hydrolysis of 8-oxo-dGTP, a deleterious nucleoside that increases with aging in mitochondrial DNA (Souza-Pinto et al., 1999) (Figure 7C). *NUDT1* overexpression extends lifespan in mice (De Luca et al., 2013). Yet another intriguing candidate is *GSTT1A*, which encodes a gluthathione S-transferase, a class of redox homeostasis enzymes that regulate lifespan in worms and mice (Ayyadevara et al., 2007; Pesch et al., 2004) (Figure 7C). Finally, this region comprises genes encoding two transcription factors (*STAT3, STAT5.1(2of6)*) that have been implicated in regulating 'inflammaging' (De-Fraja et al., 2000) (Figure 7C).

Of the 31 genes underlying the lifespan QTL, 15 harbored non-synonymous coding differences between the P0 individuals (Figures 7C, S7A, and Table S7A). For six of these 15 genes (*ATXN7L1, GRN, HIPK2(11of26), IFI35, TTYH3A* and *ZNF800A*), the coding differences occur in otherwise well-conserved residues, are also observed in the resequenced MZM-0403 individual, and the variant in the longer-lived P0 corresponds to the consensus from other species (Figure S7B–S7D; Table S7A). The presence of several of these variants in cross GxM P0 founders was confirmed by Sanger sequencing (Table S7E). Interestingly, one of the variants in the turquoise killifish GRN (W449 in the shorter-lived strain and C449 in the longer-lived strain) is within a motif that plays a key role in protein folding and that is mutated in frontotemporal dementia (FTD) (Wang et al., 2010) (Figure 7D). Consistently, this turquoise killifish variant is predicted to have functional consequences (Figure 7D and Tables S7G,H). This variant is also found in wild fish captured during our expedition to Mozambique and Zimbabwe in 2010 (Figure S7E and Table S7F), indicating that it is not a spurious mutation that arose in the laboratory or from the bottleneck of a rare allele. Thus, coding variations in GRN or other candidates may underlie differences in lifespan between strains of this species, although we cannot exclude that non-coding variants are responsible for these phenotypic difference (Figure S7A).

We wondered if the number of aging genes within the lifespan QTL was higher than expected by chance. Statistical analysis showed that the lifespan QTL region is significantly enriched for known aging-related genes (GenAge, mouse and human combined, Table S4A) compared to the rest of LG-3 ($p = 6.4 \times 10^{-4}$, Fisher's exact test) or to all linkage groups ($p = 2.1 \times 10^{-4}$, Fisher's exact test) (Figure 7E). This significant enrichment for known aging and longevity genes in the region underlying the lifespan QTL suggests that a haplotype block containing a cluster of genes, rather than a single gene, might be involved in the observed lifespan difference. Sex-determining regions are usually regions of suppressed recombination, and indeed the recombination frequencies are the lowest in the region associated with sex determination (Figure S7F and S7G). Intriguingly, the recombination frequencies are lower than expected throughout the entire LG-3, including the lifespan QTL region (Figures 7E, S7F, and S7G), compared to the rest of the genome. This haplotype block might have formed because of the suppressed recombination in this region, due to its proximity with the sex-determining region (Figure 7E). While this may be fortuitous, the presence of the lifespan-determining region on the sex chromosome might have evolved to couple strategies of fast reproduction with genes involved in overall fitness.

Discussion

The genome of one of the shortest-lived vertebrate species: a resource for comparative, experimental, and evolutionary genomics

The African turquoise killifish reference genome and transcriptome, high-density genetic linkage map, and a comprehensive genome browser represent great resources for the exploration of the genetic principles and the evolution underlying several unique traits in this species, such as a compressed life cycle and embryonic diapause. These resources, combined with the rapid experimental capacity of this fish, provide an unprecedented setting for evolution and genetic studies in vertebrates.

Evolution of the lifespan differences between species

The forces that shape differences in lifespan in nature are still largely unknown. Extrinsic mortality imposes a strong constraint for rapid reproduction, which may ultimately result in the evolution of shorter adult lifespan (Chen and Maklakov, 2012). The complete desiccation of the turquoise killifish ponds during the dry season represents a potent extrinsic mortality constraint that may have resulted in the evolution of accelerated sexual maturation. Short lifespan might have evolved as the pleiotropic byproduct of rapid sexual maturation or could have resulted from relaxed negative selection on several genes. The abundance of transposable elements in the turquoise killifish genome might have played a role in the evolution of its compressed lifespan.

Intriguingly, genes implicated in insulin/IGF signaling and genome maintenance are under positive selection or uniquely changed in exceptionally long-lived mammals such as the naked mole rat (Kim et al., 2011), Brandt's bat (Seim et al., 2013), and bowhead whale (Keane et al., 2015). *IGF1R* has also been shown to be under positive selection in the genome of a short-lived primate, the marmoset (The Marmoset Genome Sequencing and Analysis Consortium, 2014), raising the exciting possibility that the same gene may be modified to evolve compressed or extended life trajectories under different environmental constraints. IGF1R, which controls organismal growth, sexual maturity, fertility, and fitness, may be particularly 'tunable' for the evolution of different life strategies.

Genetic architecture of lifespan between turquoise killifish strains

A region associated to lifespan differences between turquoise killifish strains contains a cluster of known aging and longevity genes. The gene encoding progranulin (*GRN*) is particularly interesting as it influences lifespan in mice (Ahmed et al., 2010) and contributes to age-related diseases in humans, in particular frontotemporal dementia (Baker et al., 2006). It will be important to determine the functional consequences of the turquoise killifish GRN variants, for example in stress response (Judy et al., 2013). Variants in non-coding genes or regulatory regions, such as enhancers, may also mediate the observed lifespan differences between fish strains. Finally, the differences in lifespan between turquoise killifish strains may be accentuated in the specific laboratory setting used in this study, and the genetic region identified may represent the interaction between genetic determinants in this particular environment.

A cluster of aging and longevity genes linked to the sex-determining region

The most important region associated to differences in lifespan between turquoise killifish strains is on the sex chromosome, consistent with a previous finding (Kirschner et al., 2012). Our high-resolution genetic linkage map enabled us to show that the lifespan QTL is linked to, but distinct from the sex-determining region, within a region of suppressed recombination. The association between the sex-determining region and the lifespan QTL raises the intriguing possibility that these regions might have co-evolved. Alternatively, conditions of partially suppressed recombination might have allowed this trait to hitchhike on the sex-determining region without any direct fitness benefit.

In conclusion, our analyses reveal a possible association between lifespan and sex determination in the turquoise killifish and identify several potential candidates that may underlie lifespan differences across taxa. The African turquoise killifish genomic and genetic resources should help further explore the genetic basis of longevity and the evolutionary forces that shape vertebrate life history traits in nature.

Experimental Procedures

Additional details are provided in the Extended Experimental Procedures.

Assembly of the turquoise killifish genome and identification of genetic variants between different strains

Genomic DNA was isolated from African turquoise killifish individuals from the inbred strain GRZ, and 10 libraries with varying insert sizes were constructed for Illumina sequencing on HiSeq2000 instruments. Assembly was performed using SGA (Simpson and Durbin, 2012) and SOAPdenovo (Luo et al., 2012), scaffolding using SSPACE Basic (Boetzer et al., 2011) and Gap-filling using GapFiller (Nadalin et al., 2012). The SGA and SOAP genome assemblies were reconciled using GARM (Soto-Jimenez et al., 2014). Higher-order scaffolding was performed using raw reads from Illumina paired-end RNA-seq libraries and the linkage map from cross GxM (see below). To identify genetic variants among different strains of the turquoise killifish, we resequenced the founders of cross GxM (GRZ female and MZM-0703 male) and a male individual from another wild-derived strain (MZM-0403). We used the GATK genotyping pipeline (McKenna et al., 2010) to call variants between turquoise killifish strains. The turquoise killifish genome browser is at http://africanturquoisekillifishbrowser.org.

De novo prediction and annotation of protein coding gene models

The MAKER2 pipeline was used to generate putative protein coding gene predictions (Holt and Yandell, 2011) supported by two ab initio gene prediction approaches. To annotate protein-coding genes from predictions, a homology-based approach using 19 other genomes (Table S2B) was implemented, leading to a final set of 28,494 turquoise killifish putative protein coding genes.

Identification of turquoise killifish genes under positive selection and prediction of their functional impact

The genes under positive selection were identified using a branch-site model implemented in PAML (Yang, 2007). Single ortholog protein families were identified using the 8 teleost fish genomes in our analysis. Proteins from each family were aligned using PRANK (Loytynoja and Goldman, 2005) and the resulting alignments were filtered in GUIDANCE (Penn et al., 2010). CODEML was then used to predict the genes and individual sites under positive selection in the turquoise killifish lineage after stringent filtering and multiple hypothesis correction. Functional impact for the residues under positive selection or the non-synonymous variants between killifish strains was assessed using two prediction algorithms: PROVEAN (Choi et al., 2012) and SIFT (Kumar et al., 2009).

Linkage map and QTL mapping

One female from the GRZ strain was crossed with one male from the MZM-0703 strain (cross GxM), and one female from the Soveia strain was crossed with a male from the GRZ strain (cross SxG). F1 fish were interbred in families to generate F2 fish and lifespan was scored as the age at death for all individuals (raised in cohorts of mixed sexes). RAD-seq libraries for 225 samples from the cross GxM and for 86 from the SxG cross were prepared as described (Etter et al., 2011). This resulted in 8,399 polymorphic markers for cross GxM. We built a linkage map for cross GxM with R/qtl using RAD-seq markers that had homozygous haplotypes in the grandparents. After stringent filtering of markers and individuals, the final linkage map comprised 193 F2 individuals and 5,757 RAD-seq markers. QTL detection was conducted using a Random Forest-based method (Clement-Ziza et al., 2014), which uses genetic markers as predictors to model the traits and where population structure is modelled as covariates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Ahmed Z, Sheng H, Xu YF, Lin WL, Innes AE, Gass J, Yu X, Wuertzer CA, Hou H, Chiba S, et al. Accelerated lipofuscinosis and ubiquitination in granulin knockout mice suggest a role for progranulin in successful aging. Am. J. Pathol. 2010; 177:311–324. [PubMed: 20522652]

Austad SN. Methusaleh's Zoo: how nature provides us with clues for extending human health span. J. Comp. Pathol. 2010; 142(Suppl 1):S10–S21. [PubMed: 19962715]

Ayyadevara S, Dandapat A, Singh SP, Siegel ER, Shmookler Reis RJ, Zimniak L, Zimniak P. Life span and stress resistance of *Caenorhabditis elegans* are differentially affected by glutathione transferases metabolizing 4-hydroxynon-2-enal. Mech. Ageing Dev. 2007; 128:196–205. [PubMed: 17157356]

Baker M, Mackenzie IR, Pickering-Brown SM, Gass J, Rademakers R, Lindholm C, Snowden J, Adamson J, Sadovnick AD, Rollinson S, et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature. 2006; 442:916–919. [PubMed: 16862116]

Ballew BJ, Yeager M, Jacobs K, Giri N, Boland J, Burdett L, Alter BP, Savage SA. Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in Dyskeratosis congenita. Hum. Genet. 2013; 132:473–480. [PubMed: 23329068]

Bluher M, Kahn BB, Kahn CR. Extended longevity in mice lacking the insulin receptor in adipose tissue. Science. 2003; 299:572–574. [PubMed: 12543978]

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011; 27:578–579. [PubMed: 21149342]

Budovsky A, Craig T, Wang J, Tacutu R, Csordas A, Lourenco J, Fraifeld VE, de Magalhaes JP. LongevityMap: a database of human genetic variants associated with longevity. Trends Genet. 2013; 29:559–560. [PubMed: 23998809]

Cellerino A, Valenzano DR, Reichard M. From the bush to the bench: the annual *Nothobranchius* fishes as a new model system in biology. Biol. Rev. Camb. Philos. Soc. 2015

Chen HY, Maklakov AA. Longer life span evolves under high rates of condition-dependent mortality. Curr. Biol. 2012; 22:2140–2143. [PubMed: 23084993]

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012; 7:e46688. [PubMed: 23056405]

Clement-Ziza M, Marsellach FX, Codlin S, Papadakis MA, Reinhardt S, Rodriguez-Lopez M, Martin S, Marguerat S, Schmidt A, Lee E, et al. Natural genetic variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. Mol. Syst. Biol. 2014; 10:764. [PubMed: 25432776]

Conneely KN, Capell BC, Erdos MR, Sebastiani P, Solovieff N, Swift AJ, Baldwin CT, Budagov T, Barzilai N, Atzmon G, et al. Human longevity and common variations in the LMNA gene: a meta-analysis. Aging Cell. 2012; 11:475–481. [PubMed: 22340368]

De Luca G, Ventura I, Sanghez V, Russo MT, Ajmone-Cat MA, Cacci E, Martire A, Popoli P, Falcone G, Michelini F, et al. Prolonged lifespan with enhanced exploratory behavior in mice overexpressing the oxidized nucleoside triphosphatase hMTH1. Aging Cell. 2013; 12:695–705. [PubMed: 23648059]

de Magalhaes JP, Budovsky A, Lehmann G, Costa J, Li Y, Fraifeld V, Church GM. The Human Ageing Genomic Resources: online databases and tools for biogerontologists. Aging Cell. 2009; 8:65–72. [PubMed: 18986374]

De-Fraja C, Conti L, Govoni S, Battaini F, Cattaneo E. STAT signalling in the mature and aging brain. Int. J. Dev. Neurosci. 2000; 18:439–446. [PubMed: 10817928]

Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P, et al. Recurrent *de novo* point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. Nature. 2003; 423:293–298. [PubMed: 12714972]

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol. Biol. 2011; 772:157–178. [PubMed: 22065437]

Flachsbart F, Caliebe A, Kleindorp R, Blanche H, von Eller-Eberstein H, Nikolaus S, Schreiber S, Nebel A. Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc. Natl. Acad. Sci. USA. 2009; 106:2700–2705. [PubMed: 19196970]

Gottlieb S, Ruvkun G. *daf-2, daf-16* and *daf-23*: genetically interacting genes controlling Dauer formation in *Caenorhabditis elegans* . Genetics. 1994; 137:107–120. [PubMed: 8056303]

Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011; 12:491. [PubMed: 22192575]

Holzenberger M, Dupont J, Ducos B, Leneuve P, Geloen A, Even PC, Cervera P, Le Bouc Y. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature. 2003; 421:182–187. [PubMed: 12483226]

Johnson SC, Rabinovitch PS, Kaeberlein M. mTOR is a key modulator of ageing and age-related disease. Nature. 2013; 493:338–345. [PubMed: 23325216]

Judy ME, Nakamura A, Huang A, Grant H, McCurdy H, Weiberth KF, Gao F, Coppola G, Kenyon C, Kao AW. A shift to organismal stress resistance in programmed cell death mutants. PLoS Genet. 2013; 9:e1003714. [PubMed: 24068943]

Kaeberlein M, Kennedy BK. Hot topics in aging research: protein translation and TOR signaling, 2010. Aging Cell. 2011; 10:185–190. [PubMed: 21176090]

Kapahi P, Chen D, Rogers AN, Katewa SD, Li PW, Thomas EL, Kockel L. With TOR, less is more: a key role for the conserved nutrient-sensing TOR pathway in aging. Cell Metab. 2010; 11:453–465. [PubMed: 20519118]

Keane M, Semeiks J, Webb AE, Li YI, Quesada V, Craig T, Madsen LB, van Dam S, Brawand D, Marques PI, et al. Insights into the evolution of longevity from the bowhead whale genome. Cell Rep. 2015; 10:112–122. [PubMed: 25565328]

Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, Franceschi C, Lithgow GJ, Morimoto RI, Pessin JE, et al. Geroscience: linking aging to chronic disease. Cell. 2014; 159:709–713. [PubMed: 25417146]

Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R. A *C. elegans* mutant that lives twice as long as wild type. Nature. 1993; 366:461–464. [PubMed: 8247153]

Kenyon CJ. The genetics of ageing. Nature. 2010; 464:504–512. [PubMed: 20336132]

Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature. 2011; 479:223–227. [PubMed: 21993625]

Kirschner J, Weber D, Neuschl C, Franke A, Bottger M, Zielke L, Powalsky E, Groth M, Shagin D, Petzold A, et al. Mapping of quantitative trait loci controlling lifespan in the short-lived fish *Nothobranchius furzeri* - a new vertebrate model for age research. Aging Cell. 2012; 11:252–261. [PubMed: 22221414]

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 2009; 4:1073–1081. [PubMed: 19561590]

Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013; 153:1194–1217. [PubMed: 23746838]

Loytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. USA. 2005; 102:10557–10562. [PubMed: 16000407]

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012; 1(1):18. [PubMed: 23587118]

The Marmoset Genome Sequencing and Analysis Consortium. The common marmoset genome provides insight into primate biology and evolution. Nat. Genet. 2014; 46:850–857. [PubMed: 25038751]

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]

Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics. 2012; 13(Suppl 14):S8. [PubMed: 23095524]

Patterson GI, Koweek A, Wong A, Liu Y, Ruvkun G. The DAF-3 Smad protein antagonizes TGF-beta-related receptor signaling in the *Caenorhabditis elegans* dauer pathway. Genes Dev. 1997; 11:2679–2690. [PubMed: 9334330]

Penn O, Privman E, Landan G, Graur D, Pupko T. An alignment confidence score capturing robustness to guide tree uncertainty. Mol. Biol. Evol. 2010; 27:1759–1767. [PubMed: 20207713]

Pesch B, Dusing R, Rabstein S, Harth V, Grentrup D, Bruning T, Landt O, Vetter H, Ko YD. Polymorphic metabolic susceptibility genes and longevity: a study in octogonarians. Toxicol. Lett. 2004; 151:283–290. [PubMed: 15177664]

Proitsi P, Lupton MK, Dudbridge F, Tsolaki M, Hamilton G, Daniilidou M, Pritchard M, Lord K, Martin BM, Johnson J, et al. Alzheimer's disease and age-related macular degeneration have different genetic models for complement gene variation. Neurobiol. Aging. 2012; 33:1843. e1849-1817. [PubMed: 22300950]

Raeder H, Johansson S, Holm PI, Haldorsen IS, Mas E, Sbarra V, Nermoen I, Eide SA, Grevle L, Bjorkhaug L, et al. Mutations in the *CEL* VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. Nat. Genet. 2006; 38:54–62. [PubMed: 16369531]

Reichwald K, Lauber C, Nanda I, Kirschner J, Hartmann N, Schories S, Gausmann U, Taudien S, Schilhabel MB, Szafranski K, et al. High tandem repeat content in the genome of the short-lived annual fish *Nothobranchius furzeri*: a new vertebrate model for aging research. Genome Biol. 2009; 10 R16.

Seim I, Fang X, Xiong Z, Lobanov AV, Huang Z, Ma S, Feng Y, Turanov AA, Zhu Y, Lenz TL, et al. Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii* . Nat. Commun. 2013; 4:2212. [PubMed: 23962925]

Simpson JT, Durbin R. Efficient de novo assembly of large genomes using compressed data structures. Genome Res. 2012; 22:549–556. [PubMed: 22156294]

Soerensen M, Dato S, Tan Q, Thinggaard M, Kleindorp R, Beekman M, Jacobsen R, Suchiman HE, de Craen AJ, Westendorp RG, et al. Human longevity and variation in GH/IGF-1/insulin signaling, DNA damage signaling and repair and pro/antioxidant pathway genes: cross sectional and longitudinal studies. Exp. Gerontol. 2012; 47:379–387. [PubMed: 22406557]

Soto-Jimenez LM, Estrada K, Sanchez-Flores A. GARM: genome assembly, reconciliation and merging pipeline. Curr. Top. Med. Chem. 2014; 14:418–424. [PubMed: 24304318]

Souza-Pinto NC, Croteau DL, Hudson EK, Hansford RG, Bohr VA. Age-associated increase in 8-oxo-deoxyguanosine glycosylase/AP lyase activity in rat mitochondria. Nucleic Acids Res. 1999; 27:1935–1942. [PubMed: 10101204]

Suh Y, Atzmon G, Cho MO, Hwang D, Liu B, Leahy DJ, Barzilai N, Cohen P. Functionally significant insulin-like growth factor I receptor mutations in centenarians. Proc. Natl. Acad. Sci. USA. 2008; 105:3438–3442. [PubMed: 18316725]

Terzibasi E, Valenzano DR, Benedetti M, Roncaglia P, Cattaneo A, Domenici L, Cellerino A. Large differences in aging phenotype between strains of the short-lived annual fish *Nothobranchius furzeri* . PLoS One. 2008; 3:e3866. [PubMed: 19052641]

Valdesalici S, Cellerino A. Extremely short lifespan in the annual fish *Nothobranchius furzeri* . Proc. Biol. Sci. 2003; 270(Suppl 2):S189–S191. [PubMed: 14667379]

Valenzano DR, Kirschner J, Kamber RA, Zhang E, Weber D, Cellerino A, Englert C, Platzer M, Reichwald K, Brunet A. Mapping loci associated with tail color and sex determination in the short-lived fish *Nothobranchius furzeri* . Genetics. 2009; 183:1385–1395. [PubMed: 19786620]

Vilchez D, Morantte I, Liu Z, Douglas PM, Merkwirth C, Rodrigues AP, Manning G, Dillin A. RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. Nature. 2012; 489:263–268. [PubMed: 22922647]

Vogel H, Lim DS, Karsenty G, Finegold M, Hasty P. Deletion of Ku86 causes early onset of senescence in mice. Proc. Natl. Acad. Sci. USA. 1999; 96:10770–10775. [PubMed: 10485901]

Wang J, Van Damme P, Cruchaga C, Gitcho MA, Vidal JM, Seijo-Martinez M, Wang L, Wu JY, Robberecht W, Goate A. Pathogenic cysteine mutations affect progranulin function and production of mature granulins. J. Neurochem. 2010; 112:1305–1315. [PubMed: 20028451]

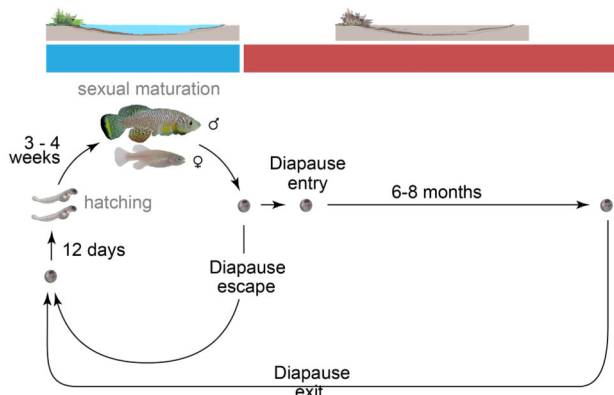Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 2007; 24:1586–1591. [PubMed: 17483113]
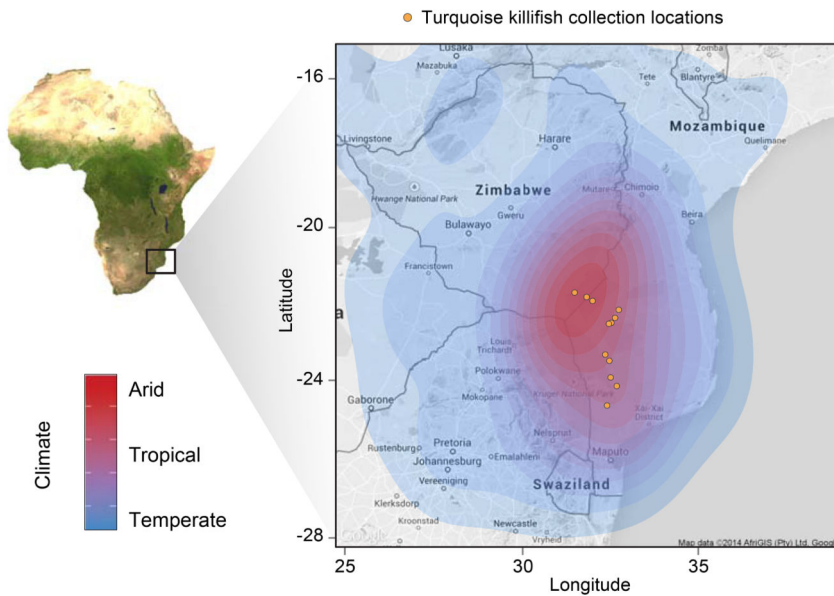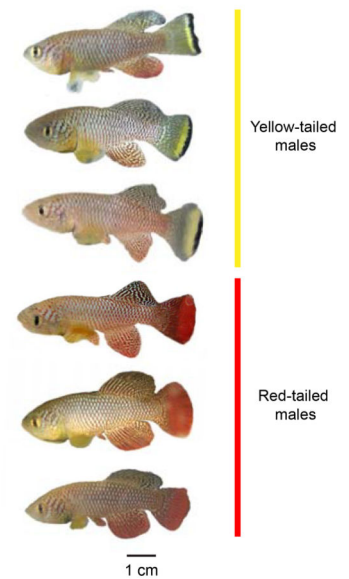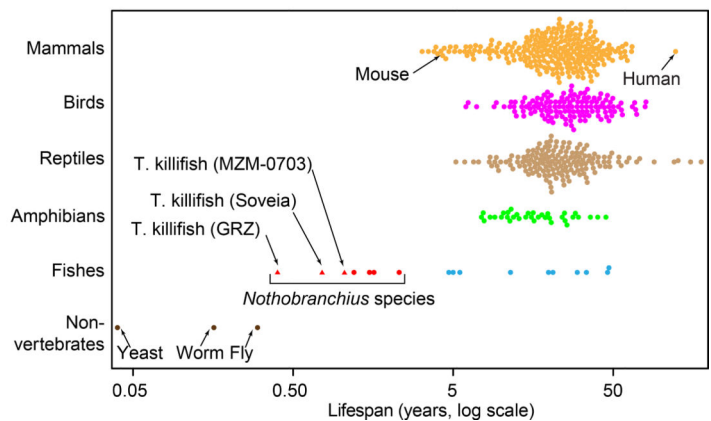
Figure 1. The African turquoise killifish is a naturally short-lived vertebrate with multiple strains

A) Geographical distribution of turquoise killifish in the wild (orange dots). The climate of the area measured by the Koeppen-Geiger index (concentric contours).

B) Examples of different morphological types in turquoise killifish males: yellow-tailed and red-tailed.

C) Life cycle of the turquoise killifish. The turquoise killifish achieves sexual maturation and reproduces during the wet season. Diapausing embryos can survive through the dry season, when the ponds are desiccated. Diapause can be skipped in the laboratory, resulting in a short generation time.

D) *Nothobranchius* species are among the shortest-lived vertebrates. Lifespan data for turquoise killifish strains are from our experimental data (age of the 10th percentile survivors) and are depicted by triangles. Lifespan of the strains can vary depending on

housing conditions. Maximum lifespan data for the other organisms are from the AnAge database.

Figure 2. De novo sequencing and assembly of the reference genome of the African turquoise killifish

A) Assembly statistics for draft genome (NotFur1) for the reference turquoise killifish strain (GRZ). Scaffolds that are not captured by the linkage map remain unplaced. CEGs: core eukaryotic genes. See also Figures 6C, S1B, S1C.
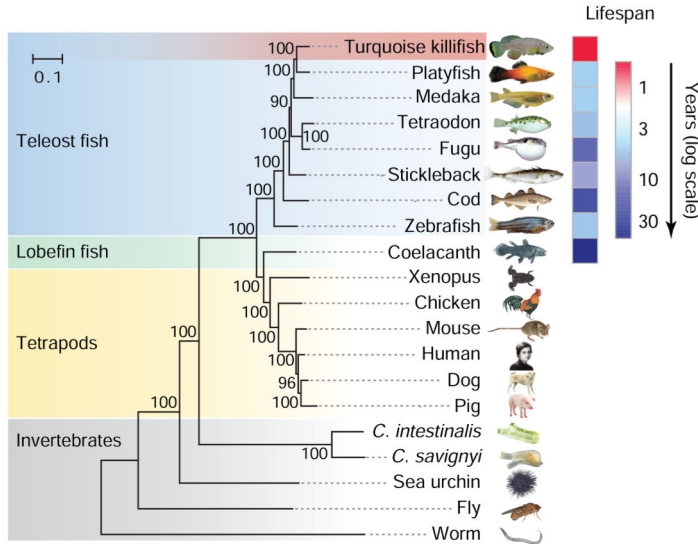
B) Number of annotated coding and non-coding genes in the reference turquoise killifish genome. High-confidence protein coding genes are genes with homologs in at least 10 species. lncRNA: long non-coding RNA; ncRNA: non-coding RNA; miRNA: microRNA. See also Figures S2E, S2F and Tables S1A,B.
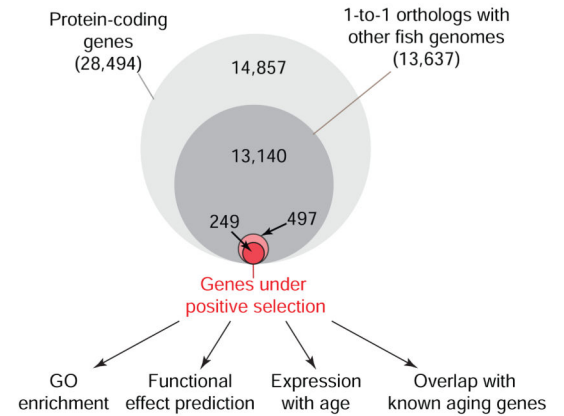
C) Repetitive element composition in the reference turquoise killifish genome. LINE: long interspersed nuclear element; SINE: short interspersed nuclear element; LTR: long terminal repeat. Type I transposons are RNA-mediated. Type II transposons are DNA-mediated. See also Tables S1C–E.

D) Example of a genomic region containing insulin-like growth factor 1 receptor (*IGF1RA*).
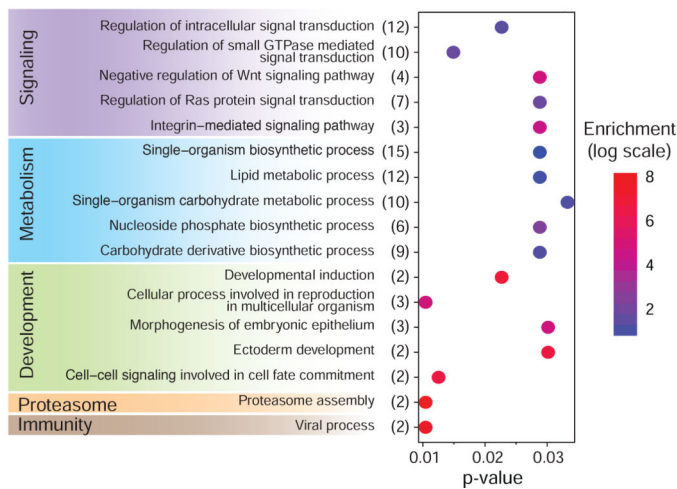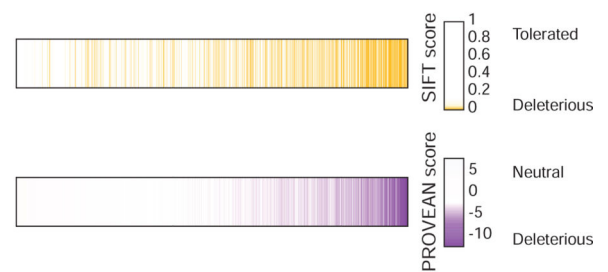
Figure 3. Evolutionary analysis of the turquoise killifish genome

A) Phylogenetic tree of 20 animal species, including the turquoise killifish, based on 619 one-to-one orthologs (Table S2C). Number on nodes: level of confidence (% bootstrap support). Scale bar: evolutionary distance (substitution per site). Maximum lifespan data are from our experimental data (turquoise killifish) or from the AnAge database (other fish species), and represented as a heat map.

B) Proportion and analysis of the genes under positive selection in the turquoise killifish compared to 7 other fish species after multiple hypothesis correction (FDR < 5%). See also Figure S3A.

C) Selected GO term enrichment for the genes under positive selection in the turquoise killifish. The number of genes associated with each category is indicated in brackets after the term description, and enrichment values are indicated in colored scale. See also Table S3C.

D) Predicted functional effect on the protein of residues under positive selection in the turquoise killifish have based on SIFT (top row) and PROVEAN (bottom row). Residues are ordered from left to right based on the rank-product of the SIFT and PROVEAN scores. Only sites scored by both methods are displayed. See also Figure S3B, Table S3D, and S4G.

A Residues under positive selection in INSRA

B Residues under positive selection in IGF1R(1of2)

C Aging and longevity candidates in turquoise killifish and other vertebrates

D Residues and variants in IGF1R(1of2) in turquoise killifish and other organisms

E Residues and variants in LMNA(3of3) in turquoise killifish and human
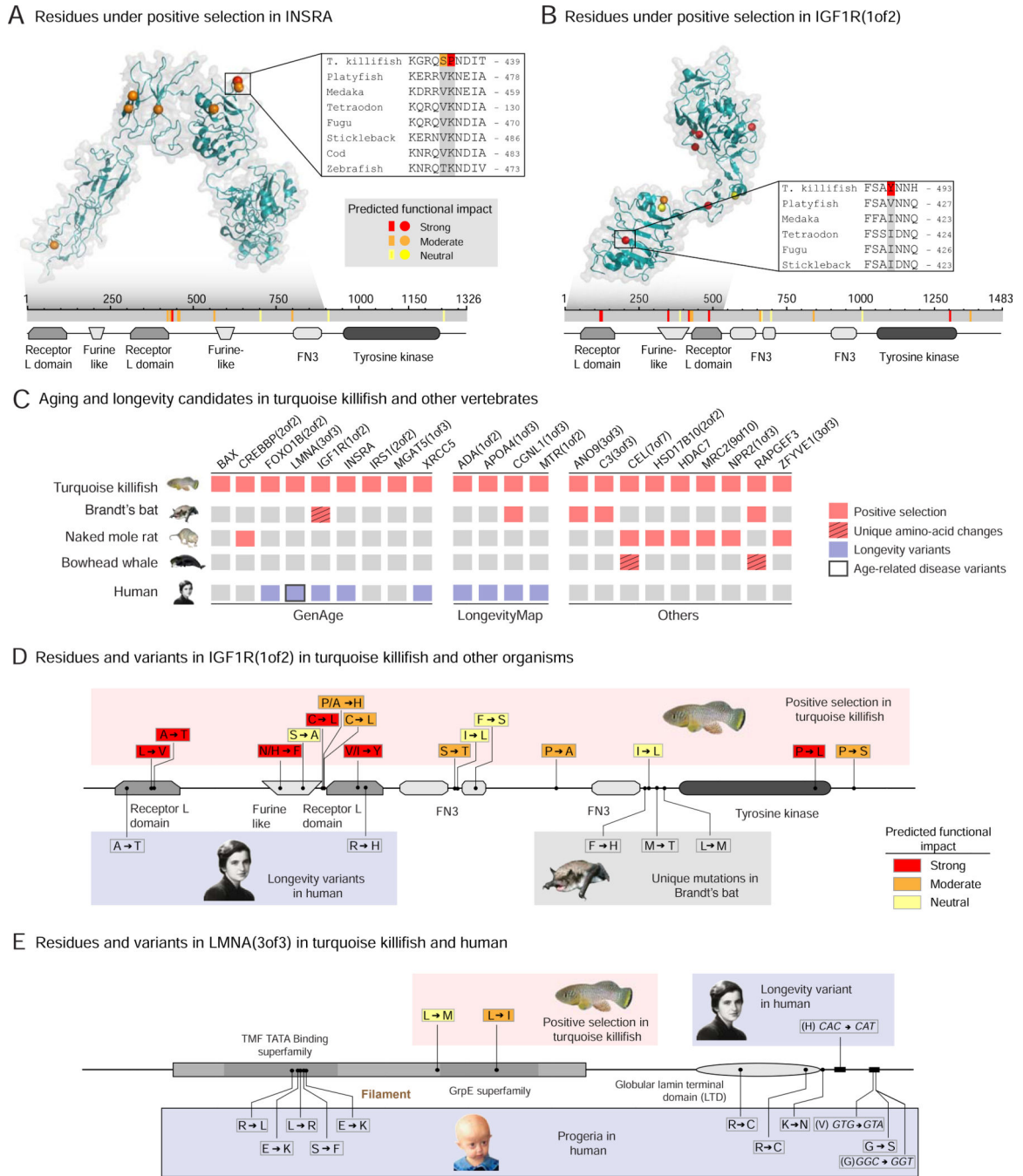
Figure 4. Aging and longevity genes under positive selection in the turquoise killifish and with variants in long-lived species or in humans

A-B) Location of residues under positive selection and with putative functional consequences in insulin receptor A (INSRA) (A), IGF1R(1of2) (B) in the turquoise killifish. Top panels: crystal structure of human orthologs. Color represents the strength of functional impact. Grey shadow: region of the protein with available crystal structure. Insert: alignment of an example residue with strong functional effect in the turquoise killifish and other fish. Bottom panels: schematic of the residues mapped on the turquoise killifish protein sequence (grey). Colored bars: residues under positive selection with different functional impacts. The

conserved protein domains and functional sites are also indicated. FN3: Fibronectin type-III repeats.

(C) Aging and longevity candidates under positive selection in the short-lived turquoise killifish and their variation in long-lived animal species and in humans. Left: aging-related genes from the GenAge database (human and mouse combined, Table S4A). Middle: genes identified in association studies in humans from the LongevityMap database (Table S4A). Right: other genes that are also under positive selection or uniquely changed in other species with extreme longevity phenotype (naked mole rat, Brandt's bat, bowhead whale).

D) Location and variants of residues under positive selection in the turquoise killifish for IGF1R(1of2), and their location and variants in long-lived species and in human centenarians. Top: turquoise killifish variants, with the changed amino acid on the right. Color represents the strength of functional impact. Bottom: variants associated with centenarians in humans or residues with unique amino acid changes in long-lived Brandt's bat mapped on the turquoise killifish sequence. The variants correspond to the amino acids on the right.

E) Location and variants of residues under positive selection in the turquoise killifish for LMNA(3of3), and their location and variants in progeria and human centenarians. Top left: turquoise killifish variants, with the changed amino acid on the right. Top right: variants associated with centenarians in humans mapped onto the turquoise killifish sequence. Bottom: Variants in Hutchinson-Gilford Progeria Syndrome mapped onto the turquoise killifish sequence. These variants are the amino acids on the right. For the turquoise killifish, color represents the strength of functional impact.

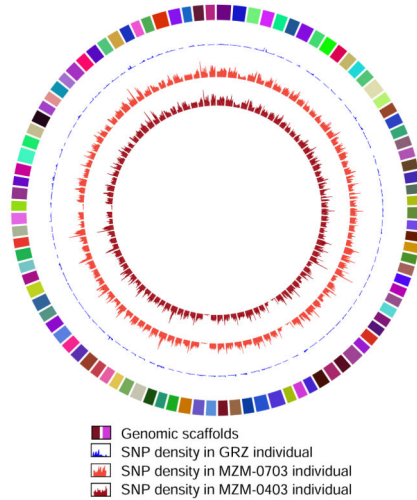A Resequencing of turquoise killifish individuals from various strains

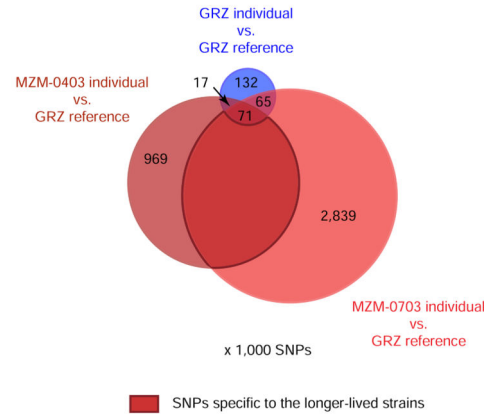| Sequencing	Strain		Captive lifespan [c]	Year / country of capture	# of individuals	Coverage	SNPs
Reference genome [a]	GRZ		Shorter-lived	1968 / Zimbabwe	9	80X	-
Genome resequencing [b]	GRZ		Shorter-lived	1968 / Zimbabwe	1	19X	284,386
	MZM-0703		Longer-lived	2007 / Mozambique	1	54X	4,671,729
	MZM-0403		Longer-lived	2004 / Mozambique	1	6X	2,754,094

a Data used for reference assembly
b Data from individual fish not used for the reference assembly
c This study; Terzibasi et al, 2008; Kirschner et al, 2012.

B SNP density over the 100 longest scaffolds

C Unique and shared SNPs between resequenced individuals

Genomic scaffolds
SNP density in GRZ individual
SNP density in MZM-0703 individual
SNP density in MZM-0403 individual

GRZ individual vs. GRZ reference
MZM-0403 individual vs. GRZ reference
17 | 132 | 65 | 71
969
2,839
MZM-0703 individual vs. GRZ reference
x 1,000 SNPs

SNPs specific to the longer-lived strains

D Example genes with non-synonymous SNPs in individuals from longer-lived strains

|	Hallmark of aging	Genes with SNPs specific to the longer-lived strains [a]	Example genes that have variants with predicted functional effect
	Altered intercellular communication	45	*GRN*, *TNFB(3of3)*, *PDGFRA*
	Genomic instability	39	*BRCA1*, *TP53BP1*, *ERCC6*
	Deregulated nutrient sensing	22	*GHR(1of4)*, *IRS4*, *FOXO4*
	Cellular senescence	13	*MYC(2of2)*, *EGR1(1of5)*
	Epigenetic alterations	8	*MED1(1of2)*, *NCOR1(1of2)*
	Mitochondrial dysfunction	8	*POLG*, *GSR*
	Stem cell exhaustion	8	*MGAT5(1of3)*
	Telomere attrition	8	*TERT*
	Loss of proteostasis	7	*HSF1, HSPA9*

a One gene may belong to more than one hallmark

Figure 5. Genetic variation in individuals from different strains of the turquoise killifish

A) Resequencing of individuals from different turquoise killifish strains (GRZ, MZM-0703, MZM-0403) with different reported lifespans in specific laboratory environments. SNPs: Single Nucleotide Polymorphisms.

B) Circos plot of the 100 longest scaffolds showing SNP density in resequenced individuals from different turquoise killifish strains (GRZ, MZM-0703, MZM-0403) versus the reference GRZ assembly.

C) Unique and shared SNPs between resequenced individuals from reported shorter-lived (GRZ) or longer-lived strains (MZM-0703 and MZM-0403) versus the reference GRZ assembly. Values in the Venn diagram should be multiplied by 1000 and are rounded for concision.

D) Non-synonymous SNPs specific to individuals from the longer-lived strains (MZM-0403 and MZM-0703) in genes encompassing the hallmarks of aging. Aging-related genes were obtained from the GenAge database (human and mouse combined, Table S4A). All presented aging-related genes had at least one variant with predicted functional effect by both SIFT and PROVEAN (bolded genes) or by SIFT or PROVEAN (non bolded genes) (see also Table S5B).
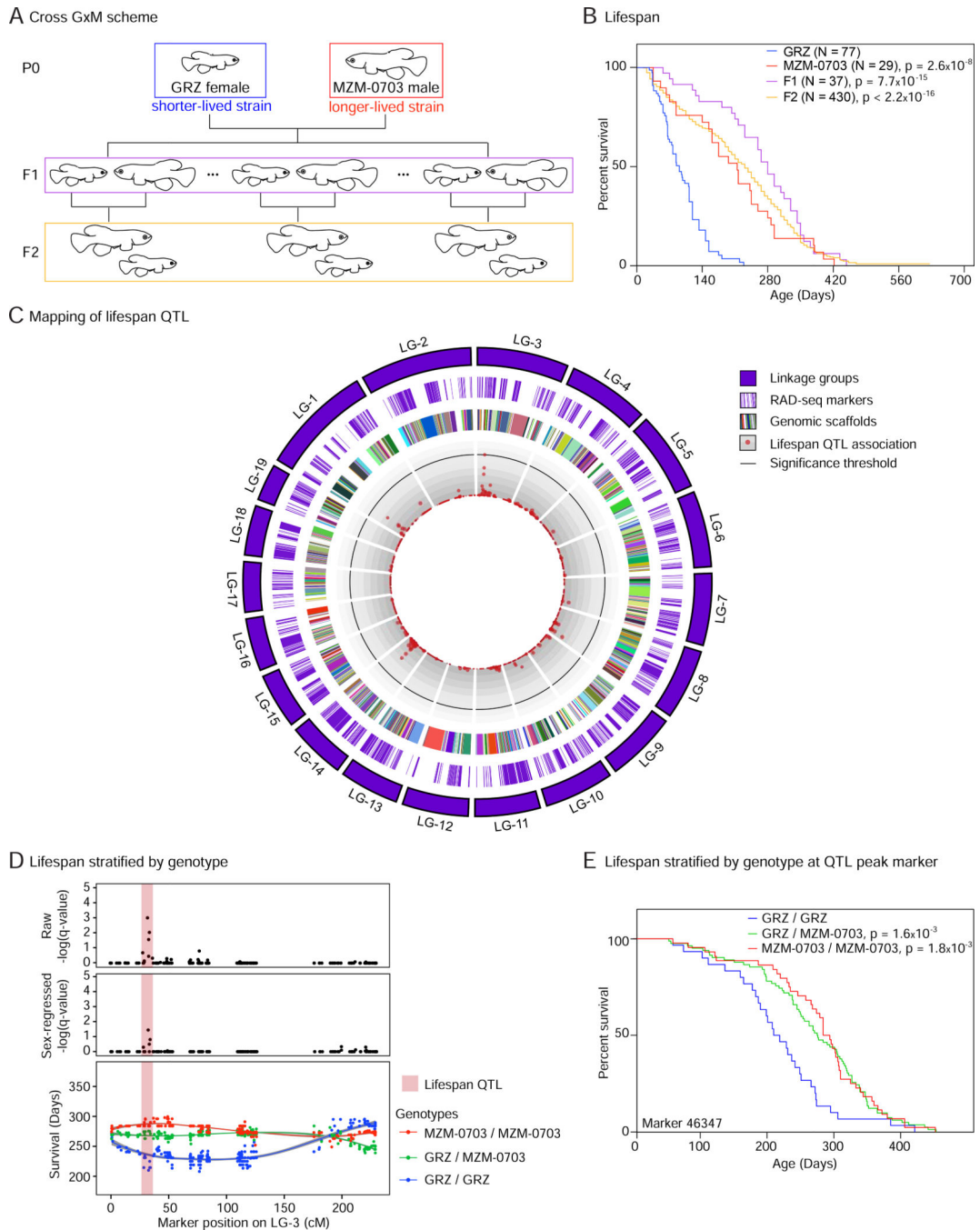
Figure 6. Genetic architecture of lifespan using a cross between shorter-lived and longer-lived strains of the turquoise killifish

A) Scheme of cross GxM. A female from the shorter-lived GRZ strain was crossed with a male from the longer-lived MZM-0703 strain (P0) to generate F1 progeny. F1 individuals were mated to generate F2 progeny.

B) Lifespan of the parental strains, F1, and F2 progeny of cross GxM in the captive conditions used in this study (pooled males and females). p-values for differential survival compared to GRZ individuals in Log-Rank tests are indicated. See Table S6A for complete statistics.

C) Circos plot representing the linkage map of cross GxM and association of markers with lifespan by quantitative trait locus (QTL) analysis. The linkage map is composed of 19 linkage groups (LG). The association of each RAD-seq marker with differences in lifespan is represented as –log10 of the Random Forest Analysis q-value (red dots). The 5% FDR significance threshold is denoted by a black line. There is one marker above this threshold in LG-3 (lifespan QTL). Genomic scaffold length is scaled to genetic distance (cM) and not physical distance (bp).

D) The lifespan QTL is non-transgressive. Upper panel: Raw –log10 (Random Forest Analysis q-value) for association of markers to individual fish lifespan. Middle panel: –log10 of the Random Forest Analysis q-values for association of markers to lifespan after sex regression to account for the possible effect of sex as a confounding variable. Bottom panel: Survival stratified by genotype associated with each marker on LG-3. Homozygotes with alleles coming from the long-lived MZM-0703 grandparent (red) exhibit highest survival at the ~35cM position on LG-3, whereas homozygotes with alleles coming from the short-lived GRZ grandparent (blue) exhibit lowest survival. Light red rectangle: lifespan QTL region.

E) Lifespan of fish with different genotypes at the marker that is most significantly associated to the lifespan QTL (RAD-seq marker 46347). p-values for differential survival compared to individuals with the GRZ/GRZ genotype in Log-Rank tests are indicated. See also Table S6B for complete statistics.

A Mapping of sex-determining region

B Marker association with lifespan or sex

C Genes underlying the lifespan QTL

D Impact of turquoise killifish GRN non-synonymous variants

E Cluster of aging genes in a region of suppressed recombination
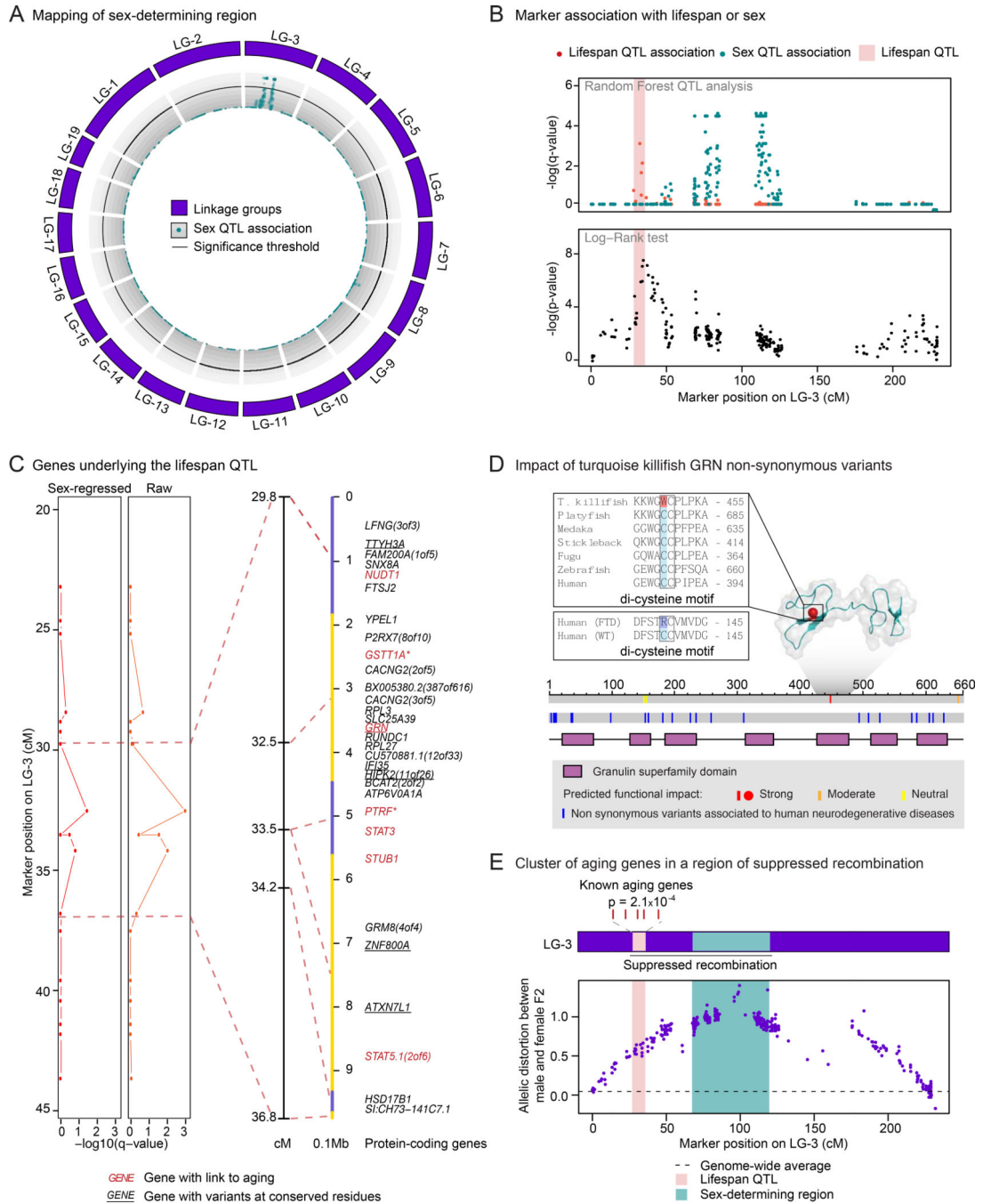
Figure 7. The lifespan QTL is linked to the sex-determining region and contains a cluster of aging genes

A) Circos plot representing the association of markers with sex by QTL analysis in cross GxM. The association of each RAD-seq marker with sex is represented as –log10 of the Random Forest Analysis q-value (turquoise dots). The 5% FDR significance threshold is denoted by a black line. There is a cluster of markers above this threshold in LG-3 (sex-determining region).

B) The lifespan QTL is distinct from the sex-determining region. Upper panel: Random Forest analysis for marker association with lifespan (red) or sex (turquoise). Lower panel:

Log-Rank survival analysis in the F2 generation of cross GxM between homozygotes with alleles coming from GRZ grandparent vs. the MZM-0703 grandparent at each marker. Light red rectangle: Lifespan QTL region.

C) Identification of the genes underlying the lifespan QTL on LG-3. Left panels: Sex-regressed or raw −log10 of the Random Forest Analysis q-value for association with lifespan. Dashed lines delimit the lifespan QTL. Right panels: lifespan QTL region on LG3 and corresponding anchored genomic scaffolds in alternating yellow and slate blue colors. Markers are linked to the mid-points of the scaffolds. Genes in red have been previously linked to aging in the GenAge database (human and mouse combined, Table S4A) or manually curated from the literature (asterisks, Table S7A). Underlined genes have non-synonymous variants at evolutionary-conserved residues. See also Figure S7B–S7G.

D) Location of residues with putative functional consequences and associated to human neurodegenerative diseases on GRN in the turquoise killifish. Color represents the strength of functional impact. Top panel: NMR structure of human orthologous domain. Grey shadow: GRN domain with available NMR structure. Top insert: alignment of the residue with strong functional effect W449 in the turquoise killifish with other species (see also Figure S7D and Table S7H). Bottom insert: region surrounding a mutation found in human frontotemporal dementia (FTD) patients, involving an analogous di-cysteine motif residue. Bottom panel: schematic of the residues mapped on the turquoise killifish protein sequence (grey).

E) Cluster of known aging genes in the lifespan QTL region, in a region of suppressed recombination. Top panel: schematic of the enrichment for known aging-related genes (from GeneAge, human and mouse combined) in the lifespan QTL region ($p = 2.1 \times 10^{-4}$, in Fisher's exact test, compared to rest of the genome, $p = 6.4 \times 10^{-4}$ in Fisher's exact test, compared to the rest of LG-3). Bottom panel: measure of suppressed recombination by allelic distortion between the male and female F2 progeny at each marker on LG-3. See also Figure S7H–S7I. Dash line indicates the genome-wide average for allelic distortion. Light red rectangle: lifespan QTL region. Turquoise rectangle: sex-determining region.