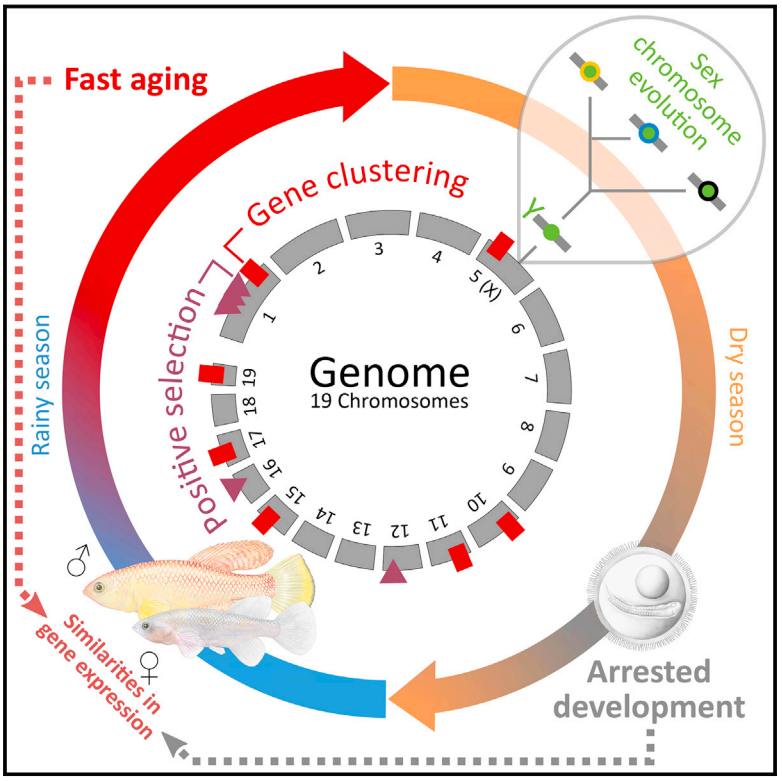
Resource

Cell

Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish

Graphical Abstract



Highlights

- The genome sequence of a very short-lived fish is a resource for aging research
- The sex chromosomes display features of early mammalian XY evolution
- Aging-related genes are clustered in specific genomic regions
- Transcriptional profiles show similarities between developmental arrest and aging

Authors

Kathrin Reichwald, Andreas Petzold, Philipp Koch, ..., Alessandro Cellerino, Christoph Englert, Matthias Platzer

Correspondence

matthias.platzer@leibniz-fli.de

In Brief

The turquoise killifish has a lifespan of only 4–12 months and yet its aging shares many similarities with that of humans. We sequenced and analyzed the killifish genome and provide insights into its biology. We detected very early stages of sex chromosome evolution, identified the sex-determining master gene, found clustering of aging-related genes in the genome, identified genes under positive selection, and discovered that similar gene sets are regulated during developmental arrest of embryos and aging.

Accession Numbers

KG817100
KG959958

Reichwald et al., 2015, Cell *163*, 1527–1538
December 3, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2015.10.071

CellPress

Resource

Cell

Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish

Kathrin Reichwald,[1,14] Andreas Petzold,[1,14,16] Philipp Koch,[1,14] Bryan R. Downie,[1,14] Nils Hartmann,[1,14] Stefan Pietsch,[1] Mario Baumgart,[1] Domitille Chalopin,[2,17] Marius Felder,[1] Martin Bens,[1] Arne Sahm,[1] Karol Szafranski,[1] Stefan Taudien,[1] Marco Groth,[1] Ivan Arisi,[3] Anja Weise,[4] Samarth S. Bhatt,[4] Virag Sharma,[5,6] Johann M. Kraus,[7] Florian Schmid,[7,8] Steffen Priebe,[9] Thomas Liehr,[4] Matthias Görlach,[1] Manuel E. Than,[1] Michael Hiller,[5,6] Hans A. Kestler,[1,7,10] Jean-Nicolas Volff,[2] Manfred Schartl,[11,12] Alessandro Cellerino,[1,13,15] Christoph Englert,[1,10,15] and Matthias Platzer[1,15,*]

[1]Leibniz Institute on Aging-Fritz Lipmann Institute (FLI), Jena 07745, Germany

[2]Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, CNRS UMR5242, Université Claude Bernard Lyon 1, 69364 Lyon Cedex, France

[3]Genomics Facility, European Brain Research Institute (EBRI) Rita Levi-Montalcini, Rome 00143, Italy

[4]Jena University Hospital, Institute of Human Genetics, Friedrich Schiller University, Jena 07743, Germany

[5]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden 01307, Germany

[6]Max Planck Institute for the Physics of Complex Systems, Dresden 01307, Germany

[7]Medical Systems Biology, Ulm University, Ulm 89069, Germany

[8]International Graduate School in Molecular Medicine at Ulm University (GSC270), Ulm 89069, Germany

[9]Leibniz Institute for Natural Product Research and Infection Biology-Hans-Knoell-Institute (HKI), Jena 07745, Germany

[10]Faculty of Biology and Pharmacy, Friedrich Schiller University Jena, Jena 07743, Germany

[11]Department of Physiological Chemistry, Biocenter, University of Wuerzburg, Wuerzburg 97074, Germany

[12]Comprehensive Cancer Center Mainfranken, University Hospital Wuerzburg, Wuerzburg 97074, Germany

[13]Laboratory of Biology, Scuola Normale Superiore, Pisa 56126, Italy

[14]Co-first author

[15]Co-senior author

[16]Present address: Deep Sequencing Group SFB 655, Biotechnology Center, Dresden University of Technology, Dresden 01307, Germany

[17]Present address: Department of Genetics, University of Georgia, Athens, GA 30602, USA

*Correspondence: matthias.platzer@leibniz-fli.de

http://dx.doi.org/10.1016/j.cell.2015.10.071

SUMMARY

The killifish *Nothobranchius furzeri* is the shortest-lived vertebrate that can be bred in the laboratory. Its rapid growth, early sexual maturation, fast aging, and arrested embryonic development (diapause) make it an attractive model organism in biomedical research. Here, we report a draft sequence of its genome that allowed us to uncover an intra-species Y chromosome polymorphism representing—in real time—different stages of sex chromosome formation that display features of early mammalian XY evolution "in action." Our data suggest that *gdf6Y*, encoding a TGF-β family growth factor, is the master sex-determining gene in *N. furzeri*. Moreover, we observed genomic clustering of aging-related genes, identified genes under positive selection, and revealed significant similarities of gene expression profiles between diapause and aging, particularly for genes controlling cell cycle and translation. The annotated genome sequence is provided as an online resource (http://www.nothobranchius.info/NFINgb).

INTRODUCTION

The turquoise killifish *Nothobranchius furzeri* (Jubb, 1971) is an annual fish that inhabits seasonal freshwater ponds in the southeast of Africa. It is characterized by rapid growth, early sexual maturation, and an exceptionally short lifespan reflecting the adaptation to the ephemeral nature of the habitat (Blažek et al., 2013; Cellerino et al., 2015; Genade et al., 2005). Several laboratory strains exist differing in their origin and lifespan. The GRZ strain comes from a semi-arid habitat in Zimbabwe (Figures 1A, 1C, and 2A) (Jubb, 1971), where its founders were collected in 1969 and have a maximum lifespan of 4–6 months. To date, this is the shortest maximum lifespan reported for a vertebrate bred in captivity (Valdesalici and Cellerino, 2003). Strains from semi-arid or more humid regions in Mozambique (e.g., MZM-0403 and MZM-0410) (Figure 1B) and the borderland between Mozambique and Zimbabwe (MZZW-0701) have a longer maximum lifespan of ~1 year (Terzibasi et al., 2008; Tozzini et al., 2013). These strains were established only several years ago and are genetically heterogeneous, whereas GRZ is highly inbred (Reichwald et al., 2009). In spite of the short lifespan, both GRZ and MZM strains show typical signs of aging, i.e., a decline in cognitive and behavioral capacity accompanied by aging-related histological changes (Di Cicco et al., 2011; Terzibasi et al., 2007) as well as aging-related telomere shortening and impairment of mitochondrial function (Hartmann et al.,

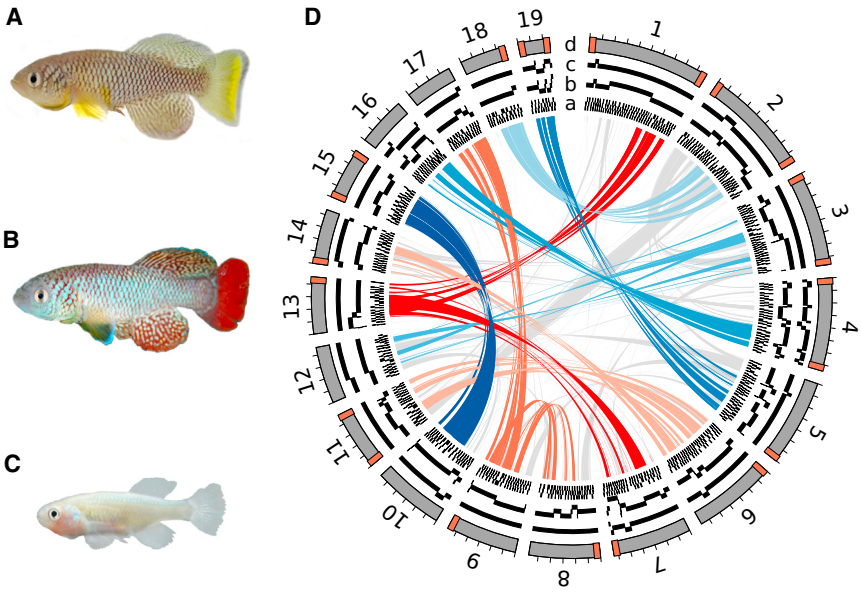


Figure 1. The Turquoise Killifish and the Genome Assembly

(A) Adult GRZ male.

(B) Adult MZM-0403 male.

(C) Adult GRZ female.

(D) Circles: the stepwise assembly of the reference sequence is represented from the inner to the outer circle. a: scaffolds obtained by applying programs ALLPATHS-LG and KILAPE. b: super-scaffolds built upon integration of optical mapping data. c: genetic scaffolds generated by linkage map integration. d: synteny groups defined upon analyses of synteny in medaka and stickleback. Synteny groups are sorted by length and numbered accordingly. Chromosome ends identified by optical mapping are marked in orange. The distance between two ticks is 10 Mb. Center: pairs of paralogous genes for synteny groups with a 1:1 (1:2) relation are connected by blue (red) lines; different hues define different chromosomal pairs (trios). Grey lines indicate gene pairs that do not follow our classification of chromosomal paralogy.

See also Figures S1 and S2 and Data S1.

2009, 2011). Lifespan determination in *N. furzeri* is polygenic; four quantitative loci relevant for lifespan are presently known (Kirschner et al., 2012).

Due to its fast development, *N. furzeri* can reach sexual maturity in <3 weeks and first signs of sexual dimorphism are apparent at 2 weeks after hatching (Blažek et al., 2013). In vertebrates, the gonads are usually the last organ system to develop into the functional adult structure. In fish, gonad differentiation commences only at late larval stages or even after metamorphosis and full functionality is reached at puberty (Devlin and Nagahama, 2002). Also, the sex determination system that provides the decision whether the undifferentiated gonad anlage of the embryo will develop later into a testis or an ovary is very plastic and can differ between closely related fish species or even within species (Volff et al., 2007). The fact that sex determination systems can change easily or arise rapidly during fish evolution together with the necessary rapid development of the reproductive system observed in *N. furzeri*, raises the question whether fast lifecycle and short lifespan influenced the evolution of the primary sex-determining (SD) gene and the sex chromosomes. Thus far, the segregation analyses of four sex-linked markers in crosses of GRZ and MZM-0403 is concordant with an XY sex determination system (Kirschner et al., 2012; Valenzano et al., 2009). The identical morphology (homomorphy) of the putative sex chromosomes (Reichwald et al., 2009) pointed to their young age and possibly to a situation of "sex chromosome evolution in action."

To survive the dry season, embryos of *N. furzeri* are protected from dehydration by a desiccation-resistant chorion and can enter into a state of developmental arrest termed diapause; the latter being a well-known adaptation in animal species to overcome unfavorable conditions. In *N. furzeri*, the arrest may occur at three distinct developmental stages (diapause I, II, and III) and can last for more than a year. Also in the nematode *Caenorhabditis elegans*, a larval arrest is observed (dauer larvae), and genes relevant for entering and maintaining the dauer state affect lifespan (Kenyon et al., 1993). We therefore analyzed whether differentially expressed genes (DEGs) in *N. furzeri* diapause versus non-diapause embryos are regulated in aging.

Recently, protocols for transgenesis (Hartmann and Englert, 2012; Valenzano et al., 2011) and CRISPR/Cas9-mediated mutagenesis have been established for *N. furzeri* (Harel et al., 2015). These tools, together with the short lifespan, make *N. furzeri* a very attractive vertebrate model to study aging, developmental arrest, and the interrelationship between both phenotypes. Here, we report a high-quality draft sequence of the *N. furzeri* genome. We provide insights into the very early evolutionary stages of an XY sex determination system, reveal clustering of aging-related genes in specific genomic regions, identify genes under positive selection and detect common expression profiles in diapause and aging.

RESULTS AND DISCUSSION

Assembly and Annotation of a High-Quality Draft Genome Sequence with Long-Range Contiguity

Today's challenge in genome analysis is generating a reference sequence of high quality and long-range contiguity. The *N. furzeri* project required special efforts because the genome is large and repeat-rich (Reichwald et al., 2009). In these two aspects, it resembles the zebrafish genome, for which a high-quality reference sequence was published only recently (Howe et al., 2013). We sequenced genomic DNA from *N. furzeri* females of the highly inbred GRZ strain (Figures 1A and 1C) in which all autosomes and the X chromosome are nearly homozygous. Using Illumina and Roche next-generation sequencing (NGS) technologies, we obtained whole-genome shotgun (WGS) data from 17 paired-end and mate-pair libraries amounting to 236 Gb (158-fold coverage, based on a genome-size estimate of 1.5 Gb; Figure S1B; Data S1A and S1B). Further, we sequenced

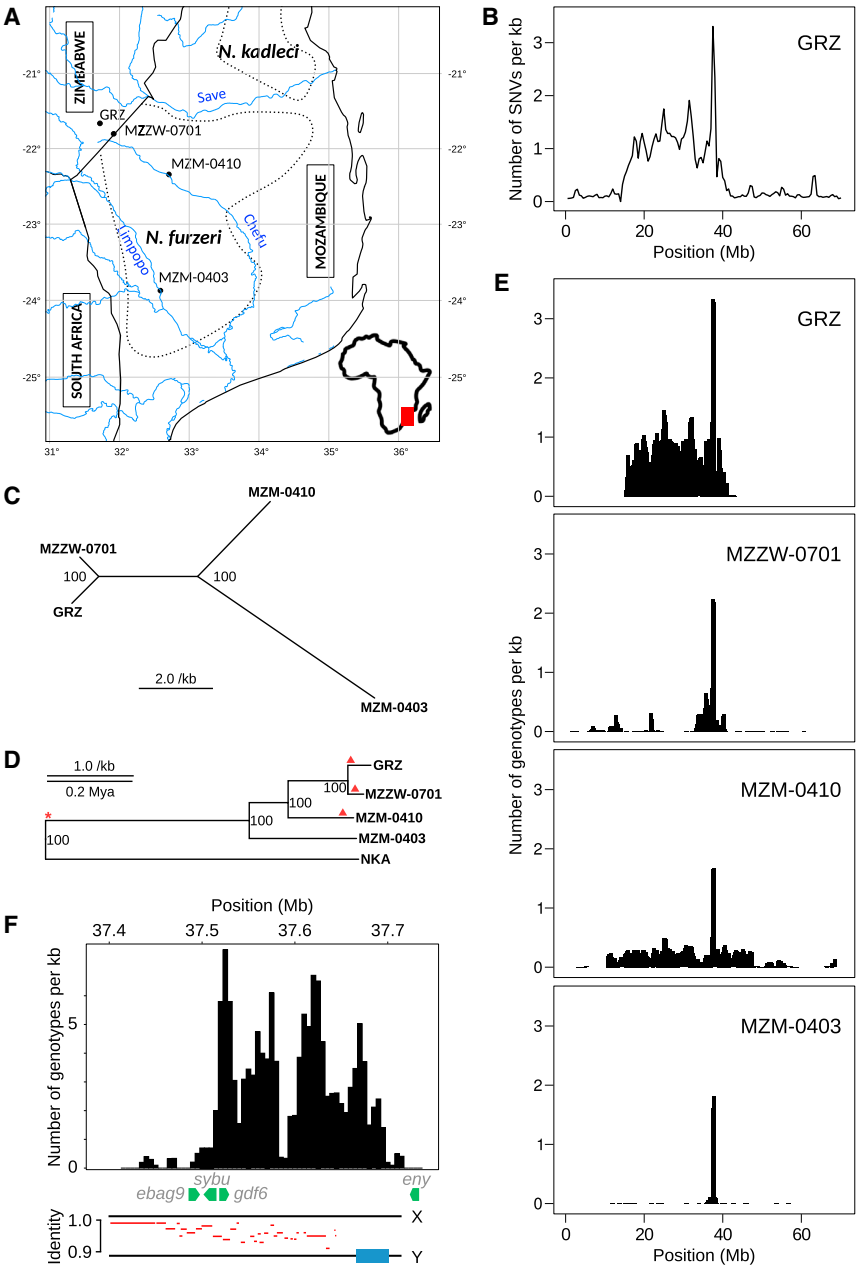


Figure 2. Phylogeny and Sex Chromosome Analyses of *N. furzeri* Strains GRZ, MZZW-0701, MZM-0410, and MZM-0403

(A) Geographic origin of *N. furzeri* strains is indicated by dots. The distribution range of *N. furzeri* and *N. kadleci* is marked by dotted lines.

(B) SNV density profile for synteny group (sgr) 05 obtained by aligning GRZ male WGS reads to the female reference sequence (sliding window: 1 Mb, step size: 500 kb).

(C) Phylogenetic relationship between strains based on WGS variation data.

(D) Phylogenetic tree of *N. furzeri* strains rooted by their sister species *N. kadleci* (NKA) based on exonic variations obtained by RNA-seq (Data S4I). The divergence time of *N. furzeri* and *kadleci* was estimated as 0.75 Mya (Dorn et al., 2014) and used for scaling. Red marks indicate the primary (asterisk) and secondary (triangle) events leading to the suppression of recombination shown in (E) and (F).

(E) Genotype density profiles for sgr05 of the strains. Genotype data were filtered for SNV positions in a given strain where all females are homozygous and all males are heterozygous (sliding window: 500 kb, step size: 250 kb).

(F) Top: zoom into the genotype density profile of the sex-determining region (SDR) in MZM-0403 (sliding window: 10 kb, step size: 5 kb). Genes annotated in the SDR of the GRZ female reference sequence are shown as green arrows. Bottom: identity plot (red lines) of BAC-based X and Y chromosome-specific sequences (black lines). The blue box represents a Y-specific 35 kb tandem repeat cluster composed of repeat units of 634 nt and 150 nt.

See also Figure S3 and Data S1I and S2A–S2C.

genomic insert ends of 81,393 BACs and fosmids to assist in the assembly and to provide a physical resource of the *N. furzeri* genome (5.3-fold clone coverage; Data S1C and S1D). To build the assembly, a five-step strategy was applied: we started with ALLPATHS-LG (Gnerre et al., 2011), continued with scaffolding, integrated optical and three genetic linkage maps, and finished with comparative synteny mapping in two closely related fish species (Table 1). The incorporation of optical mapping data remarkably improved the assembly contiguity (30-fold, Figure S1C). The genome assembly, in the following referred to as reference sequence, comprises 1.24 Gb (scaffold N50: ~0.5 Mb, optical N50: ~16 Mb, synteny N50: ~57 Mb), of which 87% were assigned to 19 synteny groups (sgrs) (Figure 1D). For 15 sgrs, we identified the corresponding *N. furzeri* chromosomes by fluorescence in situ hybridization using BAC probes (Figure S2A; Data S1E). Further, optical mapping data indicate that the assembly reached 22 of 38 chromosome ends (Figures 1D and S1D).

We built a comprehensive catalog of repetitive elements using Sanger/Illumina WGS reads and the genome assembly. Based on the Sanger data, we determined a repeat content of 64.6%, comprising 42.1% dispersed and 22.5% tandem repeats. This was confirmed by non-assembled NGS WGS data (Figure S2B). The *N. furzeri* reference sequence, however, contains only 35% repeats. In particular, tandem repeats are underrepresented (2% instead of 22.5%). This is most likely caused by the short NGS reads that collapse during the assembly process. Dispersed repeats amount to 33%, with LINEs being most abundant (8.4%) attributable to a recent expansion in this class of retrotransposons (Figure S2C; Data S1F). Finally, we confirmed the high quality of the reference sequence by PacBio WGS- and BAC sequencing (Data S1H and S1I) showing that gaps

Table 1. Statistics of the Stepwise Assembly

Assembly Step	Number of Scaffolds	Total Length (bp)	Fraction of N[a] (%)	Longest Assembly Unit (bp)	N50 (bp)
A ALLPATHS-LG	15,930	900,823,930	9.9	1,451,049	132,538
B Scaffolding + gap filling	7,675	943,595,854	9.2	3,869,209	494,454
C Optical map integration	6,012	1,230,898,532	30.4	44,272,285	15,858,201
D Genetic map integration	5,924	1,239,698,532	30.9	96,068,516	48,234,189
E Synteny integration	5,896	1,242,498,532	31.0	98,476,147	57,367,160
Anchoring within the Final Assembly					
Chromosomes/synteny groups	19	1,078,719,814	33.64	98,476,147	63,666,967
Autosomes	18	1,008,464,687	33.42	98,476,147	57,680,405
X chromosome	1	70,255,127	36.78	70,255,127	70,255,127
Unassigned	5,877	163,778,718	13.94	1,706,182	81,864

See also Data S1.

[a]Unresolved nucleotide positions, stands for A, C, G, or T.

contained in the genome assembly are almost entirely composed of repeats (83.1%).

We performed gene annotation using comprehensive RNA sequencing (RNA-seq) and microRNA sequencing (miRNA-seq) datasets as well as protein homology and in silico prediction tools (Figure S2D; Data S1K–S1N). We annotated 26,141 protein-coding genes with 59,154 transcripts, and 59 rRNA, 453 tRNA, 184 small nucleolar RNA (snoRNA), 598 miRNA, and 117 other non-protein coding RNA (ncRNA) genes (a detailed description of the miRNome will be reported elsewhere; M. Baumgart, I.A., and A.P., unpublished data). The teleost genome duplication (TGD) is reflected by the presence of 2,229 paralogous gene pairs, representing 17% of the *N. furzeri* protein-coding genes; further, we identified five pairs of putatively paralogous chromosomes with a 1:1 and three triads with a 1:2 relationship (Figure 1D; Data S1O).

To assess the completeness of the reference sequence with respect to the non-repetitive fraction of the genome, we used the Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al., 2007) and searched in *N. furzeri* for orthologs of 248 highly conserved genes present in most eukaryotic genomes. Of these, we detected 98% in the reference sequence with 95% being completely covered (Data S1P). Furthermore, we could align 91% of the *N. furzeri* transcript catalog (Petzold et al., 2013) with the reference sequence strongly suggesting a highly complete representation of the genic fraction of the genome. Moreover, the PacBio-based estimate of the repeat content in gaps confirms that ~90% of the non-repetitive genome fraction is represented in the assembly. The annotated genome reference sequence is accessible at the *N. furzeri* Information Network Genome Browser (NFINgb, http://www.nothobranchius.info/NFINgb). Its long-range contiguity, chromosomal scale assembly, and completeness of genic regions allow studying the biology of the *N. furzeri* genome.

Insights into Early Events of XY Sex Chromosome Evolution

To map the SD region (SDR) in the reference sequence, which represents a GRZ female genome, we performed additional WGS sequencing of four GRZ males (Data S2B). Because the GRZ strain is highly inbred, we expected genomic variations predominantly in the region of suppressed recombination between male and female sex chromosomes. Accordingly, male single nucleotide variations (SNVs) were mainly confined to a region on sgr05 (Figures 2B and S3A) that bears the only four sex-linked markers identified so far (Kirschner et al., 2012; Valenzano et al., 2009). This male-specific region of the Y chromosome (MSY) encompasses 26.1 Mb (sgr05: 15,031,832–41,162,746) and exhibits a distinct peak in variation density at position 37.6 Mb. PCR/Sanger sequencing-based validation of sex-linkage for selected SNVs pointed to an intra-species sex chromosome polymorphism between *N. furzeri* strains. For example, variations in the syntabulin gene (*sybu*) are associated with sex in GRZ, MZZW-0701, and MZM-0410 but not in MZM-0403, whereas SNVs up to 42 kb upstream of *sybu* show sex-linkage in all strains (Data S2A).

By analyzing the intra-species variations by additional WGS data from males and females of MZZW-0701, MZM-0410, and MZM-0403 in more detail (Data S2B), we identified ~3.3 million SNVs (accessible at NFINgb). Using those SNVs to determine the phylogenetic relationship between strains, we found a good agreement with the geographic location of collection sites (Figures 2A and 2C). Rooting of the phylogenetic tree revealed that MZM-0403 belongs to a different lineage than the three other strains (Figure 2D), thus confirming the deep geographic structuring of the species (Bartáková et al., 2013; Dorn et al., 2011). We next searched genome-wide for signs of suppressed recombination and identified the most prominent region in all strains on sgr05 (Figures 2E and S3A). In GRZ, the SNV and genotype density profiles coincide (Figures 2B and 2E) suggesting that the same genetic signal of suppressed sex chromosomal recombination was detected with both approaches. While the size of the MSY differs considerably between strains, ranging from 196 kb to 37 Mb (Data S2C), the position of the variation peak is identical. To date, intra-species sex chromosome polymorphisms have been observed only in exceptional cases and only by using cytogenetic methods, e.g., in guppy (Nanda et al., 2014).

Comparative variation analyses of this remarkable strain-specific Y chromosome polymorphism indicate a two-step scenario for its evolution. First, an ancient event in the common ancestor of all strains led to suppressed recombination in a 196-kb region and the emergence and/or fixation of a SD signal. This stage of early sex-chromosome evolution is conserved in MZM-0403 (Figure 2F, top). In all four strains, the highest number of sex chromosome-specific SNVs was accumulated in the 196-kb region, indicating that recombination suppression shielded in the ancestral state the newly evolving SD gene from cross-over and the proto-Y from losing its identity. To shed light on the mechanism of recombination suppression, we sequenced X- and Y-specific BACs harboring this region using PacBio technology (Data S1I). The BAC-based X-specific assemblies confirmed the reference sequence. In addition, we obtained a corresponding Y-specific region encompassing a 35 kb tandem-repeat cluster (Figure 2F, bottom) that similarly to the MSY of the medaka fish (Kondo et al., 2006) may prevent recombination in flanking regions.

Secondary events encompassing larger regions (7–37 Mb), yet containing the primary SDR, occurred independently in each of the three northern strains. By applying FISH analysis, we identified an inversion as the secondary cross-over barrier in MZM-0410 (Figure S3B). Thus, the individually structured *N. furzeri* Y chromosomes seem to reflect the first stages of the mammalian XY evolution that has shaped these chromosomes by consecutive inversions into evolutionary strata over 320 million years (Lahn and Page, 1999). Also, the sex chromosomes of the flatfish *Cynoglossus semilaevis* estimated to be ~30 million years old, have most likely diverged due to suppression recombination by a large inversion (Chen et al., 2014). For *N. furzeri* we estimate the occurrence of the secondary recombination suppression in GRZ around 70 thousand years ago (kya), in MZZW-0701 50 kya and in MZM-0410 38 kya by dating the primary event to the species split between *N. furzeri* and *N. kadleci* at 750 kya (Dorn et al., 2014) (Figure 2D; Data S2D). Although this is a rough estimate, we conclude that the secondary events are very young in evolutionary terms compared to previously studied SD systems.

Our data demonstrate that during early sex chromosome evolution, a whole set of different Ys can be created. In-depth analyses of Y polymorphisms in species with older Y chromosomes will allow studying whether in a second phase the most successful Y might make a sweep through the species. Such a sweep would then lead to a situation noticed for mammalian Ys where only minor sequence variations mark the Y haplotypes in a later phase of Y chromosome evolution (Ellegren, 2003). Future studies will clarify whether population-genetic fragmentation (Bartáková et al., 2013), short lifespan, annualism, and/or the multiple specific adaptations of *N. furzeri* facilitated its unprecedented Y chromosome polymorphism.

Tracing the Emergence of a Novel Sex-Determining Gene: *gdf6Y*

We next attempted to identify the SD gene in *N. furzeri*. The minimal MSY was observed in MZM-0403 encompassing 196 kb and coinciding with the peak of Y-specific sequence variation at position 37.6 Mb in sgr05 (Figures 2E and 2F). This region contains only one annotated gene, *gdf6*, encoding growth differentiation factor 6, a member of the TGF-β family. We propose *gdf6Y* as symbol for the gene in the MSY. In GRZ, the *gdf6Y* coding sequence (CDS) differs from *gdf6* on the X chromosome in 22 SNVs and a 9-bp deletion, resulting in 15 amino acid (aa) exchanges and a 3 aa deletion (Figure S4A). All non-synonymous SNVs and the deletion are conserved between strains (Data S3A and S3B). Remarkably, the part of *gdf6Y* coding for the C-terminal 120 aa and homologous to the mature human GDF6, contains five non-synonymous but no synonymous substitutions indicating that positive selection acted on this part of the protein. The mature growth factor is highly conserved between vertebrates, and all male substitutions affect aa conserved between the *N. furzeri* X-chromosomal Gdf6 and its human ortholog (Figure 3A). Scanning all 339 genes in the 26.1 Mb MSY of GRZ confirms the sequence variations in *gdf6Y* as by far strongest signal of local positive selection (Data S3C).

To evaluate the impact of these aa changes, we performed homology modeling using the structure of the human receptor-bound GDF5 dimer (Kotzsch et al., 2009). Four of the five aa differing between mature Gdf6 and Gdf6Y reside in the modeled region (Figure 3A). Two of them (Gdf6Y/Gdf6: R405/Q408 and V407/E410) point outward into the solvent and reside at the edge of a β sheet (Figures 3B and S4B) that undergoes an induced fit upon formation of the GDF5:receptor complex (Kotzsch et al., 2009). The other two (T364/I367 and V372/M375) are located in a helix being part of the protomer interface but also contacting the receptor (Kotzsch et al., 2009). Hence, all four X/Y variable aa might have a bearing on protein interactions, either during dimerization or in the process of forming complexes with receptor(s).

Comparative analyses of *gdf6*/*gdf6Y* transcript levels revealed biallelic expression in early developmental stages of male and female GRZ and a significantly higher overall expression in males starting at day 3 post-hatching (Figures S4C and S4D; Data S3D). In RNA-seq data of adult ovaries, we found few *gdf6* reads, whereas in testes only *gdf6Y* mRNAs were detected at a considerable level. A possible explanation for the male-specific expression from the Y-chromosomal locus is a *gdf6Y*-specific deletion of 241 bp (sgr05: 37,526,406–37,526,646) in the 3′UTR including a potential mir-430 binding site (Figures S4E–S4G). In fish, mir-430 is an important regulator of germline-specific gene expression (Mishima et al., 2006). It is tempting to speculate that this deletion was the primary event marking the inception of the XY differentiation.

Gdf6Y expression peaks shortly after hatching; this is a time period when sex determination occurs in many fish species. Gdf6 is a member of the TGF-β family known to play a predominant role in developmental processes. Other members of the TGF-β family, e.g., the anti-Mullerian hormone (AMH) and the gonadal soma-derived growth factor (GSDF), as well as their receptors are important factors in sexual development of mammals and other vertebrates and function as master male sex determinants in several fish species (Josso and Clemente, 2003; Kikuchi and Hamaguchi, 2013; Morrish and Sinclair, 2002; Myosho et al., 2012; Rondeau et al., 2013). Gdf9 and Bmp15 are important players in ovarian development of mammals (Otsuka et al., 2011) and fish (Clelland and Kelly, 2011). Gdf6 has not

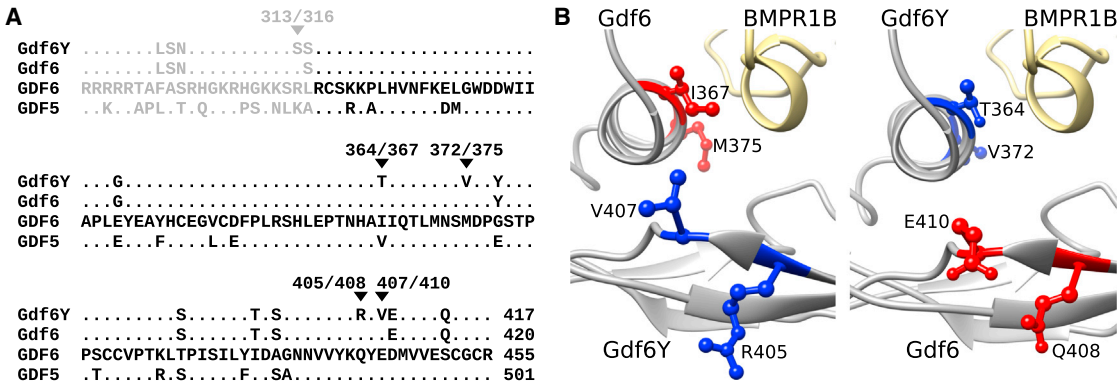


Figure 3. Gdf6Y/Gdf6 Homology Modeling

(A) ClustalW alignment of C-terminal, highly conserved 125 aa of *N. furzeri* Gdf6Y and Gdf6 as well as human GDF6 and GDF5. Amino acids (aa) identical to GDF6 are shown as dots. Amino acids varying between Gdf6Y and Gdf6 are highlighted by filled triangles and their numbers. The first 22 aa depicted in gray were not included in the modeling because they are missing in the reference structure.

(B) Detailed ribbon representations of two regions (left, right) of the modeled Gdf6Y/Gdf6 hetero-dimer (gray) receptor (yellow) complex given in Figure S4B. The four Gdf6Y/Gdf6 variable aa covered by the model are shown with side chains in blue for Gdf6Y and red for Gdf6. In the dimer, these aa are located spatially close to each other in the two regions shown.

See also Figure S4 and Data S3.

been described in the context of gonad development so far; how it acts as a master sex regulator in *N. furzeri* warrants further investigation.

Genomic Positional Enrichment of Aging-Related Genes

Recently, data have accumulated suggesting that eukaryotic genes located in physical proximity may be co-regulated and/or have similar functions. Correlations between chromosomal position and membership of functional gene sets were identified for yeast (Santoni et al., 2013) and human (Thévenin et al., 2014) genomes. Hence, chromosomal and spatial co-localization in the nucleus may indicate co-regulation. It was previously shown that 3D chromatin structure couples nuclear compartmentalization of chromatin domains with the control of gene activity (Guelen et al., 2008) and thus contributes to cell-specific gene expression (Zullo et al., 2012). In this context, it is noteworthy that cellular senescence is associated with modifications of the global chromatin interaction network (Chandra et al., 2015). To our knowledge, it has not yet been investigated whether genes relevant for organismal aging are clustered in genomic regions.

Taking advantage of the long-range contiguity of the *N. furzeri* reference sequence, we set out to study whether aging-related genes show positional gene enrichment (PGE) in sgrs. To this end, we identified aging-related DEGs in three tissues (brain, liver, and skin) by applying two different approaches: (1) we compared young versus old MZM-0410 (5 weeks versus 39 weeks, corresponding to 10% versus 75% of maximum lifespan), and (2) we compared GRZ versus MZM-0410 at 12 weeks. As aging rates differ between these strains (Terzibasi et al., 2008), the same chronological age in the second approach corresponds to 50% of the maximum lifespan in GRZ and 24% in MZM-0410 (Data S4A–S4G).

In total, we detected ten PGE regions. Four of those are based on DEGs obtained by the first approach and six were identified by the second approach (false discovery rate [FDR] < 0.05, scan statistics; Data S4H). These regions are located on seven sgrs, extend over 2.6–9.2 Mb, and contain 11–23 DEGs. On three sgrs, two PGE regions each overlap non-randomly ($p = 0.0012$, resampling test) indicating that the same genomic features were detected by different approaches and in samples from different organs. One of the latter PGE regions located on sgr10 and detected based on DEGs in skin aging (Figure 4A) is enriched for the GO term "response to wounding" (FDR < 0.05, Fisher's exact test). The genes are downregulated in aging (Figure 4B) thus suggesting their co-regulation and providing a link to the well-accepted aging-related phenotype of decreased regenerative capacity (Conboy et al., 2005). These findings demonstrate that *N. furzeri* genes related to aging are distributed non-randomly in the genome and that positional clustering may allow their co-regulation.

Positively Selected Genes in *N. furzeri*

The availability of high-quality genomic reference sequences facilitates the identification of genes under positive selection. To identify genes potentially relevant for adaptation of life-history traits we analyzed *N. furzeri* in comparison with *N. pienaari* because these sympatric species show convergent evolution of short lifespan (Tozzini et al., 2013). Therefore, we generated CDS data for *N. pienaari* and, additionally, for four longer-lived Nothobranchius species as well as the non-annual killifish *Aphyosemion striatum* as outgroup by RNA-seq of brain samples (Data S4I). The consensus tree based on multi-species CDS alignments matched well their reported phylogeny (Dorn et al., 2014) (Figure 4C). To avoid assembly errors, only de novo assembled *N. furzeri* transcripts that show 100% identity to the reference sequence (n = 23,108; corresponding to 11,748 genes) were analyzed. Accordingly, for *N. pienaari* we included transcripts showing at least 99% coverage and 98% identity to the *N. furzeri* reference sequence (n = 5,576; corresponding to 5,363 genes). We identified seven genes under

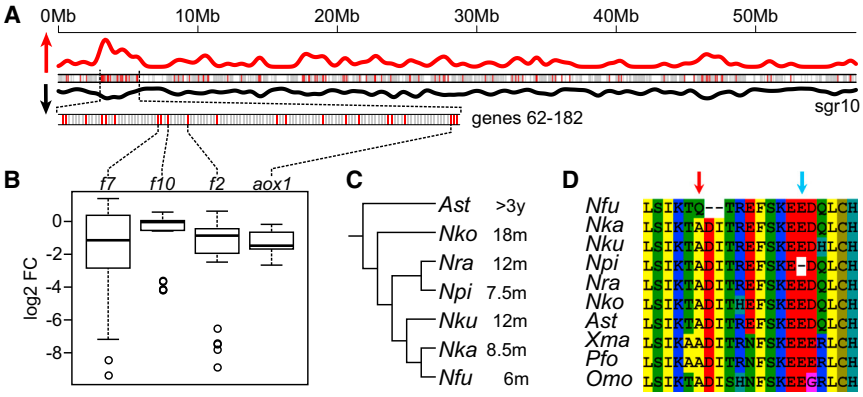


Figure 4. Positional Gene Enrichment and Positive Selection

(A) Schematic representation of synteny group (sgr) 10 and a region of positional gene enrichment. The genes in the sgr are represented by vertical bars: red, differentially expressed; gray, not differentially expressed. The density of all genes on the sgr (black line) and those differentially expressed (red line) is shown (kernel density estimation, Gaussian kernel). Arrows indicate the direction of increasing values.

(B) Relative downregulation of four DEGs with the GO annotation "response to wounding" in aging skin. Gene symbols *f2*, *f7*, and *f10* stand for coagulation factors II, VII, and X. *aox1*, aldehyde oxidase 1. Boxes, first and third quartiles; horizontal line, median; whiskers, most extreme value within 1.5× of inter-quartile range; dots, outliers. Expression differences were calculated by pairwise comparisons (n = 25) between the samples.

(C) Phylogram of the species used for transcriptome sequencing based on Dorn et al. (2014). For each species, the captive median lifespan is reported: *A. striatum* (unpublished), *N. korthause* (Baumgart et al., 2015), *N. rachovii*, *N. pienaari*, *N. kuhntae*, *N. furzeri* (Tozzini et al., 2013), and *N. kadleci* (Ng'oma et al., 2014).

(D) Alignment of the Id3 C terminus. The red arrow indicates aa under positive selection in *N. furzeri* followed by a two aa deletion. The blue arrow indicates the *N. pienaari*-specific deletion. The background color of each aa relates to the chemical nature of its side chain.

See also Data S4A–S4M.

positive selection in *N. furzeri* and one in *N. pienaari* (FDR < 0.05, Data S4J) highlighting the importance of a reference sequence for evolutionary analyses. Remarkably, five of these genes are either up- or downregulated in aging in at least one of three MZM-0410 organs (brain, liver, skin at 39 versus 5 weeks; Data S4A and S4K–S4M).

The signature of selection for *id3* (inhibitor of DNA binding 3, dominant negative helix-loop-helix protein) is particularly interesting. *Id3* is upregulated during aging in brain and skin and is also a key component of TGF-β signaling. TGF-β regulates inflammation, is involved in aging-related diseases such as tumorigenesis, fibrosis, glaucoma, and osteoarthritis (Krieglstein et al., 2012), and regulates life-history traits in *C. elegans* (Luo et al., 2010; Shaw et al., 2007). In *N. furzeri*, the gene shows signs of positive selection; i.e., a radical substitution of a non-polar by a charged aa followed by a 2-aa deletion (Figure 4D). Interestingly, at 10-aa distance in *N. pienaari* one evolutionarily conserved aa is deleted suggesting convergent evolution.

Another interesting gene under positive selection is *ikbip* (I Kappa B Kinase Interacting Protein), a pro-apoptotic gene (Hofer-Warbinek et al., 2004) downregulated in skin aging. Apoptosis is relevant for both diapause and aging. Diapausing killifish embryos are resistant to apoptosis (Meller and Podrabsky, 2013), but apoptosis is induced in aging *N. furzeri* (Di Cicco et al., 2011; Ng'oma et al., 2014). Apoptosis-related genes were shown to be age-regulated across tissues in a meta-analysis of mammalian aging (de Magalhães et al., 2009). Studies of larger taxonomical samples, including genomic and transcriptomic sequence datasets, are needed for further investigation of positive selection and convergent evolution in Nothobranchius species.

Overlap of Transcriptional Changes in Developmental Arrest and Aging

Last, we assessed the potential relation between developmental arrest (diapause) and aging in *N. furzeri*. Focusing on diapause II at the somite stage, we determined gene expression changes between arrested and non-arrested embryos at a comparable morphological stage using RNA-seq (Figure 5A). We identified 1,256 down- and 971 upregulated genes in arrested GRZ and MZM-0403 embryos (FDR < 0.05, DEseq and edgeR; Data S4O). In the set of downregulated genes, pattern specification processes including embryonic development of different organs and processes associated with cell proliferation were enriched (p < 0.05, hypergeometric test). Processes enriched in upregulated genes were more diverse and included translational elongation, ribosome biogenesis, metabolism and regulation of cellular component movement (Figure S5A). Decreased rates of cell proliferation and changes in the metabolic status have also been observed in diapause embryos of the South American killifish *Austrofundulus limnaeus* (Podrabsky and Culpepper, 2012). Upregulation of genes involved in translational and ribosomal processes, however, was unexpected. A possible explanation is the need for immediate cellular activity once environmental conditions trigger the exit from diapause.

We then analyzed whether there were similar gene expression changes in diapause and aging. To this end, we again employed MZM-0410 RNA-seq data (brain, liver, skin; 5/12/20/27/39 weeks; Data S4A) and focused on genes showing a monotonic increase or decrease of transcript levels in aging (Data S4P–S4R). In brain, the number of genes that were either down- or upregulated in both aging and diapause was significantly higher than the number of genes downregulated in brain aging and upregulated in diapause or vice versa (p < 0.001, chi-square test; Figure 5B). We therefore concentrated on the first two groups with highest DEG numbers and found that all significantly enriched processes in the group of downregulated genes were associated with cell-cycle progression and DNA replication (Figure 5C). Previous work suggests that brain aging in *N. furzeri* is associated with reduced mitotic activity of adult neuronal stem cells (Tozzini et al., 2012). Unexpectedly, the

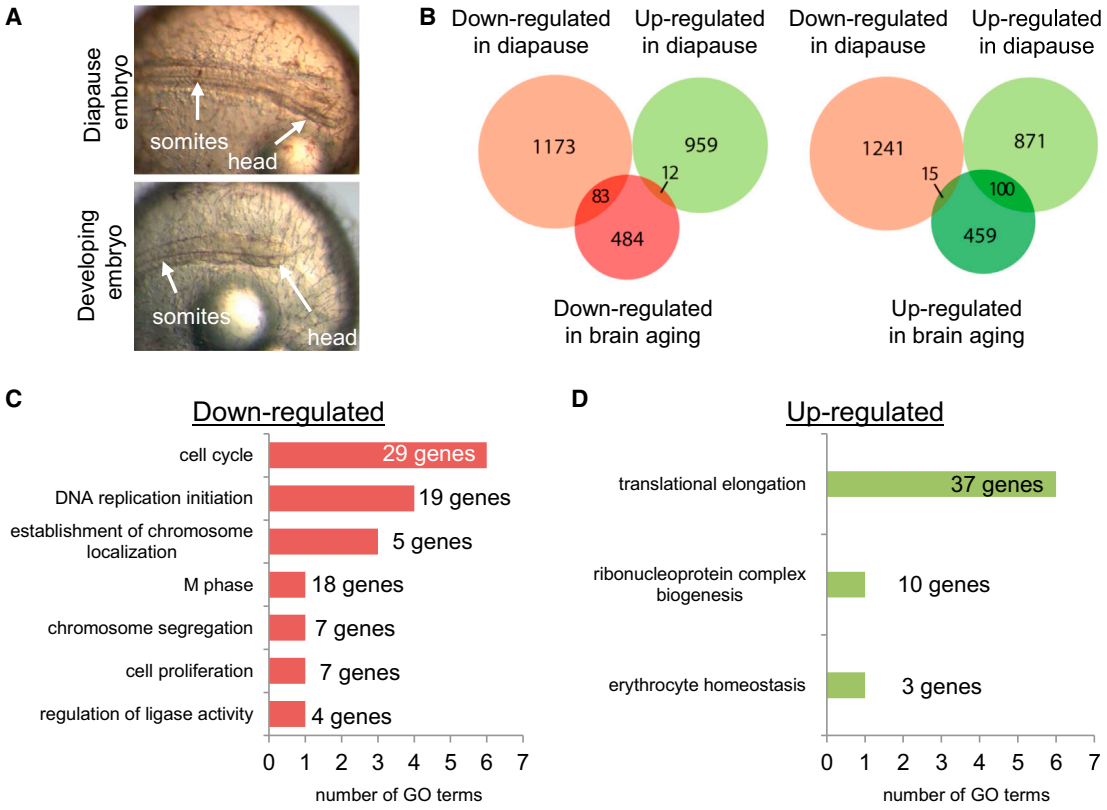


Figure 5. RNA-Seq Analyses of Diapause Embryos and Brain Aging

(A) The embryo (upper picture) has arrested in diapause II for 9 months, whereas the non-arrested embryo (lower picture) exhibiting a comparable morphological stage has an age of 6 days post fertilization.

(B) Venn-analyses of genes downregulated (light red) and upregulated (light green) in diapause as well as monotonic downregulated (dark red) and upregulated (dark green) in brain aging.

(C) Enrichment analyses of genes downregulated in diapause and brain aging. Numbers of involved genes and GO terms are shown for each biological process.

(D) Enrichment analyses of genes upregulated in diapause and brain aging.

See also Figure S5 and Data S4A and S4N–S4U.

two major processes enriched in upregulated genes in diapause and brain aging were translational elongation and ribonucleoprotein complex biogenesis (Figure 5D). The small number of overlapping DEGs between diapause and liver aging prevented further analysis (Figure S5B). Similar to brain, we identified in skin a significantly higher number of genes that were either up- or downregulated both in diapause and aging than genes regulated in opposite ways ($p < 0.001$, chi-square test, Figure S5B). Analysis of consistently downregulated genes showed enrichment of diverse processes. In the respective set of upregulated genes, however, again translational elongation and ribosome biogenesis were enriched (Figure S5C). Previously, aging-related upregulation of genes encoding translational and ribosomal proteins has been reported for human brain, muscle, and kidney suggesting a compensatory mechanism for aging-related increase in protein damage (Zahn et al., 2006). To our knowledge, a common expression profile for vertebrate developmental arrest and aging has not been described before.

In the nematode *C. elegans*, a link between developmental arrest, the so-called dauer larvae, and longevity has been identified. When mutated, some genes affecting dauer formation such as *daf-2* (a homolog of the insulin and IGF-1 receptor) increase lifespan (Kenyon et al., 1993; Lin et al., 1997; Ogg et al., 1997; Shaw et al., 2007). Moreover, the gene expression profile of dauer larvae shows similarities to the expression profile of long-lived adult mutants (McElwee et al., 2004). At present, the absence of long-lived mutants prevents such kinds of analysis in *N. furzeri*. Our comparison of gene expression changes between *N. furzeri* diapause embryos and *C. elegans* dauer larvae (Wang and Kim, 2003) revealed little overlap (Data S4V). This does not seem surprising, given the long evolutionary distance between the two species and their different habitats. The identification of e.g., *daf-16/FoxO4* being upregulated in embryonic arrest of both species, however, indicates commonalities between the two processes and calls for further analyses, e.g., genomic manipulation of the *FoxO4* locus in *N. furzeri*.

In conclusion, the high-quality draft sequence of the *N. furzeri* genome provided here and the availability of several *N. furzeri* strains that differ in lifespan represent excellent resources for studying and identifying genes involved in aging and longevity.

Furthermore, the novel genomic engineering tools now available in *N. furzeri* such as the CRISPR/Cas system (Harel et al., 2015) will allow the generation of mutant lines at a large scale providing a platform for drug screening and sophisticated models to study aging as well as aging-related and other diseases and to develop novel therapies.

EXPERIMENTAL PROCEDURES

Additional details are provided in the Supplemental Experimental Procedures.

Animal Material

Sample acquisition was carried out in accordance with the "principles of laboratory animal care" and the current version of the German Law on the Protection of Animals.

De Novo Genome Sequencing and Assembly

Two adult female GRZ were sequenced using Illumina technology and assembled with ALLPATHS-LG. In parallel, two adult male GRZ were sequenced using Roche technology; these data served for long-range scaffolding and gap filling. Further, optical mapping (OpGen; http://www.opgen.com) was performed in one adult female GRZ. By combining restriction maps obtained with this procedure and sequence scaffolds, superscaffolds were formed. These were manually ordered in genetic scaffolds based on own genetic maps (Kirschner et al., 2012; Ng'oma et al., 2014). Finally, by synteny analyses in medaka and stickleback, genetic scaffolds were arranged in sgrs.

Repeat Annotation

Repeats are identified by (1) RepeatModeler in the reference sequence, (2) RepeatMasker, RepeatScout (Price et al., 2005) for assembled Sanger sequences generated by whole-genome sample sequencing, and (3) RepARK (Koch et al., 2014) for WGS Illumina reads. Subsequently, libraries were merged in a *N. furzeri*-specific repeat library and finally used to annotate the reference sequence by RepeatMasker and TandemRepeatFinder (Benson, 1999).

Gene Annotation and Identification of Paralogs

Protein-coding genes were annotated based on (1) ab initio gene prediction, (2) protein sequence similarity, and (3) Illumina RNA-seq data. Results were combined into CDS models with EVM and UTRs, and transcripts were constructed with PASA (Haas et al., 2008). Gene symbols and functions were annotated using homologous proteins of medaka, platyfish, stickleback, tetraodon, and zebrafish obtained from Ensembl (Cunningham et al., 2015). InterProScan75 (Zdobnov and Apweiler, 2001) was used to identify protein domains and to retrieve Gene Ontology annotations.

MiRNA genes were identified from Illumina miRNA-seq data. To detect rRNA genes, BLAT searches using known *N. furzeri* rRNA sequences (Reichwald et al., 2009) as queries were performed. In addition, miRNA, tRNA, rRNA, and other non-protein-coding genes were identified using ab initio gene prediction tools.

TGD-derived paralogs were identified with Ensembl Compara. First, *N. furzeri* genes were used to find orthologs in medaka, platyfish, stickleback, tetraodon, and zebrafish. Next, Ensembl gene IDs served as queries in Ensembl Compara to detect pairwise paralogous relationships. Any pair of duplicated genes originating before the teleost split was discarded. Finally, *N. furzeri* genes related to the same orthologous gene were also included.

Genomic Resequencing of *N. furzeri* Strains and Variation Calling

Illumina WGS reads generated for all strains were mapped to the reference sequence with Bowtie2 (Langmead and Salzberg, 2012) (minimum mapping quality score of 11). Regions with alignment gaps were realigned with GATK (McKenna et al., 2010) and duplicate reads marked with Picard Tools (http://picard.sourceforge.net). Sequence variations and genotypes were called with GATK. Selected genomic regions were resequenced in additional specimens by PCR and Sanger technology as described (Reichwald et al., 2009).

Overrepresentation Analysis

Zebrafish orthologs of *N. furzeri* genes were retrieved using BLAST. Human orthologs were fetched with R package orthology. GO enrichment analysis was done using DAVID (Huang et al., 2009) and summarized by REViGO (Supek et al., 2011).

Positional Gene Enrichment

Aging-related DEGs were identified by Illumina RNA-seq. Scan statistics (Glaz et al., 2001) were used to test if an observed accumulation of k DEGs on a sgr containing N genes is likely to happen by chance. The scan statistic S is the maximal k in any interval W of fixed size w (w = 0.1 × N). Subsequently, an overrepresentation analysis for each detected genomic region was performed.

Positive Selection

Protein-coding sequences of *N. kadleci*, *N. korthausae*, *N. kuhntae*, *N. pienaari*, *N. rachovii*, *A. striatum*, and *N. furzeri* were assembled de novo using Illumina RNA-seq data. Prank (Löytynoja and Goldman, 2008) alignments of orthologous CDS were filtered by Gblocks (Talavera and Castresana, 2007) and in-house software. Then, the improved branch-site test of positive selection was applied as described (Zhang et al., 2005). Ka/Ks ratios were calculated for all CDS pairs in the SDR both in total and in 333 nt windows sampled using a step size of 99 nt.

Gene Expression Analysis in Diapause Embryos

In total, 287 diapause and 239 non-diapause embryos were collected at the somite stage. Approximately 30 embryos per state were pooled resulting in eight diapause and eight non-diapause samples. Total RNA was extracted and sequenced by Illumina RNA-seq. Significant DEGs were identified and an overrepresentation analysis was performed.

ACCESSION NUMBERS

The accession number for the *N. furzeri* genome project including genome assembly and NGS data (WGS, RNA-seq, and BAC-seq) reported in this paper is BioProject: PRJEB5837. The accession numbers for the *N. furzeri* GRZ genomic insert end sequences of BACs and fosmids are GenBank: KG817100 to KG959958. The accession number for assembled Sanger WGS sequences is BioProject: PRJNA29535. Accession numbers of individual datasets are given in Data S1, S2, S3, and S4.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and four data sets and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.10.071.

AUTHOR CONTRIBUTIONS

C.E., K.R., and M.P. initiated, managed, and drove the genome project. K.R., N.H., M.Ba., S.T., and M.Gr. prepared the samples. K.R., S.T., and M.Gr. performed the sequencing. A.P., P.K., B.R.D., V.S., and M.H. performed the genome assembly and annotation. P.K., B.R.D., D.C., and J.N.V. performed the repeat analysis. M.Ba., M.Gr., A.C., and M.P. performed the mRNA analysis. I.A., M.Ba., A.P., and A.C. performed the miRNA analysis. K.R., A.W., S.S.B., and T.L. performed the chromosome FISH. K.R., A.P., P.K., M.F., K.S., N.H., M.S., C.E., and M.P. performed the sex chromosome evolution analysis. M.Go. and M.E.T. performed protein structure modeling. J.M.K., F.S., S.Pr., P.K., H.A.K., A.C., and M.P. performed the PGE analysis. A.S., M.Be., A.P., B.R.D., A.C., and M.P. performed the positive selection analysis. N.H., S.Pi., and C.E. performed the diapause analysis. All authors contributed to data interpretation. K.R., A.P., P.K., N.H., M.S., A.C., C.E., and M.P. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Silke Foerste, Ivonne Goerlich, Ivonne Heinze, Christin Hahn, Cornelia Luge, Sabine Matz, Martin Neumann, and Bernd Senf for technical

assistance. We thank Karl Lenhard Rudolph for discussions and Cornelia Platzer for critical reading of the manuscript. This work was supported by the Leibniz Association (WGL: PAKT-2006-FLI to C.E. and M.P., and SAW-2012-FLI to M.P.), the German Research Foundation (DFG: RE 3505/1-1 to K.R., HA 6214/2-1 to N.H., and SFB 1074 project Z1 to H.A.K.), the German Federal Ministry of Education and Research (BMBF: JenAge 0315581A/C to A.C., C.E., and M.P.; 031A099 to M.H.; Gerontosys II, Forschungskern SySTaR, project ID 0315894A to H.A.K.), the European Community's Seventh Framework Program (FP7/2007-2013 under grant agreement 602783 to H.A.K.), and the Italian Ministry of Higher Education (FIRB: RBAP10L8TY to I.A.).

Received: June 3, 2015
Revised: August 11, 2015
Accepted: October 21, 2015
Published: December 3, 2015

REFERENCES

Bartáková, V., Reichard, M., Janko, K., Polačik, M., Blažek, R., Reichwald, K., Cellerino, A., and Bryja, J. (2013). Strong population genetic structuring in an annual fish, Nothobranchius furzeri, suggests multiple savannah refugia in southern Mozambique. BMC Evol. Biol. *13*, 196.

Baumgart, M., Di Cicco, E., Rossi, G., Cellerino, A., and Tozzini, E.T. (2015). Comparison of captive lifespan, age-associated liver neoplasias and age-dependent gene expression between two annual fish species: Nothobranchius furzeri and Nothobranchius korthause. Biogerontology *16*, 63–69.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

Blažek, R., Polačik, M., and Reichard, M. (2013). Rapid growth, early maturation and short generation time in African annual fishes. Evodevo *4*, 24.

Cellerino, A., Valenzano, D.R., and Reichard, M. (2015). From the bush to the bench: the annual Nothobranchius fishes as a new model system in biology. Biol. Rev. Camb. Philos. Soc. Published online April 28, 2015. http://dx.doi.org/10.1111/brv.12183.

Chandra, T., Ewels, P.A., Schoenfelder, S., Furlan-Magaril, M., Wingett, S.W., Kirschner, K., Thuret, J.Y., Andrews, S., Fraser, P., and Reik, W. (2015). Global reorganization of the nuclear landscape in senescent cells. Cell Rep. *10*, 471–483.

Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D., Volff, J.N., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat. Genet. *46*, 253–260.

Clelland, E.S., and Kelly, S.P. (2011). Exogenous GDF9 but not Activin A, BMP15 or TGFβ alters tight junction protein transcript abundance in zebrafish ovarian follicles. Gen. Comp. Endocrinol. *171*, 211–217.

Conboy, I.M., Conboy, M.J., Wagers, A.J., Girma, E.R., Weissman, I.L., and Rando, T.A. (2005). Rejuvenation of aged progenitor cells by exposure to a young systemic environment. Nature *433*, 760–764.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. Nucleic Acids Res. *43*, D662–D669.

de Magalhães, J.P., Curado, J., and Church, G.M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics *25*, 875–881.

Devlin, R.H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture *208*, 191–364.

Di Cicco, E., Tozzini, E.T., Rossi, G., and Cellerino, A. (2011). The short-lived annual fish Nothobranchius furzeri shows a typical teleost aging process reinforced by high incidence of age-dependent neoplasias. Exp. Gerontol. *46*, 249–256.

Dorn, A., Ng'oma, E., Janko, K., Reichwald, K., Polačik, M., Platzer, M., Cellerino, A., and Reichard, M. (2011). Phylogeny, genetic variability and colour polymorphism of an emerging animal model: the short-lived annual Nothobranchius fishes from southern Mozambique. Mol. Phylogenet. Evol. *61*, 739–749.

Dorn, A., Musilová, Z., Platzer, M., Reichwald, K., and Cellerino, A. (2014). The strange case of East African annual fishes: aridification correlates with diversification for a savannah aquatic group? BMC Evol. Biol. *14*, 210.

Ellegren, H. (2003). Levels of polymorphism on the sex-limited chromosome: a clue to Y from W? BioEssays *25*, 163–167.

Genade, T., Benedetti, M., Terzibasi, E., Roncaglia, P., Valenzano, D.R., Cattaneo, A., and Cellerino, A. (2005). Annual fishes of the genus Nothobranchius as a model system for aging research. Aging Cell *4*, 223–233.

Glaz, J., Naus, J., and Wallenstein, S. (2001). Scan Statistics (New York: Springer).

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., et al. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. USA *108*, 1513–1518.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., and van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.

Harel, I., Benayoun, B.A., Machado, B., Singh, P.P., Hu, C.K., Pech, M.F., Valenzano, D.R., Zhang, E., Sharp, S.C., Artandi, S.E., and Brunet, A. (2015). A platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate. Cell *160*, 1013–1026.

Hartmann, N., and Englert, C. (2012). A microinjection protocol for the generation of transgenic killifish (Species: Nothobranchius furzeri). Dev. Dyn. *241*, 1133–1141.

Hartmann, N., Reichwald, K., Lechel, A., Graf, M., Kirschner, J., Dorn, A., Terzibasi, E., Wellner, J., Platzer, M., Rudolph, K.L., et al. (2009). Telomeres shorten while Tert expression increases during ageing of the short-lived fish Nothobranchius furzeri. Mech. Ageing Dev. *130*, 290–296.

Hartmann, N., Reichwald, K., Wittig, I., Dröse, S., Schmeisser, S., Lück, C., Hahn, C., Graf, M., Gausmann, U., Terzibasi, E., et al. (2011). Mitochondrial DNA copy number and function decrease with age in the short-lived fish Nothobranchius furzeri. Aging Cell *10*, 824–831.

Hofer-Warbinek, R., Schmid, J.A., Mayer, H., Winsauer, G., Orel, L., Mueller, B., Wiesner, Ch., Binder, B.R., and de Martin, R. (2004). A highly conserved proapoptotic gene, IKIP, located next to the APAF1 gene locus, is regulated by p53. Cell Death Differ. *11*, 1317–1325.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature *496*, 498–503.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Josso, N., and Clemente, Nd. (2003). Transduction pathway of anti-Müllerian hormone, a sex-specific member of the TGF-beta family. Trends Endocrinol. Metab. *14*, 91–97.

Jubb, R.A. (1971). A new Nothobranchius (Pisces, Cyprinodontidae) from Southeastern Rhodesia. J Am Killifish Association *8*, 12–19.

Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A C. elegans mutant that lives twice as long as wild type. Nature *366*, 461–464.

Kikuchi, K., and Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution. Dev. Dyn. *242*, 339–353.

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

Koch, P., Platzer, M., and Downie, B.R. (2014). RepARK–de novo creation of repeat libraries from whole-genome NGS reads. Nucleic Acids Res. *42*, e80.

Kondo, M., Hornung, U., Nanda, I., Imai, S., Sasaki, T., Shimizu, A., Asakawa, S., Hori, H., Schmid, M., Shimizu, N., and Schartl, M. (2006). Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res. *16*, 815–826.

Kotzsch, A., Nickel, J., Seher, A., Sebald, W., and Müller, T.D. (2009). Crystal structure analysis reveals a spring-loaded latch as molecular mechanism for GDF-5-type I receptor specificity. EMBO J. *28*, 937–947.

Krieglstein, K., Miyazono, K., ten Dijke, P., and Unsicker, K. (2012). TGF-β in aging and disease. Cell Tissue Res. *347*, 5–9.

Lahn, B.T., and Page, D.C. (1999). Four evolutionary strata on the human X chromosome. Science *286*, 964–967.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lin, K., Dorman, J.B., Rodan, A., and Kenyon, C. (1997). daf-16: An HNF-3/forkhead family member that can function to double the life-span of Caenorhabditis elegans. Science *278*, 1319–1322.

Löytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science *320*, 1632–1635.

Luo, S., Kleemann, G.A., Ashraf, J.M., Shaw, W.M., and Murphy, C.T. (2010). TGF-β and insulin signaling regulate reproductive aging via oocyte and germline quality maintenance. Cell *143*, 299–312.

McElwee, J.J., Schuster, E., Blanc, E., Thomas, J.H., and Gems, D. (2004). Shared transcriptional signature in Caenorhabditis elegans Dauer larvae and long-lived daf-2 mutants implicates detoxification system in longevity assurance. J. Biol. Chem. *279*, 44533–44543.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Meller, C.L., and Podrabsky, J.E. (2013). Avoidance of apoptosis in embryonic cells of the annual killifish Austrofundulus limnaeus exposed to anoxia. PLoS ONE *8*, e75837.

Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. Curr. Biol. *16*, 2135–2142.

Morrish, B.C., and Sinclair, A.H. (2002). Vertebrate sex determination: many means to an end. Reproduction *124*, 447–457.

Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., and Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, Oryzias luzonensis. Genetics *191*, 163–170.

Nanda, I., Schories, S., Tripathi, N., Dreyer, C., Haaf, T., Schmid, M., and Schartl, M. (2014). Sex chromosome polymorphism in guppies. Chromosoma *123*, 373–383.

Ng'oma, E., Reichwald, K., Dorn, A., Wittig, M., Balschun, T., Franke, A., Platzer, M., and Cellerino, A. (2014). The age related markers lipofuscin and apoptosis show different genetic architecture by QTL mapping in short-lived Nothobranchius fish. Aging (Albany, N.Y.) *6*, 468–480.

Ogg, S., Paradis, S., Gottlieb, S., Patterson, G.I., Lee, L., Tissenbaum, H.A., and Ruvkun, G. (1997). The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. Nature *389*, 994–999.

Otsuka, F., McTavish, K.J., and Shimasaki, S. (2011). Integral role of GDF-9 and BMP-15 in ovarian function. Mol. Reprod. Dev. *78*, 9–21.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067.

Petzold, A., Reichwald, K., Groth, M., Taudien, S., Hartmann, N., Priebe, S., Shagin, D., Englert, C., and Platzer, M. (2013). The transcript catalogue of the short-lived fish Nothobranchius furzeri provides insights into age-dependent changes of mRNA levels. BMC Genomics *14*, 185.

Podrabsky, J.E., and Culpepper, K.M. (2012). Cell cycle regulation during development and dormancy in embryos of the annual killifish Austrofundulus limnaeus. Cell Cycle *11*, 1697–1704.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21* (*Suppl 1*), i351–i358.

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

Rondeau, E.B., Messmer, A.M., Sanderson, D.S., Jantzen, S.G., von Schalburg, K.R., Minkley, D.R., Leong, J.S., Macdonald, G.M., Davidsen, A.E., Parker, W.A., et al. (2013). Genomics of sablefish (Anoplopoma fimbria): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. BMC Genomics *14*, 452.

Santoni, D., Castiglione, F., and Paci, P. (2013). Identifying correlations between chromosomal proximity of genes and distance of their products in protein-protein interaction networks of yeast. PLoS ONE *8*, e57707.

Shaw, W.M., Luo, S., Landis, J., Ashraf, J., and Murphy, C.T. (2007). The C. elegans TGF-beta Dauer pathway regulates longevity via insulin signaling. Curr. Biol. *17*, 1635–1645.

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE *6*, e21800.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. *56*, 564–577.

Terzibasi, E., Valenzano, D.R., and Cellerino, A. (2007). The short-lived fish Nothobranchius furzeri as a new model system for aging studies. Exp. Gerontol. *42*, 81–89.

Terzibasi, E., Valenzano, D.R., Benedetti, M., Roncaglia, P., Cattaneo, A., Domenici, L., and Cellerino, A. (2008). Large differences in aging phenotype between strains of the short-lived annual fish Nothobranchius furzeri. PLoS ONE *3*, e3866.

Thévenin, A., Ein-Dor, L., Ozery-Flato, M., and Shamir, R. (2014). Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. Nucleic Acids Res. *42*, 9854–9861.

Tozzini, E.T., Baumgart, M., Battistoni, G., and Cellerino, A. (2012). Adult neurogenesis in the short-lived teleost Nothobranchius furzeri: localization of neurogenic niches, molecular characterization and effects of aging. Aging Cell *11*, 241–251.

Tozzini, E.T., Dorn, A., Ng'oma, E., Polačik, M., Blažek, R., Reichwald, K., Petzold, A., Watters, B., Reichard, M., and Cellerino, A. (2013). Parallel evolution of senescence in annual fishes in response to extrinsic mortality. BMC Evol. Biol. *13*, 77.

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

Valenzano, D.R., Kirschner, J., Kamber, R.A., Zhang, E., Weber, D., Cellerino, A., Englert, C., Platzer, M., Reichwald, K., and Brunet, A. (2009). Mapping loci associated with tail color and sex determination in the short-lived fish Nothobranchius furzeri. Genetics *183*, 1385–1395.

Valenzano, D.R., Sharp, S., and Brunet, A. (2011). Transposon-Mediated Transgenesis in the Short-Lived African Killifish Nothobranchius furzeri, a Vertebrate Model for Aging. G3 (Bethesda) *1*, 531–538.

Volff, J.N., Nanda, I., Schmid, M., and Schartl, M. (2007). Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sex Dev. *1*, 85–99.

Wang, J., and Kim, S.K. (2003). Global analysis of dauer gene expression in Caenorhabditis elegans. Development *130*, 1621–1634.

Zahn, J.M., Sonu, R., Vogel, H., Crane, E., Mazan-Mamczarz, K., Rabkin, R., Davis, R.W., Becker, K.G., Owen, A.B., and Kim, S.K. (2006). Transcriptional profiling of aging in human muscle reveals a common aging signature. PLoS Genet. *2*, e115.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847–848.

Zhang, J., Nielsen, R., and Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. *22*, 2472–2479.

Zullo, J.M., Demarco, I.A., Piqué-Regi, R., Gaffney, D.J., Epstein, C.B., Spooner, C.J., Luperchio, T.R., Bernstein, B.E., Pritchard, J.K., Reddy, K.L., and Singh, H. (2012). DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. Cell *149*, 1474–1487.