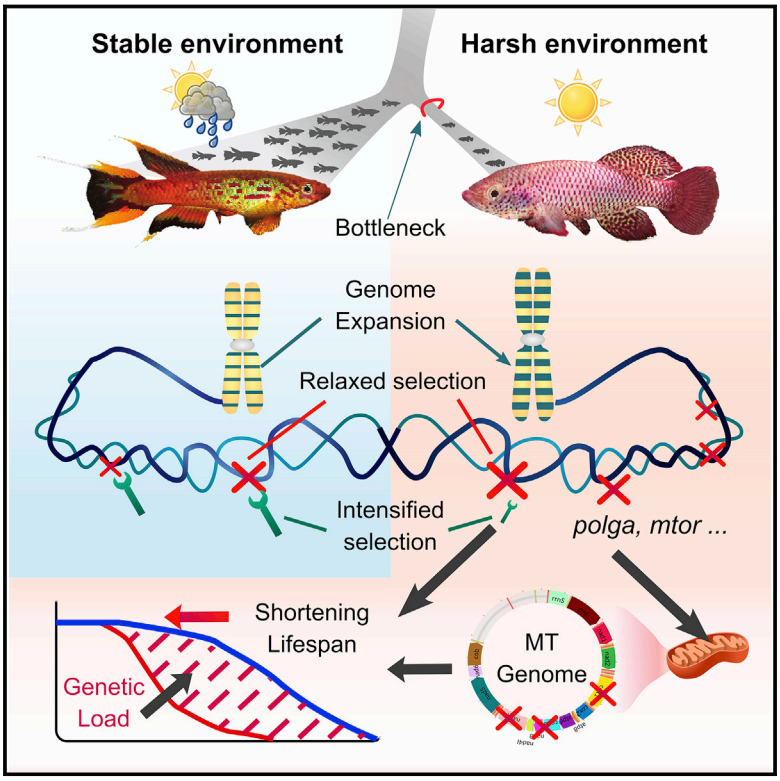
Article

Cell

Relaxed Selection Limits Lifespan by Increasing Mutation Load

Graphical Abstract



Highlights

- Whole-genome sequences of 45 African killifish species
- Relaxed selection leads to larger genomes in short-lived, annual killifish
- Population bottlenecks affect the distribution of aging variants in fish and humans
- Natural *polg* variants in annual killifish cause high rate of mtDNA mutations

Authors

Rongfeng Cui, Tania Medeiros, David Willemsen, ..., Martin Graef, Martin Reichard, Dario Riccardo Valenzano

Correspondence

rcui@age.mpg.de (R.C.), dvalenzano@age.mpg.de (D.R.V.)

In Brief

Studying the genomic changes accompanying the repeated evolution of short lifespan in 45 African killifish species reveals that neutral genetic drift, rather than adaptive evolution, explains the accumulation of deleterious gene variants affecting lifespan and aging.

Cui et al., 2019, Cell *178*, 1–15
July 11, 2019
https://doi.org/10.1016/j.cell.2019.06.004

CellPress

Please cite this article in press as: Cui et al., Relaxed Selection Limits Lifespan by Increasing Mutation Load, Cell (2019), https://doi.org/10.1016/j.cell.2019.06.004

Article

Cell

Relaxed Selection Limits Lifespan by Increasing Mutation Load

Rongfeng Cui,[1,*] Tania Medeiros,[1] David Willemsen,[1] Leonardo N.M. Iasi,[1] Glen E. Collier,[2] Martin Graef,[1,4] Martin Reichard,[3] and Dario Riccardo Valenzano[1,4,5,*]
[1]Max Planck Institute for Biology of Ageing, 50931 Cologne, Germany
[2]Department of Biological Science, University of Tulsa, Tulsa, OK 74104, USA
[3]The Czech Academy of Sciences, Institute of Vertebrate Biology, 603 65 Brno, Czech Republic
[4]CECAD, University of Cologne, 50931 Cologne, Germany
[5]Lead Contact
*Correspondence: rcui@age.mpg.de (R.C.), dvalenzano@age.mpg.de (D.R.V.)
https://doi.org/10.1016/j.cell.2019.06.004

SUMMARY

To uncover the selective forces shaping life-history trait evolution across species, we investigate the genomic basis underlying adaptations to seasonal habitat desiccation in African killifishes, identifying the genetic variants associated with positive and relaxed purifying selection in 45 killifish species and 231 wild individuals distributed throughout sub-Saharan Africa. In annual species, genetic drift led to the expansion of nuclear and mitochondrial genomes and caused the accumulation of deleterious genetic variants in key life-history modulating genes such as *mtor*, *insr*, *ampk*, *foxo3*, and *polg*. Relaxation of purifying selection is also significantly associated with mitochondrial function and aging in human populations. We find that relaxation of purifying selection prominently shapes genomes and is a prime candidate force molding the evolution of lifespan and the distribution of genetic variants associated with late-onset diseases in different species.

INTRODUCTION

Species in nature vastly differ in lifespan, from short-lived mayflies and rotifers, to extremely long-lived clams, whales, and rockfish (Finch, 1994). The genomic underpinnings shaping the evolution of lifespan among species still escape our understanding (Charlesworth, 2009), and while positive selection plays a prominent role in adaptations to new environmental niches, the role of genetic drift in the evolution of lifespan has not been fully explored. To address this gap in knowledge, we investigate the genomic basis of life history trait evolution in African killifishes, a clade of teleosts that colonized a wide range of habitats, ranging from rainforests to arid savanna woodlands, characterized by seasonal water availability and variable annual temperature (Murphy and Collier, 1997). Periodic droughts pose an upper limit to adult survival, shaping life history trait evolution in killifishes, which evolved a set of unique adaptations, such as explosive growth and embryonic diapause (Vrtílek et al., 2018). Annual killifishes are known for their extremely short lifespan, which in the shortest-lived species *Nothobranchius furzeri* (Turquoise Killifish) ranges from 4 to 8 months—the shortest recorded lifespan for a vertebrate in captivity (Valenzano et al., 2015). The short lifespan of annual killifishes is also accompanied by a broad set of aging-related dysfunctions, such as adult onset of cancer, age-dependent locomotor decline (Smith et al., 2017), and neurodegeneration (Cellerino et al., 2016), making them an attractive model system to study vertebrate aging and age-related diseases (Valenzano et al., 2017). In African killifishes, at least two independent evolutionary transitions to annual environments occurred at the ancestral branches of the genera *Nothobranchius* and *Callopanchax* (Murphy et al., 1999) and facultative annualism (semi-annualism) in *Fundulopanchax* and *Aphyosemion kunzi* (Murphy and Collier, 1999). Repeated independent transitions to dry and wet habitats also occurred in closely related *Nothobranchius* populations (Blažek et al., 2017; Valenzano et al., 2015). We investigate the genomic changes that occurred upon repeated adaptations to annual and non-annual environments by adopting an unbiased genome-wide approach between and within species, quantifying the occurrence of both positive and relaxed selection. Applying our approach to Hominini, we find that the occurrence of deleterious genetic variants in chimpanzee reaches higher frequency than in longer-lived human. Analysis of human genomic data additionally reveals that genes under relaxed selection in human populations are enriched for aging and mitochondrial genes. Detection of genes under relaxed selection may help uncover novel, previously undetected, risk variants for aging-related diseases. Together, our results support a general scenario where relaxed selective constraints significantly contribute to shaping mortality rates and aging-related dysfunctions, revealing a predominant role for neutral drift (Kimura, 1968; Ohta, 1973) in shaping life history across species.

RESULTS

Harsh Climate Predicts Genome Expansion in Annual Killifishes

To identify the genomic changes associated with evolution in annual environments, we performed *de novo* genome assembly and annotation of three selected Nothobranchiidae species

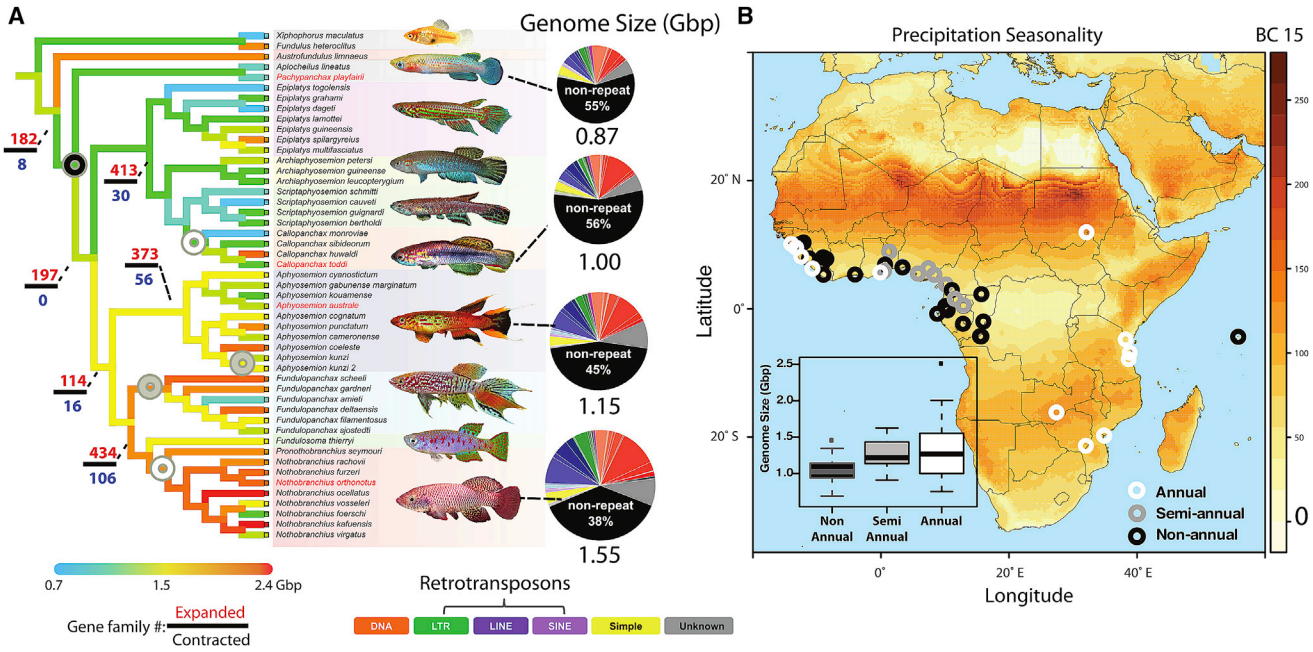


Figure 1. Phylogeny and Distribution of Annualism in African Killifishes

Figure360▷ For a Figure360 author presentation of Figure 1, see https://doi.org/10.1016/j.cell.2019.06.004.

(A) Genomic phylogeny of African killifishes and outgroup species built from concatenated protein-coding nucleotide alignments. Branch color represents genome size inferred by mapping reads to single-copy BUSCO. The number of expanded/contracted gene families is marked on the corresponding nodes (see also Figure S1A). Species names in red were sequenced at high coverage and their genome was assembled and annotated. Pie charts show proportions of repeat types (orange, DNA transposons; green, LTR; dark purple, LINE; light purple, SINE; yellow, simple repeats; gray, unknown repeat type) (see also Figures S1D–S2H). Fish photo credit: Anthony Terceira.

(B) Distribution of sequenced killifishes. White: annuals (*Callopanchax* and *Nothobranchius*); gray, semi-annuals (*Fundulopanchax* and *Aphyosemion kunzi*); black, non-annuals. Color gradient on map corresponds to precipitation seasonality (bioclim variable 15, see Figures S2A and S2B). Inset: boxplot of genome sizes for annuals, semi-annuals and non-annuals. Edges of boxplots indicate the 1st and 3rd quartiles; whiskers are the data range excluding outliers.

See also Figure S3.

(*Nothobranchius orthonotus* [NOR], annual; *Aphyosemion australe* [AAU], non-annual; *Callopanchax toddi* [CTO], annual) and one Aplocheilidae outgroup (*Pachypanchax playfairii*, PLP, non-annual) using Illumina pair-end and mate-pair libraries (Mendeley Data). These draft assemblies have continuity and protein-coding gene benchmark statistics matching previously published teleost genomes (scaffold N50: 1.0–3.3 Mb, BUSCO) (Simão et al., 2015) (completeness of annotated gene sets 95%–97%) (Mendeley Data). Repeat annotation showed a drastic recent expansion of transposable elements (TE) in the annual species NOR compared to its non-annual sister group AAU (Figures S1D–S1H; Mendeley Data). The increase in genome size in *Nothobranchius* is mainly attributable to TEs given that only 589 Mb of the assembly is not masked by repeatmasker, similar to the non-annual outgroups (478–517 Mb), suggesting 61.5% of its 1.53 Gb genome are repeats (Figures 1A and S1D–S1H). By re-sequencing 41 more killifish species originating across sub-Saharan Africa (Mendeley Data), we inferred a highly supported molecular phylogeny with the addition of two non-annual outgroups, *Xiphophorus maculatus* (Schartl et al., 2013) and *Fundulus heteroclitus* (Reid et al., 2016), and a South American annual killifish, *Austrofundulus limnaeus* (Wagner et al., 2018; Figure 1A; Mendeley Data). We detected a significantly positive correlation between "annual" life style and five bioclim variables (Kriticos et al., 2014), including precipitation seasonality (Figures 1B and S2A; Mendeley Data). Additionally, we estimated genome sizes for all 45 species from read coverage in single-copy BUSCO genes (Mendeley Data) and found a positive correlation with annual life style ($p = 0.023$, PGLS, Mendeley Data) and with eight climate variables (Figures S2A and S2B; Mendeley Data), controlling for phylogenetic non-independency (Mendeley Data). Namely, larger genomes are correlated with more variable annual temperatures ($p = 0.0019$) (Figures S2A and S2B) and precipitation ($p = 0.034$) (Figures S2A and S2B). We found that intronic TE expansion (Mendeley Data) was lower in highly expressed genes (Figures S2C–S2H), which are generally under strong selective constraints (Lawrie et al., 2013). Hence, our findings suggest that genome expansion in species with an annual life cycle is likely due to a genome-wide weakening of selective constraints (Lefébure et al., 2017).

Estimating branch-specific gene birth/death rates for 11 fish species from Ensembl and for the four *de novo* assembled killifish genomes (Mendeley Data), we found that genome size expansion in the annual genus *Nothobranchius* is not accompanied by a net expansion of gene family size despite a faster gene turn-over rate (Figures 1A, S1A, and S1C; Mendeley Data). Gene Ontology (GO) analyses show that the expanded gene families in

NOR are enriched for developmental and reproductive functions, such as cell cycle (p = 8.52e−236, false discovery rate [FDR] = 1.01e−233, fold = 12.93), G-protein receptors (driven by olfactory receptors, p = 1.11e−8, FDR = 6.73e−8, fold = 2.55), and female gametogenesis (driven by *zona pellucida* proteins, p = 3.79e−5, FDR = 1.50e−4, fold = 6.62) (Mendeley Data). Contracted gene families are enriched for immune system processes (p = 2.79e−46, FDR = 6.61e−44, fold = 3.55), including B cell-mediated immunity (p = 5.78e−21, FDR = 1.37e−19, fold = 10.18), but also for reproduction and olfactory receptors, suggesting the latter two classes of gene families experienced both gene losses and gains (Mendeley Data).

Relaxation of Selection Is Rampant in Annual Killifishes

To further investigate the occurrence of relaxation of selection, we determined orthology of 14,709 killifish genes with stringent quality control (Mendeley Data), and used the RELAX (Wertheim et al., 2015) descriptive model to infer a relaxation parameter k for every gene in every species (Mendeley Data**;** k > 1 indicates intensified selection, i.e., positive or purifying selection, k < 1 suggests relaxed selection). The species-wise median values of k negatively correlated with genome size (p = 0.00279, PGLS, Mendeley Data)—supporting that genome size expansion is correlated with relaxed selection—and significantly correlated with precipitation and temperature variables in the predicted directions (Figures S3A and S3B; Mendeley Data). Overall, genome size increases while purifying selection weakens in drier and cooler climates, as well as in regions with larger seasonal fluctuations of temperature and precipitation (Figures S2A and S3A).

Next, we used RELAX to detect genes under relaxed or intensified selection (Mendeley Data). Compared to their closest-related non-annual genera, the annual genera *Nothobranchius* and *Callopanchax* have relatively more genes under relaxed selection (k < 1) than under intensified selection (k > 1) (resampling p < 0.001) (Figures 2A and 2B), confirmed by accounting for test performances (Mendeley Data). We estimate that 39.7% of the genes in *Nothobranchius* and 17% in *Callopanchax* are under relaxation of selection (Mendeley Data)—these estimates become 36.4% and 24.8% when excluding codons with multiple differences (CMDs; Venkat et al., 2018) to minimize false calls for positive selection (Mendeley Data). Genes under intensified purifying selection were estimated to be near zero for both annual genera, compared to 2%–6% in non-annual outgroups (Mendeley Data). Hence, while both relaxed selection and intensified selection likely occurred, relaxation of selection dominates genome evolution in annual species. Despite a high number of shared genes under relaxed selection between the two annual genera, the overlap is expected by chance (Figure S4G, p = 0.59, Fisher's exact test, p = 0.71 excluding CMDs, Figure S5L), suggesting no detectable convergent evolution of genes under relaxed selection, although convergence on a functional pathway level exists (Figure S7A). The overlap of genes under intensified selection between *Callopanchax* and *Nothobranchius* was only significant (Figure S4H) when CMDs were included (Figure S5K; Mendeley Data). RNA sequencing (RNA-seq) (Figure 2C; Mendeley Data) and codon usage bias (Figure S4A; Mendeley Data), an indicator of translational optimization (Hershberg and Petrov, 2008), show that highly expressed genes have larger "k" value (i.e., are under stronger selective constraint) (Lawrie et al., 2013). Based on differential force of natural selection predicted by life history theory, genes contributing to early life fitness are expected to be under stronger selection (Charlesworth, 2000). Consistent with this prediction, genes differentially upregulated in young *Nothobranchius furzeri* (brain and skin) have larger k values—hence are less relaxed (Figures 2E, 2F, and S4C–S4F; Mendeley Data), supporting that genes highly expressed during development are less prone to relaxed selection upon evolutionary transition to annualism. The fitness cost of the private mutations carried in the genes under relaxed selection in *Nothobranchius* can be measured by their degree of conservation in distantly related proteins in other taxa, such as the Consurf score (Ashkenazy et al., 2016). Indeed, genes under relaxed selection have a significantly higher median (Figure 2D; Mendeley Data; Wilcoxon rank-sum test, p = 8.7e−6) and maximum Consurf scores (Figure S4B)—unlikely driven upward by sites under positive selection (Mendeley Data)—suggesting that amino acid changes in the relaxed genes have likely higher phenotypic severity. All correlations are robust to CMD exclusions (Mendeley Data). Besides repeat expansion, our results show that annual species also have a significantly larger proportion of genes under relaxed selection compared to the non-annual counterparts, and relaxation of selection can lead to fixation of deleterious mutations in genes that are otherwise highly conserved.

Positive Selection of Developmental and Reproductive Genes

We scanned all 46 genomes for positively selected genes using the branch-site test implemented in CodeML (Yang, 2007; Mendeley Data). We estimated fewer positively selected genes (5.16% in *Nothobranchius* and 0.45% in *Callopanchax*) compared to genes under relaxation, after accounting for power, false-positive rate and CMD sites (Mendeley Data). The number of positive selected genes is lower than in *Drosophila* (six species, ~28% tests, p < 0.05) and higher than in the human branch (one species, ~0.03% tests, p < 0.05) (Venkat et al., 2018).

In harsh environments, positive selection is expected to be stronger in genes controlling early life history stages, including genes involved in development and reproduction (Partridge and Barton, 1993). Indeed, our GO analysis supports that development and gonadotropin GO terms are enriched in annual species (Figures S7B and S7C; Mendeley Data). Numerous genes related to development (e.g., *egf* and *fshr*), growth (e.g., *igf2r*, *ghrb*, *trh*, and *tshr*), and reproduction (e.g., *vtg5*, *zp3a.1*, and *zp3c*) were positively selected (Mendeley Data). The mitochondrial translation pathway was enriched for positively selected genes in both *Callopanchax* and *Nothobranchius* regardless of CMD exclusion, in agreement with a previous scan in a smaller dataset in *Nothobranchius* (Sahm et al., 2016; Mendeley Data). However, most of the genes involved in mitochondrial translation also showed molecular signatures of relaxation of selection. After excluding genes detected with both relaxation of selection (k < 1, p < 0.05) and positive selection, the mitochondrial-related pathways were no longer enriched (Mendeley Data), suggesting that positive selection in mitochondrial genes may be compensatory for accumulated deleterious mutations. Alternatively,

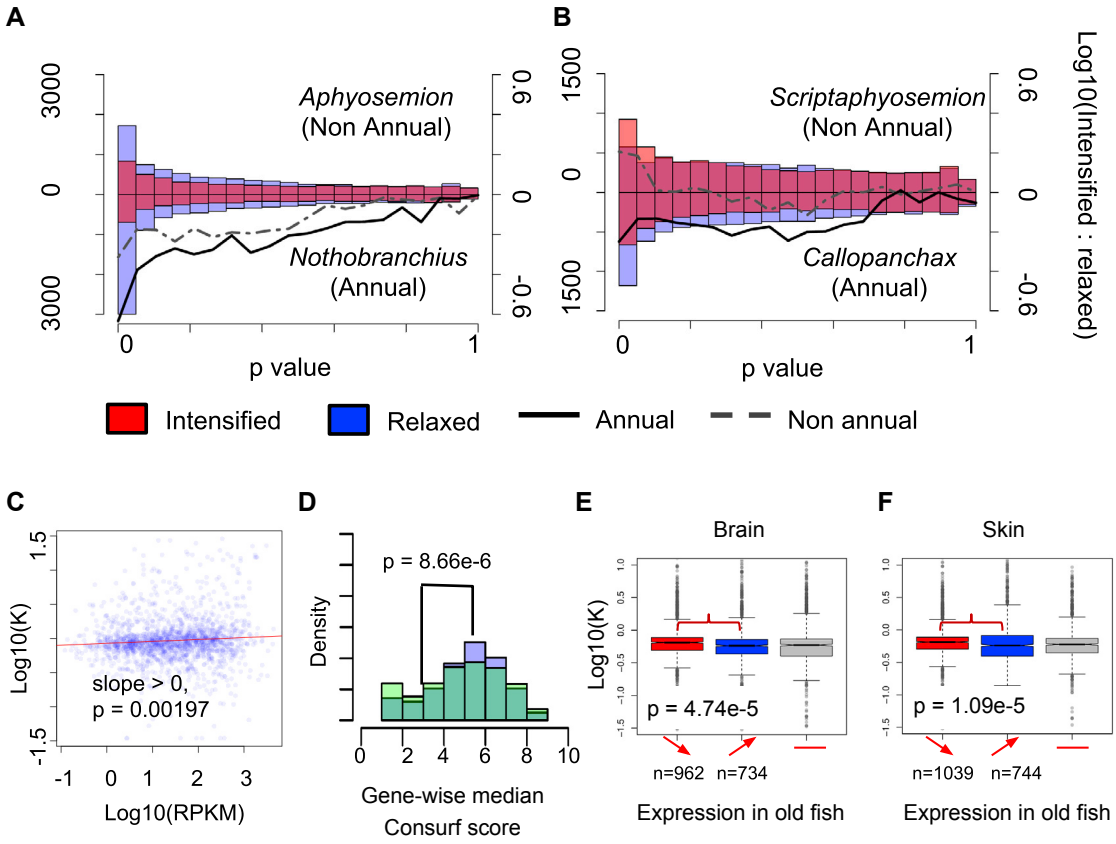


Figure 2. Relaxation of Selection in Annual Killifishes

Figure360▷ For a Figure360 author presentation of Figure 2, see https://doi.org/10.1016/j.cell.2019.06.004.

(A and B) p value distributions of the RELAX tests of non-annual (top half) and annual (bottom half) sister genera. (A) Contrast between *Nothobranchius* and *Aphyosemion* and (B) Contrast between *Callopanchax* and *Scriptaphyosemion* are shown. Blue bars, genes under relaxed selection ($k < 1$); red bars, intensified genes ($k > 1$). Dotted (non-annual) and solid (annual) lines are the log ratios of $k > 1$: $k < 1$ gene counts (see also Figures S5G and S5H).

(C) Average gene expression level is positively correlated with the relaxation parameter k (see also Figure S4A); p value determined by linear regression.

(D) Distributions of the gene-wise median Consurf scores in relaxed (blue bars) and non-significant (green bars) genes (see Figure S4B for gene-wise maximum scores); p values determined by Wilcoxon rank-sum test.

(E and F) Genes highly expressed during young age (DESeq, $p < 0.05$) in brain (E) and skin (F) experienced less relaxed selection in brain and skin (see Figures S4E and S4F for gene expression clustering of brain and skin samples and Figures S4C and S4D for liver). Edges of boxplots indicate the 1st and 3rd quartiles; whiskers are the data range excluding outliers. p values determined by Wilcoxon rank-sum test.

See also Figure S5.

adaptive positive selection in mitochondrial function can be facilitated by earlier relaxation of selection (Wright, 1932).

In line with the analysis of genes under intensified selection in both annual genera *Nothobranchius* and *Callopanchax*, the overlap of 79 positively selected genes (Figure S4I) ($p = 1.83e{-}5$, Fisher's exact test) was not robust to CMD site exclusion ($p = 0.24$, Fisher's exact test) (Figure S5M; Mendeley Data)

Genes under positive selection in annual environments are associated with growth, reproduction, and egg structures, likely playing an important role in the unique adaptations of annual fish to their extreme environments.

Convergent Amino Acid Substitutions in Relaxed and Positively Selected Genes

We scanned the genomes for convergent amino acid substitutions exclusively shared by at least two species of *Nothobranchius* and *Callopanchax* (Mendeley Data, FDR by simulation ~50%) and found that genes under relaxed selection were enriched for convergent amino acids before and after CMD exclusion (before exclusion: 14/185, $p = 0.0015$, after exclusion: 25/255, $p = 4.067e{-}07$). Enrichment for convergent amino acids in positively selected genes was still present after CMDs exclusion (before CMD exclusion: 11/79, $p = 3.66e{-}5$, Fisher's exact test; after CMD exclusion: 2/7, $p = 0.015$). Our findings suggest that convergent amino acid substitutions occur also in the context of relaxation of selection.

Populations Repeatedly Underwent Relaxed Selection in Dry Regions

Next, we asked whether relaxation of selection also affected the microevolutionary scale. We contrasted paired populations of the *N. rachovii* and *N. orthonotus* species complex (Figure 3B;

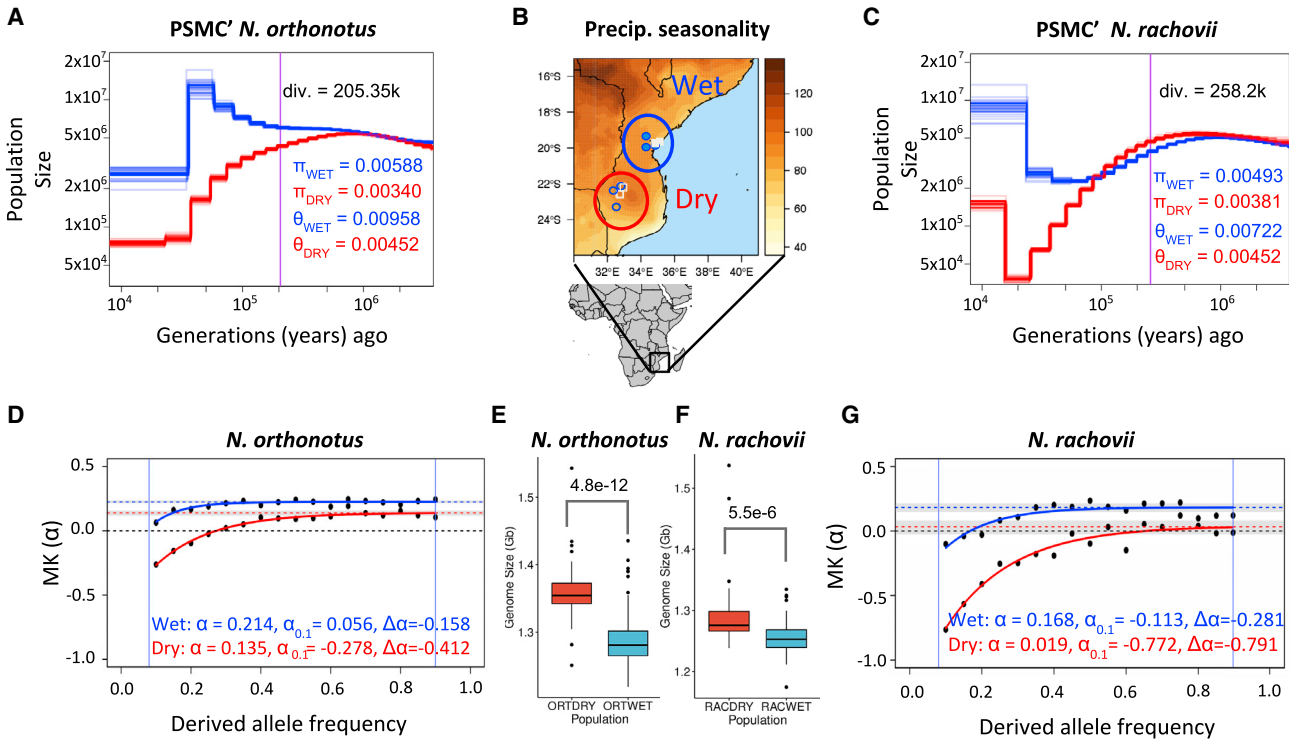


Figure 3. Relaxed Selection within Species in a Two-Species Comparison

Figure360▷ For a Figure360 author presentation of Figure 3, see https://doi.org/10.1016/j.cell.2019.06.004.

(A and C) Demographic histories inferred by PSMC of two population contrasts from dry (red) and wet (blue) localities. (A) *N. orthonotus* species complex (ORTWET, ORTDRY) and (C) *N. rachovii* species complex (RACWET, RACDRY). Solid lines, inferred demography by PSMC; purple line, divergence estimated with SMC++ (Mendeley Data). Faint lines are 30 bootstrapped runs by resampling 5 Mb chunks of the genome.

(B) Map of the sampling localities of the wet and dry populations. Squares are *N. orthonotus* and circles are *N. rachovii*.

(D and G) Asymptotic McDonald-Kreitman (MK) α statistics binned by derived allele frequencies in the range 5%–90%, in 5% intervals for *N. orthonotus* (D) and *N. rachovii* (G). The dotted horizontal lines are the asymptotic α estimated with a previous method (Messer and Petrov, 2012) (95% CI marked by gray area), representing an unbiased estimate of the proportion of mutations driven to fixation by positive selection. The level of relaxation ($\Delta\alpha$) is measured by the difference between the asymptotic α and the α value measured at the 5%–10% frequency bin ($\alpha_{0.1}$), where slightly deleterious polymorphisms are enriched.

(E and F) Both "dry" populations have a larger genome size on average than their "wet" sister populations in *N. orthonotus* (E) and *N. rachovii* (F). p values determined by Wilcoxon rank-sum test.

See also Figure S6.

Mendeley Data), representing recent independent adaptations to wet and dry climates (Blažek et al., 2017). We sequenced ~58 individuals per population to ~2.7× per individual for a total of 231 individuals at 638× coverage, and one additional individual per population at 25–37× coverage (total, 119×) (Mendeley Data). We first inferred the historical demography from the high-coverage individuals using pairwise sequentially Markovian coalescent (PSMC) and the split time from all samples with SMC++ (Mendeley Data). All populations experienced bottlenecks between 10 and 100 killoannum (ka) and populations from dry climates experienced more severe bottlenecks than their wet counterparts (Figures 3A and 3C). The two *N. rachovii* populations diverged earlier, around 258.2 ka, while the two *N. orthonotus* populations diverged 203.35 ka. Strikingly, the estimated genome sizes are larger in both populations from dry environments compared to their "wet" sister populations (Figures 3E and 3F), recapitulating the phylogenetic pattern (Figure S2B). We computed the asymptotic McDonald-Kreitman α (Messer and Petrov, 2013), defined as the proportion of sites driven to fixation by positive selection since divergence from the outgroup species *Fundulosoma thierryi* (*FTH*), pooling all sites in the protein-coding genes (Mendeley Data). Among the four populations, the asymptotic α varied between 0.019–0.214, with a lower asymptotic α for both species in populations from drier climates. The α values at lower allele frequencies (5%–10% derived allele frequency) are much more negative in the two dry populations (Figures 3D and 3G), possibly reflecting segregation of more slightly deleterious mutations (Messer and Petrov, 2013) in the populations from dry climates. This result is unchanged when another outgroup species (*Nothobranchius virgatus* [NVG]) was used (Mendeley Data). The distribution of fitness effects (DFE), estimated with Anavar (Barton and Zeng, 2018), showed that dry populations have significantly more mutations assigned to the slightly deleterious category, assuming a range of sequencing errors (Mendeley Data).

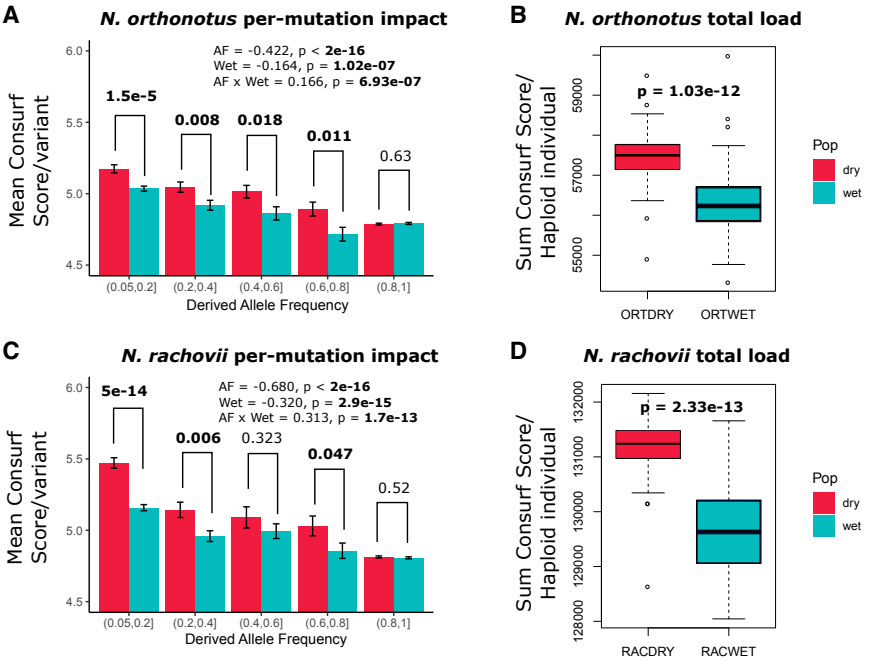


Figure 4. Deleterious Mutations Segregating at High Derived Allele Frequencies

For a Figure360 author presentation of Figure 4, see https://doi.org/10.1016/j.cell.2019.06.004. Figure360▷

Averaged Consurf scores binned into five derived allele frequency bins (AF <0.05 excluded).

(A and C) *N. orthonotus* (A) and *N. rachovii* (C), polarized by the outgroup FTH, Wilcoxon rank sum tests were used to compare each bin. Parameter estimates for the full linear model Consurf ~AF * Climate are reported. Error bars indicate ± SEM.

(B and D) The total genetic load per haploid individual is calculated by summing Consurf scores for a randomly chosen allele per site in each sampled individual for *N. orthonotus* (B) and *N. rachovii* (D). Wilcoxon rank-sum tests were used to compare between groups. Edges of boxplots indicate the 1st and 3rd quartiles; whiskers are the data range excluding outliers.

Deleterious Variants Drifting to High Frequencies Are Related to Aging

To further investigate the frequency distribution of the deleterious mutations, we contrasted the Consurf scores at different allele frequencies for the dry and wet population pairs. We found that on average the non-synonymous polymorphisms are significantly more deleterious—based on the Consurf score—from the lowest frequencies up to an allele frequency of 60%–80% for both dry populations (Figures 4A and 4C), indicating that deleterious gene variants reach high frequencies in populations from drier environments. The elevation of mean Consurf scores is unlikely driven by positively selected sites (Mendeley Data). Individuals in both dry populations also carry more total load compared to their wet counterparts (Figures 4B and 4D).

We then asked whether the frequency distribution of deleterious alleles relates to lifespan and aging in other taxa, such as Hominini. Interestingly, the lifespan of humans is approximately twice that of chimpanzees in captivity (Altschul et al., 2018), despite short divergence. We found that chimpanzees have more deleterious mutations reaching very high frequencies compared to human, some of which are completely fixed (Figure 5A), although it is also possible that the significant difference in the derived allele frequency bin in the range (0.8, 1] is influenced by the different probability of misidentifying the ancestral states (Hernandez et al., 2007). The genus *Pan* has higher per-individual load compared to *Homo* (Figure 5B), despite its larger historical population size (Prado-Martinez et al., 2013). Although life history could, in part, explain the differences in genetic load between *Pan* and *Homo*, other factors, such as higher variance in reproductive success in male Chimpanzees and recent crashes in wild *Pan* populations (Prado-Martinez et al., 2013) may also play a significant role. More overall efficient selection in the human branch (e.g., for genes contributing to delayed sexual maturation) may have removed more segregating deleterious alleles. Europeans and East Asians, who underwent more severe bottlenecks and weathered more inclement climates (Timmermann and Friedrich, 2016) compared to African populations, carry amino acid variants of higher impact up to intermediate allele frequencies (20%–40%) (Mendeley Data; Henn et al., 2015; Lohmueller et al., 2008). However, the per-individual genetic load is very similar between different human populations as previously reported (Simons et al., 2014), although Europeans and admixed Americans show statistically significant increase in total load (Figure 5B). Native American population (Mayan) was previously shown to have a higher load than Sub-Saharan Africans using a different measure of deleterious mutations (Henn et al., 2016). The 1000 Genomes American samples have an intermediate load (Figure 5B), consistent with the fact that they are admixed individuals of African, European and Native American ancestries (Montinaro et al., 2015). Despite humans being a long-lived primate, we asked whether relaxed selection is related to aging. We found that genes upregulated during human frontal lobe development (Colantuoni et al., 2011) have higher direction of selection (DoS) (Eyre-Walker and Keightley, 2007; Figure 5C) compared to genes expressed later, indicating that selection is stronger for genes expressed in early life, similar to killifish (Figures 2E and S6D–S6G). Out of 26 human populations, 21 have significantly down-shifted DoS scores for aging genes, and all populations have down-shifted DoS scores in mitochondrial genes (Figure 5D, FDR <0.2), indicating a link between relaxation of selection on the one hand and aging and mitochondrial genes on the other hand. Despite the lower power for the analysis in chimpanzee due to fewer available polymorphisms, the mitochondrial pathway shows a significant down-shift in DoS in two *Pan* populations. Interestingly, we found that genes involved in neurological system process have intensified selection in 21 human populations, highlighting possible functions that might have recently undergone adaptive evolution in our lineage

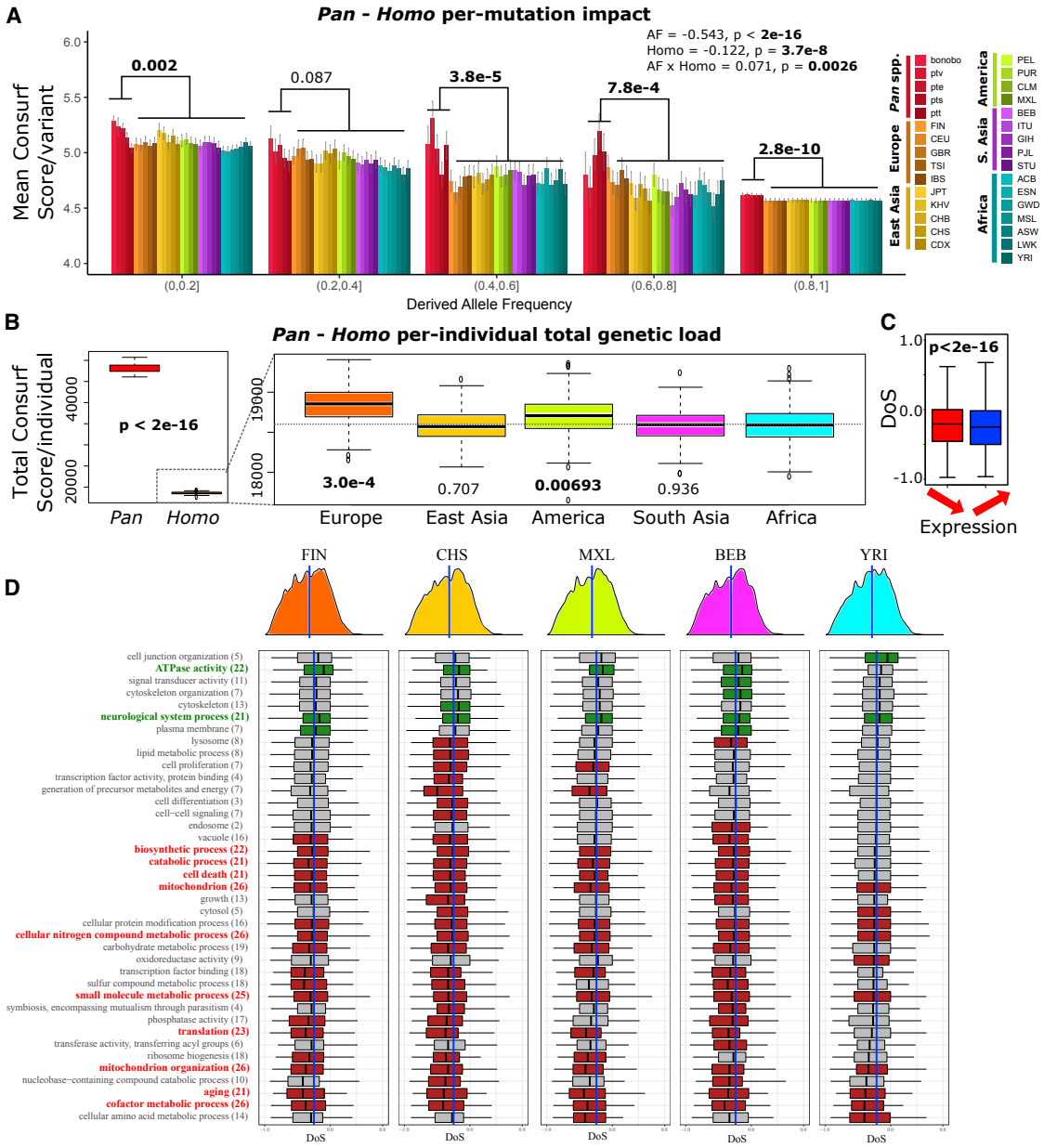


Figure 5. Genetic Load in *Homo* and *Pan*

Figure360▷ For a Figure360 author presentation of Figure 5, see https://doi.org/10.1016/j.cell.2019.06.004.

(A) Per mutation impact binned by derived allele frequencies in 5 chimpanzee and 26 human populations, a *Gorilla* sequence was used as the outgroup (see also Mendeley Data). At each frequency bin, the mixed effect model Consurf ~Genus + (1|Pop) was used to compare the differences. Parameter estimates are from the mixed effect model Consurf ~AF * Genus + (1|Pop). Error bars are ± SEM.

(B) Comparisons of per individual genetic load, measured using the summed Consurf scores, of *Pan* and different human populations. Statistics performed using the same mixed effect model as in (A).

(C) Genes highly expressed during human frontal lobe development have higher DoS in human populations. Statistics performed using the same mixed effect model as in (A).

(D) Distribution of DoS scores of genes in different GO pathways in representative human populations from each continent. Only terms significant (FDR <0.2) in at least one of the selected population are shown, number of significant populations (FDR <0.2) in parentheses. Terms significant in at least 20 populations are bolded. Green box, intensified; red box, relaxed (FDR <0.2).

(Figure 5D). Overall, we identified 453 positively selected sites in killifishes, while 2,862 sites were candidate relaxed in killifishes, 227 in chimpanzees, and 137 in humans (Mendeley Data). By comparing the Europeans or East Asians with Africans, we further identified 726 candidate deleterious variants in Europeans, 812 in East Asians, and 2,218 in Africans (Mendeley Data). We overlapped these lists with ClinVar, a reference database of clinical disease variants (Landrum et al., 2016), showing that 88 of these candidate sites in Europeans, 71 in East Asians, and 263 in Africans are annotated with known human diseases (Mendeley Data). As a complementary approach, we also report genes with a significantly deviating DoS, as well as the genes within the significant GO terms (Mendeley Data). This resource provides a list of known, as well as potentially yet undiscovered, late-acting, disease-associated genetic variants.

Overall, our results suggest that recent demography (Lohmueller et al., 2008; Simons et al., 2014) and the decreased strength of selection against late-life acting mutations, together affect the frequency distribution of deleterious mutations, which can either remain in the population at low to medium frequency or become, in extreme cases, the major allele dramatically affecting life history traits, including species-specific lifespan and aging.

Characterization of the Evolutionary Dynamics of Relaxed Genes

To understand how selection changes on individual genes across time, we computed the DoS statistics for each gene (Mendeley Data), which is an unbiased estimator that measures both direction and extent of selection, being positive under adaptive evolution and negative for deleterious segregating mutations (Eyre-Walker and Keightley, 2007). We found that genes under relaxed selection ($k < 1$, $p < 0.05$ in the RELAX test) in the ancestral branch of *Nothobranchius* are more likely to have an above-average DoS (Figures S6A and S6B; derived allele frequency 0.20–0.95, outgroup NVG; Mendeley Data). Hence, ancestrally relaxed genes were less likely to experience further relaxation in more recent timescales, which may prevent functionally indispensable genes from pseudogenizing (Brawand et al., 2014; Wright, 1932). Further, ancestrally positively selected genes are more likely to have positive DoS (Figure S6C), consistent with persistent positive selection driving subsequent substitutions. To directly analyze the relation between gene-wise divergence and selection, we computed the gene-wise fixation index F_{ST} (Weir and Cockerham, 1984) between the two *N. rachovii* populations and compared it with the DoS score. We found that the gene-wise F_{ST} was negatively correlated with DoS in the more divergent *N. rachovi* species pair (Figures S6N and S6O; Mendeley Data), but not so for the more closely related *N. orthonotus* populations (Figures S6L and S6M), consistent with purifying selection slowing population divergence. Purifying selection is predicted to be more efficient in highly recombinant regions due to effective disassociation between deleterious and beneficial alleles (Stoletzki and Eyre-Walker, 2011). Indeed, in all four populations, inferred local recombination rate positively correlates with DoS (Figures S6Q, S6R, S6T, and S6U; Mendeley Data), suggesting genes located in regions with low recombination rate are particularly susceptible to relaxed selection.

Accumulation of Deleterious Mutations in Genes Contributing to Longevity

In agreement with the phylogenetic dataset (Figures 2E and 2F; Mendeley Data), genes upregulated in young fish are less relaxed in brain and skin (Figures S6D–S6K) also in population data. Hence, genes expressed in late-life, which generally experience weaker constraints (Charlesworth, 2000; Šajina and Valenzano, 2016), are also more susceptible to relaxed selection during evolutionary transitions to annual environment. Together, our data support that purifying selection is weaker throughout the genome in dry environments—where populations are under higher extrinsic mortality—and is more relaxed in genes active during late life stages, such as aging-related genes.

GO analyses (Figure S7A; Mendeley Data) of shared genes under relaxed selection in *Nothobranchius* and *Callopanchax* revealed relaxation in several key aging pathways, namely mitochondrial RNA metabolism ($p = 6.32e{-12}$, FDR = 2.47e−9) and organization ($p = 2.25e{-8}$, FDR = 1.15e−5), insulin binding ($p = 0.0041$, FDR = 0.13), determination of adult lifespan ($p = 0.0017$, FDR = 0.18), and damaged DNA binding ($p = 0.0025$, FDR = 0.046). The distribution of DoS scores for the genes in the GO term "mitochondrion" in both dry populations is significantly more negative than in either wet population (Mendeley Data), consistent with additional deleterious mutations recently accumulated in genes involved in mitochondrial functions in dry populations. Key aging genes in the insulin-IGF1 and mTOR pathways (Johnson et al., 2013), namely *mtor* (*Nothobranchius*: $p = 6.50e{-6}$, $k = 0.70$; *Callopanchax*: $p = 0.040$, $k = 0.78$), *insr* (*Nothobranchius*: $p = 0.023$, $k = 0.69$; *Callopanchax*: $p = 0.0029$, $k = 0.49$), and *foxo3a* (*Nothobranchius*: $p = 0.0053$, $k = 0.70$; *Callopanchax*: $p = 0.026$, $k = 0.69$) are under relaxed selection in both annual genera. *Nothobranchius* have additional genes interacting with *mtor* under significant relaxed purifying selection, namely *deptor* ($p = 0.0367$, $k = 0.70$), *mapkap1* ($p = 0.022$, $k = 0.71$), *tti1* ($p = 0.025$, $k = 0.57$), *rps6kb1b* ($p = 0.034$, $k = 0.0129$), *pras40* (*akt1s1* $p = 0.010$, $k = 0.64$), two *ampk* (*prkag1*, $p = 0.049$, $k = 0.67$; *prkab1a*, $p = 0.0082$, $k = 0.65$), *eif4ebp1* ($p = 0.0074$, $k = 0.40$), and *foxo3b* ($p = 0.013$, $k = 0.66$) (Mendeley Data). Overall, upon transition to annual environments, key genes and pathways involved in aging regulation were extensively affected by relaxation of selection, leading to accumulation of deleterious functional variants. We also found that human populations, despite our species being long-lived, are susceptible to significant relaxation of selection in the aging and mitochondrial pathways (Figure 5D).

Impaired DNA Replicative Accuracy Increases Mutation Load

DNA repair is a key process involved in maintaining DNA homeostasis, preventing age-related diseases, including cancer (Jackson and Bartek, 2009). Noteworthy, both genes in the canonical mismatch repair recognition complex MutSα (*msh2*/*msh6*) (Jackson and Bartek, 2009) are under relaxed selection in *Nothobranchius* (*msh2*: $p = 0.0027$, $k = 0.62$; *msh6*: $p = 1.8e{-4}$, $k = 0.65$) and *Callopanchax* (*msh2*: $p = 4.1e{-4}$, $k = 0.46$; *msh6*: $p = 0.0038$, $k = 0.58$). Other genes involved in DNA replication and repair (Branzei and Foiani, 2008; Jackson and Bartek, 2009), namely *tep1*, *rpa1*, *lig1*, *nsmce4a*, *top3a*, and

polm are also under relaxed selection in both genera. *Nothobranchius* has additional genes involved in DNA repair that are under significantly relaxed selection, such as *rad51* ($p = 0.0069$, $k = 0.0095$), *xrcc3* ($p = 0.015$, $k = 0.16$), *msh3* ($p = 0.016$, $k = 0.68$), *pole* ($p = 0.017$, $k = 0.87$), *poli* ($p = 0.0059$, $k = 0.27$), and *xrcc4* ($p = 0.044$, $k = 0.081$). Interestingly, we found that the branch length of *Nothobranchius* (but not *Callopanchax*) at 4-fold synonymous sites is significantly longer than other killifishes ($p < 1e{-}16$, LRT, BaseML molecular clock test), suggesting that DNA replicative errors may have caused a higher SNP substitution rate in the germline, in addition to genome expansion.

Mitochondrial DNA replication is carried out by POLG (POLGα-β) both in the germline and in the soma (Figure 6A). Deleterious mutations in *polga* are known in human mitochondrial diseases and are often involved in mtDNA depletion or in the accumulation of large-scale mtDNA deletions (Gorman et al., 2016). Deleterious mutations in mouse *polga* increase mtDNA mutational rate and accelerate aging (Trifunovic et al., 2004). We found that *polga* (but not *polgb*) is under significant relaxation of selection in annual fishes of the genus *Nothobranchius* ($p = 3.72e{-}6$) and *Callopanchax* ($p = 0.0049$). Some of the annual-specific amino acid substitutions for *polga* are located in previously characterized functional domains (Figure 6B) that could affect replication accuracy and alter mutation rate in the mitochondrial genome (Farnum et al., 2014). Consistent with this prediction, we found that sequence divergence at the mtDNA synonymous sites is significantly higher in annual killifishes, deviating from a molecular clock (Figure 6E, $p = 0.034$, LRT). To functionally validate the phenotypic effects of relaxed selection for *polga*, we assessed the replicative fidelity resulting from nine annual killifish mutations in a heterologous yeast erythromycin assay, where erythromycin resistance is conferred by increased mtDNA mutational rate (Doudican et al., 2005). Yeast heterologous systems are a powerful tool to test the functional effect of human *polga* mutations (Lodi et al., 2015). Switching the cognate amino acid in the yeast *polg* gene to killifish states (Mendeley Data), we confirmed that at least 6 out of 9 tested amino acid substitutions from annual fish resulted in significantly higher mtDNA mutation rate than the corresponding non-annual variants (Figure 6D). At 5 sites, yeast, non-annual killifishes, and human all share the same amino acid. Among them, 4 out of 5 annual killifish variants caused an elevated mutation rate, whereas one (R7H) showed no difference. Yeast, non-annuals, and annuals each have a different amino acid at three sites. I489T (annual), but not I489K (non-annual), conferred an elevated mutation rate. K781S (annual) and K781A conferred higher mutations than the yeast wild type (WT) variants, but the annual variant is significantly more mutagenic. At site R263, both annual and non-annual variants elevated mutation rate to about the same level, suggesting no difference between them. At site A857, yeast shares the same amino acid as *Nothobranchius*, which differs from other killifishes and human (V857). Strikingly, the non-annual variant (i.e., V857) conferred a lower mutational rate than the yeast WT (i.e., A857), again confirming that V857A in *Nothobranchius* increases mutational rate compared to non-annual killifishes and human. Hence, we confirm that annual killifishes have accumulated slightly deleterious and nearly neutral mutations in *polga*.

We showed that increased mutation rate can exacerbate relaxed selection in regions with low recombination (Figures S6Q–S6U), a problem for the non-recombinant mtDNA (McVean and Charlesworth, 2000). Mirroring results of the nuclear genome, annual killifishes have larger mitochondrial genomes than non-annual killifishes due to expansions in *Nothobranchius* and *Callopanchax* (Figures S7D–S7H, $p = 1.86e{-}11$, PGLS). Importantly, these expansions occurred independently in the two annual genera. In *Nothobranchius*, a segment near the control region underwent tandem duplication and gradually degenerated, leaving in *N. virgatus* a remnant copy of *nad6* (Figure S7F) which is no longer detectable on NOR (Figure S7D) and other species. In *Callopanchax*, the region between the 3′ end of 16S rDNA and *atp6* likely duplicated and degenerated in different species, leaving an nad2 pseudogene (Figure S7E, *C. toddi* and *C. sibideorum*) or an *atp6* pseudogene (Figure S7G, *C. monroviae*). In addition to pseudogenization, nearly all functioning mitochondrion-encoded proteins are under significantly relaxed selection in both annual genera *Nothobranchius* (k ranging between 0.22 and 0.90) and *Callopanchax* (Figure 6C, k ranging between 0.30 and 0.85) with no sign of positive selection, a pattern also confirmed by the population data (DoS measured for each population varied between -0.070 and -0.043; when pooling site counts from all populations, DoS = -0.060, $p = 0.00338$, Fisher's exact test, outgroup: FTH). Despite nuclear and mitochondrial genomes being exposed to different selective constraints, we found that relaxation of selection in annual killifishes is pervasive and affects both nuclear and mitochondrial genomes. Mutations in a number of genes affecting mitochondrial function, such as *polga*, are strong candidates for mitochondrial genomic instability and could be involved in mitochondrial genome expansion and higher mutation rate, possibly playing a key role in the age-related dysfunctions that characterize life history in annual killifishes. Relaxation of selection in the nucleus-encoded *polga* likely exacerbated mitochondrial genetic load.

DISCUSSION

Here, we provide evidence that genome-wide relaxation of selective constraints dominates evolution of short lifespan in African killifishes and more broadly affects the distribution of disease-causing alleles across species.

Transitions to annual environments impose intense selection on genes affecting early life history traits, such as embryonic development and sexual maturation, and yield to extensive relaxation of selective constraints on genes involved in late life homeostatic processes (Figure 7). Severe bottlenecks in annual species and in populations living in more arid environments further exacerbate the effect of relaxed selection by amplifying mutation load (Henn et al., 2015; Lynch and Gabriel, 1990). Although it cannot be ruled out that positively selected genes in *Nothobranchius* (Sahm et al., 2016; Valenzano et al., 2015) (Mendeley Data) may still in part contribute to aging-related phenotypes via antagonistic pleiotropic mechanisms (Charlesworth, 2000; Williams, 1957), the proportion of genes under positive

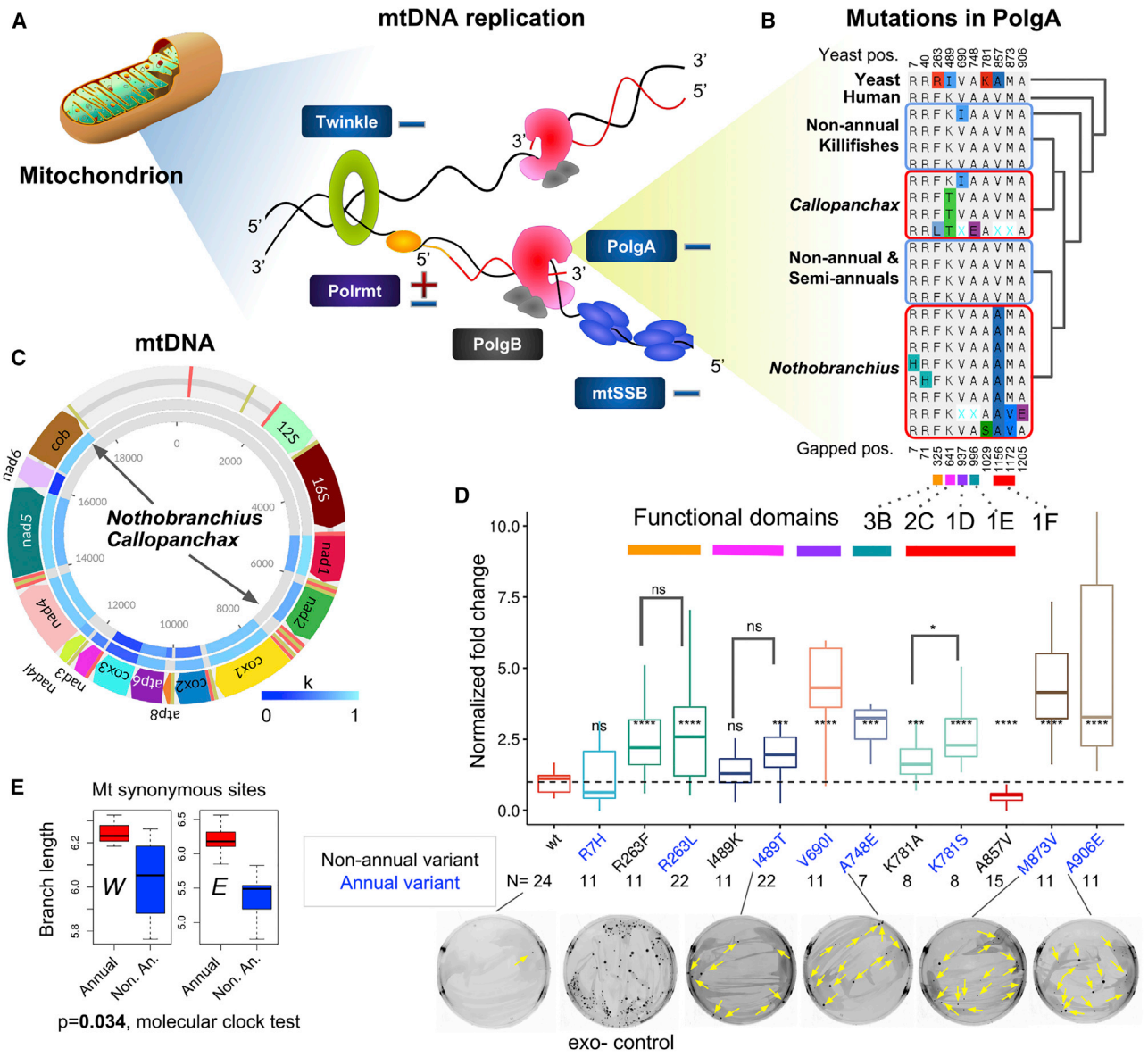


Figure 6. Relaxed Selection in *polga* Leads to Elevated mtDNA Mutation Rate

Figure360▷ For a Figure360 author presentation of Figure 6, see https://doi.org/10.1016/j.cell.2019.06.004.

(A) All four enzymes involved in the replication fork of mtDNA are relaxed in annual species (blue, negative sign). Additionally, *polrmt* is also positively selected (positive sign).

(B) Multiple protein sequence alignment of divergent sites between yeast *polg*, human *polga*, and killifish *polga*. A schematic of the phylogenetic tree is shown on the right. Six of these amino acid sites are within annotated functional domains.

(C) Relaxed selection (darker blue shade represents more "relaxed" k) for mitochondrially encoded proteins in *Nothobranchius* and *Callopanchax*. Inner circles, significant tests ($p < 0.05$, RELAX test).

(D) mtDNA replicative error rates measured for different annual killifish *polga* mutations in a yeast erythromycin assay. The normalized fold change of colonies is positively correlated with mtDNA replication errors. Representative erythromycin plates are shown for the different *polg* mutants (a yellow arrow indicates a colony). The exo-positive control has a very high mutation rate due to a lack of the exonuclease domain. If non-annuals share the same amino acid with yeast, the annual variant was directly compared to the yeast wild type (WT), otherwise, they were compared to the non-annual variants. Unpaired Wilcoxon rank-sum test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

(E) Faster mitochondrial synonymous substitution in the annual species indicated by longer branch-length in mitochondrial genomes (see Figures S7C–S7F for examples of mtDNA expansion).

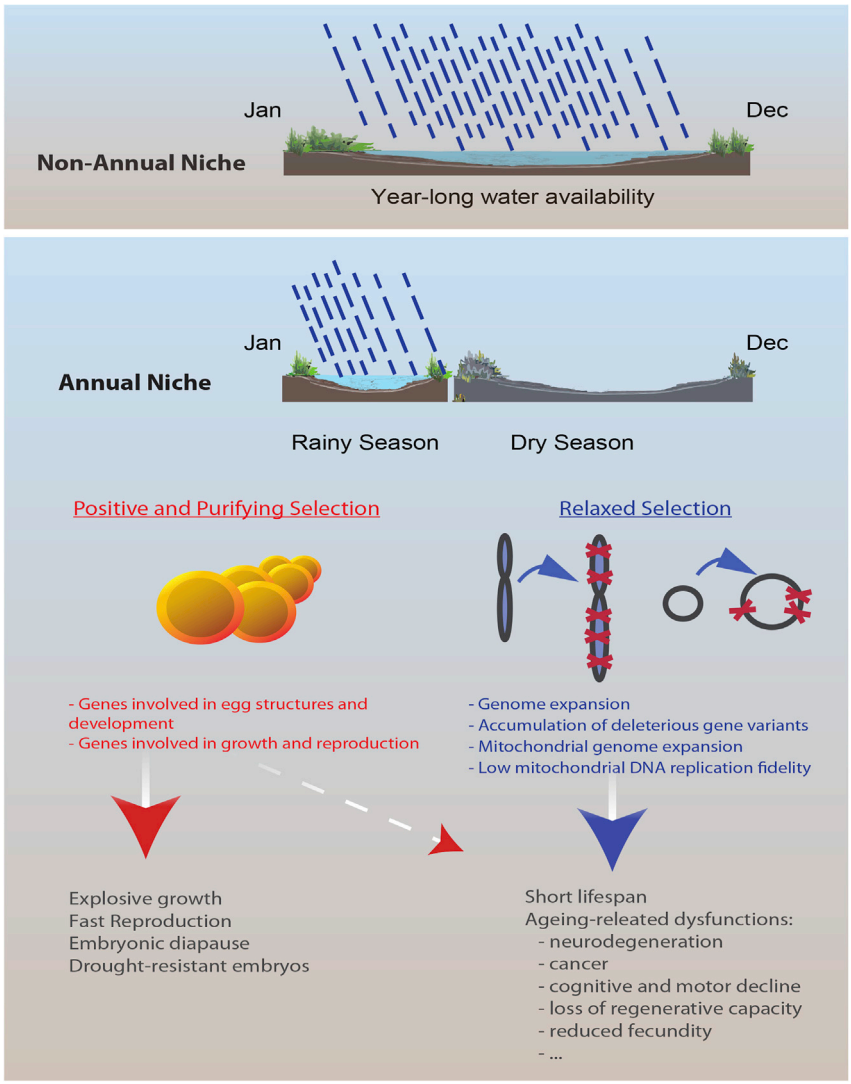


Figure 7. Schematic Representation of Main Findings and General Model

Evolutionary transitions to annual environment led to population bottlenecks and an increased chance of death due to extrinsic force of selection. On the genomic level, positive selection occurred on genes involved in growth and embryonic development, while relaxation of purifying selection resulted in expansion of transposable elements and accumulation of deleterious mutations preferentially in genes maintaining longevity. While positively selected genes may in part pleiotropically contribute to aging phenotypes (dashed red arrow), extensive relaxation of selection in genes involved in the vast majority of biological functions likely contributed to age-related dysfunctions in annual killifishes.

selection is dwarfed by those under relaxed selection. The extensive accumulation of deleterious mutations during the evolution in annual environments provides a mechanistic basis for the broad spectrum of aging phenotypes and the short lifespan that characterize annual killifishes (Figure 7; Valenzano et al., 2017).

Annual killifishes display increased mutational rate both in the nuclear and mitochondrial genomes, consistent with their size expansion. The nucleus-encoded mitochondrial DNA polymerase *polg* accumulated mutagenic variants in annual species, possibly leading to higher mitochondrial DNA mutational rate (Farnum et al., 2014). This mutational rate increase could have exacerbated mitochondrial dysfunction (Nurminen et al., 2017; Trifunovic et al., 2004). Indeed, relaxed selection on *polg* may have caused relaxation in all mitochondrially encoded genes that we observed in annual killifishes. It is conceivable that extensive relaxed selection for mitochondrial replication could subsequently generate strong positive selection for mitochondrial transcription and translation pathways to partially compensate for the dysfunctional genes. Despite between-species phenotypic differences that are believed to be largely driven by positive selection (Kern and Hahn, 2018), our results suggest that a significant proportion of positively selected genes may not exclusively contribute to species-specific adaptations but could help repair the aftermath of relaxed selective constraints.

Relaxed selection on aging-related genes is unlikely a peculiar feature of annual killifishes, but rather a general phenomenon occurring during transitions of life-history strategies and changes of demography, as suggested in the case of hominins. Our findings support that the accumulation of deleterious gene variants affecting late-life phenotypes can be evolutionary and mechanistically disconnected from selection for early life. That is, rather than being driven mainly by positive selection in mutations benefiting early life at the expense of late-life fitness, relaxed selection alone contributes to the majority of deleterious mutations affecting lifespan. Lifespan is likely limited by the cumulative action of numerous late-acting deleterious genetic variants of small effects, compatible with the nearly neutral theory of evolution (Ohta, 1973). Segregation of deleterious mutations at low-to-high frequencies, as we observed in the chimpanzee-human comparison, are consistent with differences in lifespan across these recently diverged species, although the exact cause requires further research. Strikingly, the genes under relaxed selection are enriched for aging and mitochondrial pathways across human populations, supporting that relaxation of selection is directly associated with aging-related phenotypic dysfunction, and the interpretation of biomedical genomic data for disease-associated variants can be in part explained by the recent evolutionary past of our species.

Our findings connect genomic relaxation of selection with the evolution of life history traits and help understand the origin and diffusion of disease-causing genetic variants within and among species.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Killifish samples
  - Wild *Nothobranchius* populations
  - Human and Chimpanzee data
  - Yeast strains and plasmids
- METHOD DETAILS
  - Library construction and sequencing
  - Yeast erythromycin assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Genome Assembly
  - Inferring genome size and gene family expansion
  - Phylogenetic reconstruction and climatic variable correlations
  - Bioclim variables and PGLS
  - Test for relaxation of selection and positive selection
  - Convergent AA sites
  - Consurf analysis
  - Codon bias
  - Reanalysis of a longitudinal RNaseq dataset
  - Classification of single-copy and multi-copy genes
  - Population resequencing
  - Identification of candidate sites under relaxation
  - Recombination rate estimates
  - Mitochondrial sequence assembly and tests
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.06.004.

A video abstract is available at https://doi.org/10.1016/j.cell.2019.06.004#mmc7.

ACKNOWLEDGMENTS

We thank Valenzano lab members for their support. Fish specimen contributors include Wolfgang Eberl, Winfried Grell (German Killifish Association), Melanie Stiassny, Barbara Brown, Lowell Flanders (AMNH), Stefano Valdesalici (Italian Killifish Association), Louis Hersh (University of Kentucky), and Jurij Phunkner. We thank Dmitri Petrov, David Enard, Hernan Burbano, Anne Brunet, James Stewart, and Clemens Weiss for critical reading of the manuscript. Molly Schumer provided suggestions on a simulation. The Cologne Center of Genomics (CCG) and NGX-Bio performed sequencing. We thank the IT department and the Bioinformatic Core of the MPI-AGE for their continuous support and input. Computations were run on Amalia HPC. Fish photos were provided by Anthony Terceira. The work was funded by the Max Planck Society and Max Planck Institute for Biology of Ageing (Germany). The Czech Science Foundation (Czech Republic, grant 19-1781S) contributed to part of the field work.

AUTHOR CONTRIBUTIONS

R.C. and D.R.V. conceived the study and designed the research. D.R.V. is the senior author. M.R. provided the fish tissue for population genetics. G.E.C. provided the fish tissue for the phylogenetic analysis. R.C. generated libraries for population re-sequencing, assembled and annotated the genomes, and ran all the simulations, comparative, and population genomics analyses. L.N.M.I. helped R.C. perform simulations. D.W. ran the ANGSD and GOSlim-DoS analyses. T.M. performed the erythromycin assay, supervised by M.G. R.C. and D.R.V. generated figures and wrote the manuscript. All authors read and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 22, 2018
Revised: March 18, 2019
Accepted: June 3, 2019
Published: June 27, 2019

REFERENCES

Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7-13.

Altschul, D.M., Hopkins, W.D., Herrelko, E.S., Inoue-Murayama, M., Matsuzawa, T., King, J.E., Ross, S.R., and Weiss, A. (2018). Personality links with lifespan in chimpanzees. eLife *7*, e33781.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., et al. (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science *297*, 1301–1310.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. *44* (W1), W344-50.

Auton, A., and McVean, G. (2007). Recombination rate estimation in the presence of hotspots. Genome Res. *17*, 1219–1227.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Ballesteros, J.A., and Hormiga, G. (2016). A new orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. Mol. Biol. Evol. *33*, 2117–2134.

Barton, H.J., and Zeng, K. (2018). New methods for inferring the distribution of fitness effects for INDELs and SNPs. Mol. Biol. Evol. *35*, 1536–1546.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Ar-Ganiska, J., Bely, B., Bingley, M., et al.; UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. *43*, D204–D212.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P.F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. *69*, 313–319.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. *14*, 988–995.

Blažek, R., Polačik, M., Kačer, P., Cellerino, A., Řežucha, R., Methling, C., Tomášek, O., Syslová, K., Terzibasi Tozzini, E., Albrecht, T., et al. (2017). Repeated intraspecific divergence in life span and aging of African annual fishes along an aridity gradient. Evolution *71*, 386–402.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Branzei, D., and Foiani, M. (2008). Regulation of DNA repair throughout the cell cycle. Nat. Rev. Mol. Cell Biol. *9*, 297–308.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., et al. (2014). The genomic substrate for adaptive radiation in African cichlid fish. Nature *513*, 375–381.

Browning, B.L., and Browning, S.R. (2016). Genotype imputation with millions of reference samples. Am. J. Hum. Genet. *98*, 116–126.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. *18*, 188–196.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

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

Charlesworth, B. (2000). Fisher, Medawar, Hamilton and the evolution of aging. Genetics *156*, 927–931.

Charlesworth, B. (2009). Evolution in age-structured populations, Second Edition (Cambridge University Press).

Chen, G.K., Marjoram, P., and Wall, J.D. (2009). Fast and flexible simulation of DNA sequence data. Genome Res. *19*, 136–142.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahloun, A.G., Herman, M.M., Weinberger, D.R., and Kleinman, J.E. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature *478*, 519–523.

Daley, T., and Smith, A.D. (2013). Predicting the molecular complexity of sequencing libraries. Nat. Methods *10*, 325–327.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271.

de Manuel, M., Kuhlwilm, M., Frandsen, P., Sousa, V.C., Desai, T., Prado-Martinez, J., Hernandez-Rodriguez, J., Dupanloup, I., Lao, O., Hallast, P., et al. (2016). Chimpanzee genomic diversity reveals ancient admixture with bonobos. Science *354*, 477–481.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. *45*, e18.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doudican, N.A., Song, B., Shadel, G.S., and Doetsch, P.W. (2005). Oxidative DNA damage causes mitochondrial genomic instability in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *25*, 5196–5204.

Eddy, S.R. (2008). A probabilistic model of local sequence alignment that simplifies statistical significance estimation. PLoS Comput. Biol. *4*, e1000069.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat. Rev. Genet. *8*, 610–618.

Farnum, G.A., Nurminen, A., and Kaguni, L.S. (2014). Mapping 136 pathogenic mutations into functional modules in human DNA polymerase γ establishes predictive genotype-phenotype correlations for the complete spectrum of POLG syndromes. Biochim. Biophys. Acta *1837*, 1113–1121.

Fawal, N., Savelli, B., Dunand, C., and Mathé, C. (2012). GECA: a fast tool for gene evolution and conservation analysis in eukaryotic protein families. Bioinformatics *28*, 1398–1399.

Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. *37*, 4302–4315.

Finch, C.E. (1994). Longevity, senescence, and the genome (University of Chicago Press).

Freckleton, R.P., Harvey, P.H., and Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat. *160*, 712–726.

Gordon, D., Huddleston, J., Chaisson, M.J.P., Hill, C.M., Kronenberg, Z.N., Munson, K.M., Malig, M., Raja, A., Fiddes, I., Hillier, L.W., et al. (2016). Long-read sequence assembly of the gorilla genome. Science *352*, aae0344.

Gorman, G.S., Chinnery, P.F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D.R., Zeviani, M., and Turnbull, D.M. (2016). Mitochondrial diseases. Nat. Rev. Dis. Primers *2*, 16080.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

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.

Han, M.V., Thomas, G.W.C., Lugo-Martinez, J., and Hahn, M.W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. *30*, 1987–1997.

Henn, B.M., Botigué, L.R., Bustamante, C.D., Clark, A.G., and Gravel, S. (2015). Estimating the mutation load in human genomes. Nat. Rev. Genet. *16*, 333–343.

Henn, B.M., Botigué, L.R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B.K., Martin, A.R., Musharoff, S., Cann, H., Snyder, M.P., et al. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. USA *113*, E440–E449.

Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2007). Context dependence, ancestral misidentification, and spurious signatures of natural selection. Mol. Biol. Evol. *24*, 1792–1800.

Hershberg, R., and Petrov, D.A. (2008). Selection on codon bias. Annu. Rev. Genet. *42*, 287–299.

Herwig, R., Hardt, C., Lienhard, M., and Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. Nat. Protoc. *11*, 1889–1907.

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, L., Wang, B., Chen, R., Bercovici, S., and Batzoglou, S. (2016). Reveel: large-scale population genotyping using low-coverage sequencing data. Bioinformatics *32*, 1686–1696.

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., and Otto, T.D. (2013). REAPR: a universal tool for genome assembly evaluation. Genome Biol. *14*, R47.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Johnson, S.C., Rabinovitch, P.S., and Kaeberlein, M. (2013). mTOR is a key modulator of ageing and age-related disease. Nature *493*, 338–345.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., et al. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature *447*, 714–719.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kern, A.D., and Hahn, M.W. (2018). The neutral theory in light of natural selection. Mol. Biol. Evol. *35*, 1366–1371.

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature *217*, 624–626.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.

Kriticos, D.J., Jarošik, V., and Ota, N. (2014). Extending the suite of bioclim variables: a proposed registry system and case study using principal components analysis. Methods Ecol. Evol. *5*, 956–960.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. *44* (D1), D862–D868.

Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.

Lawrie, D.S., Messer, P.W., Hershberg, R., and Petrov, D.A. (2013). Strong purifying selection at synonymous sites in *D. melanogaster*. PLoS Genet. *9*, e1003527.

Lefébure, T., Morvan, C., Malard, F., François, C., Konecny-Dupré, L., Guéguen, L., Weiss-Gayet, M., Seguin-Orlando, A., Ermini, L., Sarkissian, C., et al. (2017). Less effective selection leads to larger genomes. Genome Res. *27*, 1016–1028.

Leggett, R.M., Clavijo, B.J., Clissold, L., Clark, M.D., and Caccamo, M. (2014). NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. Bioinformatics *30*, 566–568.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lodi, T., Dallabona, C., Nolli, C., Goffrini, P., Donnini, C., and Baruffini, E. (2015). DNA polymerase γ and disease: what we have learned from yeast. Front. Genet. *6*, 106.

Lohmueller, K.E., Indap, A.R., Schmidt, S., Boyko, A.R., Hernandez, R.D., Hubisz, M.J., Sninsky, J.J., White, T.J., Sunyaev, S.R., Nielsen, R., et al. (2008). Proportionally more deleterious genetic variation in European than in African populations. Nature *451*, 994–997.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lynch, M., and Gabriel, W. (1990). Mutation load and the survival of small populations. Evolution *44*, 1725–1737.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., and Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. *2*, 1940–1955.

Maruki, T., and Lynch, M. (2015). Genotype-frequency estimation from high-throughput sequencing data. Genetics *201*, 473–486.

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.

McVean, G.A.T., and Charlesworth, B. (2000). The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics *155*, 929–944.

Medeiros, T.C., Thomas, R.L., Ghillebert, R., and Graef, M. (2018). Autophagy balances mtDNA synthesis and degradation by DNA polymerase POLG during starvation. J. Cell Biol. *217*, 1601–1611.

Mendes, F.K., and Hahn, M.W. (2016). Gene tree discordance causes apparent substitution rate variation. Syst. Biol. *65*, 711–721.

Messer, P.W., and Petrov, D.A. (2012). The McDonald-Kreitman test and its extensions under frequent adaptation: problems and solutions. arXiv, arXiv:1211.0060.

Messer, P.W., and Petrov, D.A. (2013). Frequent adaptation and the McDonald-Kreitman test. Proc. Natl. Acad. Sci. USA *110*, 8615–8620.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. *45* (D1), D183–D189.

Montinaro, F., Busby, G.B.J., Pascali, V.L., Myers, S., Hellenthal, G., and Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. Nat. Commun. *6*, 6596.

Murphy, W.J., and Collier, G.E. (1997). A molecular phylogeny for aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism. Mol. Biol. Evol. *14*, 790–799.

Murphy, W.J., and Collier, G.E. (1999). Phylogenetic relationships of African killifishes in the genera *Aphyosemion* and *Fundulopanchax* inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. *11*, 351–360.

Murphy, W.J., Nguyen, T.N., Taylor, E.B., and Collier, G.E. (1999). Mitochondrial DNA phylogeny of West African aplocheiloid killifishes (Cyprinodontiformes, Aplocheilidae). Mol. Phylogenet. Evol. *11*, 343–350.

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

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P.C., Friedman, M., and Smith, W.L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. USA *109*, 13698–13703.

Nurminen, A., Farnum, G.A., and Kaguni, L.S. (2017). Pathogenicity in POLG syndromes: DNA polymerase gamma pathogenicity prediction server and database. BBA Clin. *7*, 147–156.

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. Nature *246*, 96–98.

Paradis, E., and Schliep, K.J.B. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.

Partridge, L., and Barton, N.H. (1993). Optimality, mutation and the evolution of ageing. Nature *362*, 305–311.

Paten, B., Earl, D., Nguyen, N., Diekhans, M., Zerbino, D., and Haussler, D. (2011). Cactus: Algorithms for genome multiple sequence alignment. Genome Res. *21*, 1512–1528.

Pond, S.L.K., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676–679.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Great ape genetic diversity and population history. Nature *499*, 471–475.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.

Pryszcz, L.P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. *44*, e113.

Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., and Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics *18* (*Suppl 1*), S71–S77.

Raj, A., Stephens, M., and Pritchard, J.K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics *197*, 573–589.

Reichwald, K., Petzold, A., Koch, P., Downie, B.R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., et al. (2015). Insights into sex

chromosome evolution and aging from the genome of a short-lived fish. Cell *163*, 1527–1538.

Reid, N.M., Proestou, D.A., Clark, B.W., Warren, W.C., Colbourne, J.K., Shaw, J.R., Karchner, S.I., Hahn, M.E., Nacci, D., Oleksiak, M.F., et al. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science *354*, 1305–1308.

Revell, L. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.

Rey, C., Guéguen, L., Sémon, M., and Boussau, B. (2018). Accurate detection of convergent amino-acid evolution with PCOC. Mol. Biol. Evol. *35*, 2296–2306.

Rowan, B.A., Patel, V., Weigel, D., and Schneeberger, K. (2015). Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. G3 (Bethesda) *5*, 385–398.

Sahlin, K., Vezzi, F., Nystedt, B., Lundeberg, J., and Arvestad, L. (2014). BESST–efficient scaffolding of large fragmented assemblies. BMC Bioinformatics *15*, 281.

Sahm, A., Platzer, M., and Cellerino, A. (2016). Outgroups and positive selection: the *Nothobranchius furzeri* case. Trends Genet. *32*, 523–525.

Šajina, A., and Valenzano, D.R. (2016). An *in silico* model to simulate the evolution of biological aging. bioRxiv. https://doi.org/10.1101/037952.

Schartl, M., Walter, R.B., Shen, Y., Garcia, T., Catchen, J., Amores, A., Braasch, I., Chalopin, D., Volff, J.N., Lesch, K.P., et al. (2013). The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. Nat. Genet. *45*, 567–572.

Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925.

Sharp, P.M., and Li, W.H. (1987). The codon Adaptation Index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. *15*, 1281–1295.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.

Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–1247.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

Simons, Y.B., Turchin, M.C., Pritchard, J.K., and Sella, G. (2014). The deleterious mutation load is insensitive to recent population history. Nat. Genet. *46*, 220–224.

Simpson, J.T., and Durbin, R. (2012). Efficient de novo assembly of large genomes using compressed data structures. Genome Res. *22*, 549–556.

Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.

Smith, P., Willemsen, D., Popkes, M., Metge, F., Gandiwa, E., Reichard, M., and Valenzano, D.R. (2017). Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. eLife *6*, e27014.

Staab, P.R., Zhu, S., Metzler, D., and Lunter, G. (2015). scrm: efficiently simulating long sequences using the approximated coalescent with recombination. Bioinformatics *31*, 1680–1682.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. *33*, W465-7.

Stoletzki, N., and Eyre-Walker, A. (2011). Estimation of the neutrality index. Mol. Biol. Evol. *28*, 63–70.

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.

Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J.C., Schnable, P.S., Lyons, E., and Lu, J. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol. *16*, 3.

Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. *49*, 303–309.

Thomas, B.J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. Cell *56*, 619–630.

Timmermann, A., and Friedrich, T. (2016). Late Pleistocene climate drivers of early human migration. Nature *538*, 92–95.

Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J.N., Rovio, A.T., Bruder, C.E., Bohlooly-Y, M., Gidlöf, S., Oldfors, A., Wibom, R., et al. (2004). Premature ageing in mice expressing defective mitochondrial DNA polymerase. Nature *429*, 417–423.

Valenzano, D.R., Benayoun, B.A., Singh, P.P., Zhang, E., Etter, P.D., Hu, C.K., Clément-Ziza, M., Willemsen, D., Cui, R., Harel, I., et al. (2015). The African turquoise killifish genome provides insights into evolution and genetic architecture of lifespan. Cell *163*, 1539–1554.

Valenzano, D.R., Aboobaker, A., Seluanov, A., and Gorbunova, V. (2017). Non-canonical aging model systems and why we need them. EMBO J. *36*, 959–963.

Venkat, A., Hahn, M.W., and Thornton, J.W. (2018). Multinucleotide mutations cause false inferences of lineage-specific positive selection. Nat. Ecol. Evol. *2*, 1280–1288.

Vrtílek, M., Žák, J., Pšenička, M., and Reichard, M. (2018). Extremely rapid maturation of a wild African annual fish. Curr. Biol. *28*, R822–R824.

Wagner, J.T., Singh, P.P., Romney, A.L., Riggs, C.L., Minx, P., Woll, S.C., Roush, J., Warren, W.C., Brunet, A., and Podrabsky, J.E. (2018). The genome of *Austrofundulus limnaeus* offers insights into extreme vertebrate stress tolerance and embryonic development. BMC Genomics *19*, 155.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.

Warren, W.C., García-Pérez, R., Xu, S., Lampert, K.P., Chalopin, D., Stöck, M., Loewe, L., Lu, Y., Kuderna, L., Minx, P., et al. (2018). Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. Nat. Ecol. Evol. *2*, 669–679.

Weir, B.S., and Cockerham, C.C. (1984). ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION STRUCTURE. Evolution *38*, 1358–1370.

Wences, A.H., and Schatz, M.C. (2015). Metassembler: merging and optimizing de novo genome assemblies. Genome Biol. *16*, 207.

Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. Mol. Biol. Evol. *32*, 820–832.

Williams, G.C. (1957). Pleiotropy, natural selection, and the evolution of senescence. Evolution *11*, 398–411.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings of the VI International Congress of Genetics *1*, 356–366.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. Bioinformatics *29*, 2669–2677.

Please cite this article in press as: Cui et al., Relaxed Selection Limits Lifespan by Increasing Mutation Load, Cell (2019), https://doi.org/10.1016/j.cell.2019.06.004

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Saccharomyces cerevisiae w303 *ade2-1; leu2-3; his3-11, 15; trp1-1; ura3-1; can1-100 MAT a*	Thomas and Rothstein, 1989	YMG1
Saccharomyces cerevisiae w303 *ade2-1; leu2-3; his3-11, 15; trp1-1; ura3-1; can1-100 MAT α*	Thomas and Rothstein, 1989	YMG2
Biological Samples		
Fundulopanchax gardneri Udi mountain	Wolfgang Eberl (DKG)	N/A
Fundulopanchax amieti C89 / 31 Edea-Yahounde	Wolfgang Eberl (DKG)	N/A
Fundulopanchax scheeli	Glen Collier	N/A
Fundulopanchax deltaensis Koloware	Glen Collier	N/A
Fundulopanchax filamentosus Ikeja	Glen Collier	N/A
Fundulopanchax sjostedti Niger Delta	Jurij Phunkner	N/A
Aplocheilus lineatus Aquarium strain	Glen Collier	N/A
Pachypanchax playfairii SEY 05 La Digue	Wolfgang Eberl (DKG)	N/A
Epiplatys togolensis Palime	Glen Collier	N/A
Epiplatys grahami	Glen Collier	N/A
Epiplatys dageti Liberia	AMNH	N/A
Epiplatys lamottei Koule	Glen Collier	N/A
Epiplatys guineensis	Glen Collier	N/A
Epiplatys spilargyreius Pool Malebo	Glen Collier	N/A
Epiplatys multifasciatus, Democratic Republic of the Congo, Kinshasa	AMNH	AMNH 254532
Archiaphyosemion petersi Banco National Park	Glen Collier	N/A
Archiaphyosemion guineense, Guinea	AMNH	AMNH 260487
Archiaphyosemion leucupterygium Lola	Glen Collier	N/A
Scriptaphyosemion schmitti Juarzon	Glen Collier	N/A
Scriptaphyosemion guignardi Sougueta, Guinea	Glen Collier	N/A
Scriptaphyosemion bertholdi	Glen Collier	N/A
Scriptaphyosemion cauveti Kindia	Wolfgang Eberl (DKG)	N/A
Callopanchax sibideorum GM 97/2	Wolfgang Eberl (DKG)	N/A
Callopanchax huwaldi	Glen Collier	N/A
Callopanchax toddi GM 97 / 7 - Takhori	Wolfgang Eberl (DKG)	N/A
Callopanchax monroviae 97 / L	Glen Collier	N/A
Aphyosemion cyanostictum, Gabon	AMNH	AMNH 262946
Aphyosemion kouamense	Glen Collier	N/A
Aphyosemion australe, aquarium strain "gold"	Aquarium trade	N/A
Aphyosemion gabunense_marginatum	Wolfgang Eberl (DKG)	N/A
Aphyosemion cognatum, Congo basin	AMNH	AMNH 251785
Aphyosemion punctatum Bolo Bai	Louis Hersh (UKY)	N/A
Aphyosemion cameronense Cameroon	AMNH	N/A
Aphyosemion coeleste, Niari, Congo	AMNH	AMNH 258487
Aphyosemion kunzi	Wolfgang Eberl (DKG)	N/A
Aphyosemion kunzi BDBG 2004/10	Glen Collier	N/A
Fundulosoma thierryi Ada	Wolfgang Eberl (DKG)	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pronothobranchius seymouri Accra	Stefano Valdesalici (AIK)	N/A
Nothobranchius ocellatus Kikonkono	Glen Collier	N/A
Nothobranchius vosseleri TNKS 2013/35 Makaka	Wolfgang Eberl (DKG)	N/A
Nothobranchius foerschi	Glen Collier	N/A
Nothobranchius kafuensis Kayuni state farm	N/A	N/A
Nothobranchius virgatus Sudan	Stefano Valdesalici (AIK)	N/A
Nothobranchius rachovii Beira 98	Wolfgang Eberl (DKG)	N/A
Nothobranchius furzeri GRZ	Dario Valenzano	N/A
Nothobranchius orthonotus	Martin Reichard	N/A
Nothobranchius kuhntae wild samples	Martin Reichard	ORTWET
Nothobranchius orthonotus wild samples	Martin Reichard	ORTDRY
Nothobranchius rachovii wild samples	Martin Reichard	RACWET
Nothobranchius pienaari wild samples	Martin Reichard	RACDRY
Chemicals, Peptides, and Recombinant Proteins		
Erythromycin	Sigma-Aldrich	Cat# 114-07-8
Glycerol 99,5%	VWR	Cat# 56-81-5
Peptone from Meat	Merck	Cat# 91079-38-8
Yeast extract	MP-Biomedicals	Cat# 8013-01-2
Deposited Data		
Raw reads	NCBI	PRJNA531796
Assembled genomes	NCBI	SSNQ00000000, SSNS00000000, SSNR00000000, SSNP00000000
Genome annotations	Mendeley Data	https://doi.org/10.17632/f9phgbv4rs.1
Gene alignments	Mendeley Data	https://doi.org/10.17632/f9phgbv4rs.1
List of genes under relaxed / positive selection	Mendeley Data	https://doi.org/10.17632/f9phgbv4rs.1
Oligonucleotides		
PCR primers for yeast transformation	Mendeley Data	https://doi.org/10.17632/f9phgbv4rs.1
Recombinant DNA		
Plasmids for yeast transformation	Mendeley Data	https://doi.org/10.17632/f9phgbv4rs.1
Software and Algorithms		
Custom programs written for this study	This paper	https://github.com/valenzano-lab/killigenomics
Trimmomatic 0.32	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
Preseq 1.0.2	Daley and Smith, 2013	http://smithlabresearch.org/software/preseq/
NextClip 1.3.1	Leggett et al., 2014	https://github.com/richardmleggett/nextclip
Discovar *de novo* 52488	N/A	ftp://ftp.broadinstitute.org/pub/crd/DiscovarDeNovo/
NCBI BLAST 2.2.31+	Camacho et al., 2009	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
BESST (git version 7dba5 dd)	Sahlin et al., 2014	https://github.com/ksahlin/BESST
BESST_RNA (git version 9cc039b)	N/A	https://github.com/ksahlin/BESST_RNA
MasurCA 3.2.1	Zimin et al., 2013	https://github.com/alekseyzimin/masurca
Gapfiller 1.10	Nadalin et al., 2012	https://sourceforge.net/projects/gapfiller/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Reapr 1.0.18	Hunt et al., 2013	https://www.sanger.ac.uk/science/tools/reapr
Redundans 0.13a	Pryszcz and Gabaldón, 2016	https://github.com/lpryszcz/redundans
SGA preQC 0.10.13	Simpson and Durbin, 2012	https://github.com/jts/sga/wiki/preqc
BWA 0.6	Li and Durbin, 2010	N/A
BWA-MEM 0.7.12-r1039	Li, 2013	https://github.com/lh3/bwa
SMALT 0.5.3	Sanger Institute	https://www.sanger.ac.uk/science/tools/smalt-0
STAR 2.4.2a	Dobin et al., 2013	https://github.com/alexdobin/STAR
BUSCO v2beta4	Simão et al., 2015	https://busco.ezlab.org
Metassembler 1.3	Wences and Schatz, 2015	https://sourceforge.net/projects/metassembler/
PILON 1.22	Walker et al., 2014	https://github.com/broadinstitute/pilon
AllMaps 0.6.9	Tang et al., 2015	https://github.com/tanghaibao/jcvi/wiki/ALLMAPS
ProgressiveCactus (git version 7af8f26)	Paten et al., 2011	https://github.com/glennhickey/progressiveCactus
caper (R package, 1.0.1)	Freckleton et al., 2002	https://cran.r-project.org/web/packages/caper/index.html
RAxML 8.2.4	Stamatakis, 2006	https://cme.h-its.org/exelixis/software.html
TranslatorX 1.1	Abascal et al., 2010	https://translatorx.org/
MAFFT v7.305b	Katoh and Standley, 2013	https://mafft.cbrc.jp/alignment/software/
Muscle 3.8.31	Edgar, 2004	https://www.ebi.ac.uk/Tools/msa/muscle/
Gblocks 0.91b	Talavera and Castresana, 2007	http://molevol.cmima.csic.es/castresana/Gblocks.html
Consel 1.20	Shimodaira and Hasegawa, 2001	http://stat.sys.i.kyoto-u.ac.jp/prog/consel/
NOVOPlasty 2.6.3	Dierckxsens et al., 2017	https://github.com/ndierckx/NOVOPlasty
MITOS server	Bernt et al., 2013	http://mitos.bioinf.uni-leipzig.de/index.py
Trinity 2.1.1	Grabherr et al., 2011	https://github.com/trinityrnaseq/trinityrnaseq/wiki
Maker 3.0beta	Cantarel et al., 2008	http://yandell-lab.org/software/maker.html
Augustus 3.2.1	Stanke and Morgenstern, 2005	http://augustus.gobics.de/
CDHitEST 4.6.4	Li and Godzik, 2006	http://cd-hit.org/
Exonerate 2.2.0	Slater and Birney, 2005	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
RepeatModeler 1.73	Institute for Systems Biology	http://www.repeatmasker.org/RepeatModeler/
RepeatMasker 1.331	Institute for Systems Biology	http://www.repeatmasker.org/RMDownload.html
UPhO (git version 4ec1589)	Ballesteros and Hormiga, 2016	https://github.com/ballesterus/UPhO
MCL 14-137	Enright et al., 2002	https://micans.org/mcl/
Fasttree 2.1.9	Price et al., 2010	http://www.microbesonline.org/fasttree/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HMMER 2	Eddy, 2008	http://hmmer.org/
GeneWise 2.4.1	Birney et al., 2004	https://www.ebi.ac.uk/Tools/psa/genewise/
EvidenceModeler 1.1.1	Haas et al., 2008	https://evidencemodeler.github.io/
CAFE' 4.0.1	De Bie et al., 2006; Han et al., 2013	https://hahnlab.github.io/CAFE/download.html
TrimAl v1.4.rev15	Capella-Gutiérrez et al., 2009	http://trimal.cgenomics.org/downloads
CIWOG/GECA (git version ede81942)	Fawal et al., 2012	https://github.com/yasirs/CIWOG
GATK 3.4.46	McKenna et al., 2010	https://software.broadinstitute.org/gatk/download/archive
Samtools/Bcftools 1.2	Li et al., 2009	https://samtools.github.io/bcftools/bcftools.html
Picard v.1.119	Broad Institute	https://broadinstitute.github.io/picard/
APE 5.1	Paradis and Schliep, 2018	https://cran.r-project.org/web/packages/ape/index.html
PHYTOOLS 0.6.44	Revell, 2012	https://cran.r-project.org/web/packages/phytools/index.html
HyPhy v.2.2.5/v2.3.8	Pond et al., 2005	http://hyphy.org/download/
PAML 4.8	Yang, 2007	http://abacus.gene.ucl.ac.uk/software/paml.html
PCOC 07022018	Rey et al., 2018	https://github.com/CarineRey/pcoc
Consurf 1.0.6	Ashkenazy et al., 2016	https://github.com/Rostlab/ConSurf
Rate4Site 3.2	Pupko et al., 2002	https://m.tau.ac.il/~itaymay/cp/rate4site.html
Package-GFE (05/29/15)	Maruki and Lynch, 2015	https://github.com/Takahiro-Maruki/Package-GFE
MACs (f254a25)	Chen et al., 2009	https://github.com/gchen98/macs
scrm 1.7	Staab et al., 2015	https://github.com/scrm/scrm
fastStructure 1.0	Raj et al., 2014	https://rajanil.github.io/fastStructure/
PhyloBayes4.1	Lartillot et al., 2009	http://megasun.bch.umontreal.ca/People/lartillot/www/download.html
MSMC 2	Schiffels and Durbin, 2014; Li and Durbin, 2011	https://github.com/stschiff/msmc2
SMC++ (11e46fb)	Terhorst et al., 2017	https://github.com/popgenmethods/smcpp
Bcftools 1.6	Li et al., 2009	https://github.com/samtools/bcftools/releases/
LDHat (git version 81596e2)	Auton and McVean, 2007	https://github.com/auton1/LDhat
Anavar 1.4	Barton and Zeng, 2018	http://zeng-lab.group.shef.ac.uk/wordpress/?page_id=28
wgsim 0.3.1-r13	N/A	https://github.com/lh3/wgsim
featureCounts 1.5.0	Liao et al., 2014	http://bioinf.wehi.edu.au/featureCounts/
Deseq 2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Reveel release 2	Huang et al., 2016	http://reveel.stanford.edu/
Beagle 4.1	Browning and Browning, 2016	https://faculty.washington.edu/browning/beagle/b4_1.html
Angsd 0.918	Korneliussen et al., 2014	https://github.com/ANGSD/angsd
Uniprot Database	Bateman et al., 2015	https://www.uniprot.org/
World Climate 2 Database	Fick and Hijmans, 2017	http://worldclim.org/version2

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ConsensusPathDB version 33	Herwig et al., 2016	http://cpdb.molgen.mpg.de/
PANTHER Database 14.0	Mi et al., 2017	http://pantherdb.org/
Experimental Models: Organisms/Strains		
w303 Δ*mip1::URA3*	Martin Graef; Medeiros et al., 2018	YMG521
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1- natMX6*	Martin Graef; Medeiros et al., 2018	YMG532
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{exo-}natMX6*	Martin Graef; Medeiros et al., 2018	YMG533
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{R7H-}natMX6*	This paper	YMG651
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{V690I-}natMX6*	This paper	YMG653
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{A748E-}natMX6*	This paper	YMG802
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{k781S-}natMX6*	This paper	YMG803
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{A857V-}natMX6*	This paper	YMG804
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{M873V-}natMX6*	This paper	YMG805
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{A906E-}natMX6*	This paper	YMG654
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{R263F-}natMX6*	This paper	YMG806
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{I489K-}natMX6*	This paper	YMG807
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{K781A-}natMX6*	This paper	YMG808
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{R263L-}natMX6*	This paper	YMG652
W303 Δ*mip1::URA3* pRS315-*prMIP1-MIP1^{I48TF-}natMX6*	This paper	YMG658

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dario Riccardo Valenzano (dvalenzano@age.mpg.de). Some fish finclip samples used in the study were exhausted for DNA extraction and are no longer available.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Killifish samples

We sampled 43 African killifish species (two populations of *Aphyosemion kunzi* were included, 44 populations total) from Nothobranchiidae and 2 outgroup species from Aplocheiliidae. We chose 4 species for *de novo* genome assembly. These species were chosen to cover the 2 major phylogenetic clades of Nothobranchiidae and the outgroup. These 4 species are *N. orthonotus* (NOR, annual), *Aphyosemion australe* (AAU, non-annual), *Callopanchax toddi* (CTO, annual) and *Pachypanchax playfairii* (PLP, non-annual). Except for NOR, which is a recent-generation offspring from wild caught fish, highly inbred aquarium strains were used for the other three. Pheno-chloroform extraction or QIAGEN Dneasy Kit were used to extract gDNA from ethanol-preserved specimens.

Wild *Nothobranchius* populations

We chose two sympatric *Nothobranchius* species groups (*N. rachovii* group RAC, *N. orthonotus* group ORT), both adapted to dry and wet environments independently. For simplicity, we designate *N. pienaari* as RACDRY, *N. rachovii* as RACWET, *N. kuhntae* as ORTWET and *N. orthonotus* as ORTDRY. We sequenced from finclip DNA ~60 individuals of unknown sex (sex ratio approximately 1:1) from each population. We prepared the sequencing library with a published protocol (Rowan et al., 2015), with the modification that end-repair and A-tailing are combined into one reaction. Four individuals were sequenced to high coverage at > 25X, while others are at ~2.7X.

Human and Chimpanzee data

Called genotypes from the phase 3 1000G human genome project (Auton et al., 2015), chimpanzee resequencing data (de Manuel et al., 2016) and a Gorilla genome (Gordon et al., 2016) were downloaded from previous studies.

Yeast strains and plasmids

All *Saccharomyces cerevisiae* strains used in this study are W303 or derivatives thereof and are listed in Mendeley Data. All POLG/MIG1 alleles were generated by plasmid "gap repair"-PCR-based targeted recombination in Δmip1::URA3 and are listed in Mendeley Data. Primers used to generated specific point mutations are listed in Mendeley Data. Plasmid pRS315 was linearized prior to transformation using restriction endonucleases BamHI and HindIII and co-transformed with PCR fragments flanked by $\geq$ 40 bp homologous sequences to the 5′ and 3′ regions of the cut plasmid. To reintroduce mitochondrial genomes, strains were backcrossed, sporulated and confirmed for haploidy by analytic PCR and growth on selective media. Strains were grown in YPG medium [1% (w/v) yeast extract (Serva), 2% (w/v) bactopeptone (Merck), 3% (w/v) glycerol (Sigma)] or synthetic medium [0.7% (w/v) yeast nitrogen base (BD Difco); 2% (w/v) a-D-glucose] lacking specific amino acids for plasmid selection at 30°C.

METHOD DETAILS

Library construction and sequencing

For *de novo* assembly, 2 sets TruSeq PCR-Free libraries were made. The first set was not size-selected and sequenced to ~30X using 250x2PE reads on a Hiseq2500 Rapid mode. Preliminary analyses showed that the average insert size is only ~350bp. The second set of libraries were subjected to size selection by PippinPrep for an insert size of 450bp and sequenced to ~60X for each species using two 250X2PE runs. Mate-pair libraries were constructed with the Nextera mate-pair gel-plus protocol with minor modifications for three jump sizes: 3kb, 8kb and > 10kb. Four out of 12 libraries had a lower-than-expected insert size and/or poor complexity. These libraries were remade and sequenced with the other 8 libraries on two lanes of Hiseq 4000 (150bp x 2). This protocol was carried out by the Cologne Center of Genomics (CCG).

Depending on the available amount of gDNA, Truseq PCR-free or Truseq Nano protocols were used for the lower covered samples (5-30X) intended for reference-base variant calling. The maximal number of PCR cycles performed was 9. These libraries were sequenced using 100x2 PE on Hiseq 2500 and 150x2 Hiseq 4000.

Yeast erythromycin assay

Cells were grown in 3 mL YPG for 2 days at 30°C. Cells corresponding to 9 OD600-units were harvested by centrifugation and plated on YPG plates containing erythromycin (2 mg/ml; Sigma). To normalize for cell population, 0.01 OD600 were plated on YPG plates. Results were analyzed after 7 days at 30°C. All normalized counts from each batch were then divided by the averaged normalized count of the wild-type strain on erythromycin to control for batch effects.

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome Assembly

De novo assembly

We performed a series of library quality trimming, decontamination and deduplication (for mate-pairs). Discovar *de novo* was used to assemble contigs, and the resulting assemblies were subjected to heterozygous contig removal with Redudans, and repeated iterations of BESST scaffolding, gap-filling by GapFiller and misassembly splitting by Reapr using the mate-pair and pair-end libraries. We monitored the assembly continuity statistics and protein-encoding content quality at each iteration with BUSCO (runned with genome mode using the Teleost protein BUSCO library). We selected the assembly version with the peak statistics as the optimal assembly, and added an additional round of RNaseq scaffolding with BESST-RNA. In parallel, we also assembled the same data with MaSurCA, and repeated the same iterative process. The optimal DDN and the MaSurCA assemblies were reconciled into a meta-assembly by MetaAssembler, using the mapping statistics of the longest mate-pair library (> 10kb) as a basis to choose between local assembly regions. As a final step, we searched the assembly for closable gaps by comparing the flanking sequences around gaps, and merged them if they appeared to be nearly identical. Pilon was run using one of the 250X2PE libraries to correct for small assembly errors.

We further assigned the NOR scaffolds into linkage groups by mapping a previous RAD-Tag map from its sister species *N. furzeri*, and using orthologous proteins from the NFZ assembly as anchors. The two maps are reconciled with AllMaps. These species are closely related and have the same chromosome number. This LG-level assembly of NOR was only used in the psmc' analysis, in which we used bootstrapping to account for scaffolding errors.

Pseudogenome assembly generation

The lower-coverage resequencing data of 41 species were mapped to their closest related reference genome by BWA-mem (v.0.7.12) in pair-end mode. PCR duplicates were marked with "MarkDuplicates" tool in the Picard v.1.119 package. Reads were realigned around INDELs with the "IndelRealigner" tool in GATK v.3.4.46. We called variants with samtools v1.2 "mpileup" command, requiring a minimal mapping quality of 20 and a minimal base quality of 25. A "pseudogenome" assembly was generated by substituting reference bases with the alternative base in the reads. Uncovered regions, INDELs and sites with > 2 alleles were masked as unknown "N." The allele with more supporting reads was chosen at biallelic sites.

Gene model annotations

We extracted total RNA from whole bodies from the same ethanol-preserved individual used for *de novo* assembly. Then used the Trimmer-2 cDNA normalization kit (Evrogen) to prepare cDNA libraries. The four cDNA libraries turned into Illumina libraries by Nextera kit and sequenced on one HiSeq 4000 lane using 150bpx2. This library was mainly used for training Augustus.

Teleost protein sequences of 11 species were downloaded from Ensembl, and clustered into gene families by all-against-all blastp followed by MCL. The UPhO pipeline was used to further identify orthologs from these gene families. Each ortholog alignment we produced HMMER v.2 profiles, and aligned them to the *de novo* assemblies with Genewise.

We trained the Augustus predictor with a preliminary alignment of conspecific transcriptome data with Exonerate. We then ran Maker v.3.0beta, combining Genewise models and assembled Killifish transcriptome assemblies. Evidence modeler was run internally by Maker 3 for model reconciliation. High-quality protein models from the first round of maker models were picked by requiring complete alignment to NCBI proteins. These gene models were used to retrain Augustus. Maker was rerun with the retrained Augustus. Inspection showed that Maker often produces fused gene models between neighboring genes. We developed a pipeline to fix these artifacts, and improved the BUSCO quality score from 89.9%–92.2% in the raw Maker output to 95.66%–97.34% after the final fixing step.

Repeat masking

We ran repeat modeler on the DDN contigs to identify killfish-specific repeats. The identified repeats from 4 reference genomes were concatenated into a killifish repeat database, and merged with the repbase library. Repeatmasker was used to annotate transposable elements and other repeat types in the reference genomes.

Gene family determination and orthology assignment

We ran all-against-all blastp on predicted amino acid (AA) sequences from NOR, PLP, AAU, CTO gene models, combined with the longest AA isoforms of 11 fish species (Aparicio et al., 2002; Howe et al., 2013; Kasahara et al., 2007; Warren et al., 2018) in Ensembl (*Lepisosteus oculatus, Gadus morhua, Oreochromis niloticus, Oryzias latipes, Xiphophorus maculatus, Poecilia formosa, Takifugu rubripes, Tetraodon nigroviridis, Gasterosteus aculeatus, Danio rerio* and *Astyanax mexicanus*) and the NCBI annotation of NFZ (only the longest isoform per gene), totaling 373,018 sequences (on average 23,313 sequences/species). We clustered the results using MCL into 11847 gene families, to which 97.7% genes can be assigned. Among these, 5475 families have > = 1 gene per species, 1261 have exactly one gene per species and 8667 genes (2.3%) remained as singletons. The clustering results from MCL were used for gene family expansion/contraction analyses by CAFe'.

We further delineated orthologs from gene families with multiple paralogs using the UPhO pipeline, with a maximum likelihood tree independently built for each family by FastTree (2.1.9 SSE3, OpenMP). In total, we delineated 24563 ortholog groups for the 16 teleost species. Within these, we found that 13932 orthogroups contain exactly one gene for each of the 4 killifish species. We further combined incorrectly split orthogroups informed by Ensembl Compara database for the 11 teleost species. At the end, we can assert orthology of 14709 genes for African killifish, which were used in downstream analyses. Gene symbols and gene names are assigned from the ensembl fish genes located in the same ortholog group.

Inferring genome size and gene family expansion

Genome size

We mapped quality-trimmed reads to the single-copy BUSCO gene models and estimated the sequencing depth by bedtools. The genome size is estimated as total read bases / depth.

Gene family size analyses with CAFE'

We analyzed 11847 gene families for their expansion and contraction patterns for 16 teleost species, including the 4 newly sequenced killifishes, using the package CAFE' v2.2. We first used strictly single-copied genes (1261 genes, 601,174 AA, 4.21% missing, 16 taxa) to build a maximum likelihood tree with RAxML using 100 rapid bootstraps followed by a full ML optimization with the PROTGAMMAJTTX model. This tree has bootstrap values of 100 in all internal nodes, and recovers the expected phylogeny. This tree was only used to guide the CAFE' analyses.

We counted the number of paralogs per gene family by summing the ''completeness'' annotation and rounded to the nearest integer to account for split genes. CAFE' was run on the paralog counts with recommendations by the user manual, with an estimated error rate parameter. We tested whether gene birth and death rates significantly differ in any killifish compared to the other teleost by likelihood ratio tests and simulations outlined in the CAFE' user manual.

Expansion of TE elements in Nothobranchius

We used the GECA package to identify orthologous introns from the 14709 orthologous killifish genes. We tallied the TE types and their lengths in these introns, and contrasted the repeat lengths between AAU and NOR, both grouped by TE family or pooled. We determined the global expression levels of a gene by selecting genes with consistent expression levels in the RNaseq data among all 4 killifishes (std/mean < 0.5). We then compared the expression levels between the genes receiving 10% and 90% quantiles of TE expansions using a Wilcoxon rank sum test.

Phylogenetic reconstruction and climatic variable correlations

Cyprinodontiform phylogeny reconstruction

We extracted homologous bases in the coding regions by aligning raw reads of all killifish samples to a previous published NFZ genome (Valenzano et al., 2015) with the same approach as above. In addition, we aligned genome assemblies of *Xiphophorus maculatus*, *Austrofundulus limnaeus* and *Fundulus heteroclitus* to the NFZ reference with ProgressiveCactus, and converted the file to v*cf.* format. We then extracted protein-coding regions from the alignments based on the NFZ GFF annotations. The final alignment has 9,710,157 sites in 49 species, with 17.8% of missing data. We ran RAxML to infer the species tree, with 100 rapid bootstraps followed by a full ML optimization, using the GTRGAMMA model. This phylogeny was only used for character mapping and correlations using phylogenetic generalized linear models (PGLS). By calibrating the phylogeny with a dated tree of teleost (Near et al., 2012) we estimated a neutral mutation rate of 2.6321e-9 /bp/year in the *Nothobranchius* branch, similar to an experimentally obtained rate in cichlids (Malinsky et al., 2018).

Bioclim variables and PGLS

GPS coordinates of the collection locality (if available) or the type locality of each species were used to extract the bioclim variables within a 1km radius using the "raster" package in R. We used the phylogenetic generalized linear (PGLS) model implemented in the R package "caper" to perform correlations, correcting for phylogenetic dependency. The species phylogeny was specified as a covariate matrix assuming the Brownian diffusion model of trait evolution. We ran the model both with and without branch transformations. In case of branch transformations, we performed model selection to determine the optimal number of transformation parameters (kappa, lambda, delta) using AIC. A less complex model is preferred if it has a $\Delta AIC < 4$ than the more complex model.

Test for relaxation of selection and positive selection

Alignment quality control and gene tree inference

Using the ortholog alignments and the corresponding gene models, we extracted protein-coding gene squences, excluding unreliable parts of the gene model, and aligned them with TranslatorX 1.1, which internally calls MAFFT v7.305b for alignment and Gblocks 0.91b for exclusion of unreliable regions. The sequences for the species with abnormal read coverage ($p < 0.01$ of genome-wide coverage distribution) were excluded to avoid mismapped paralogs. We further excluded genes with significant deviation from the genus-level phylogeny (see 3.1) in an AU test (Shimodaira, 2002) to avoid false positives in the subsequent tests (Mendes and Hahn, 2016). At $p < 0.05$, 765 out of 14691 genes (0,052) were excluded, suggesting that an over-whelming support for the Cyprinodontiform phylogeny by gene trees. We used the RAxML constraint tree for the RELAX and CodeML tests.

RELAX: General descriptive model

We ran 13709 quality filtered alignments with the RELAX general descriptive model using corresponding gene trees. This model allows for a free relaxation parameter k to be inferred for every branch. The relaxation parameter k is an exponent for selection parameters between the foreground and the background branches, where $(omega_{background})^k = omega_{foreground}$. A $k > 1$ suggests selection to be more intensified in foreground branch and vice versa. We used the median k for each species for correlations with PGLS.

RELAX: hypothesis testing

To statistically test whether genes are under relaxed purifying selection, we used the likelihood ratio test in RELAX, comparing the model fixing $k = 1$ and the model allow k to be estimated. The MRCA branch and all branches within each genus were alternatively set as the foreground branches for the test, except for PLP and APL set as unclassified. We also performed an analysis where only the MRCA branch of *Nothobranchius* was included. Power and specificity simulations were performed to assess performance of these tests (Mendeley Data).

CodeML branch-site test for positive selection

We performed a likelihood ratio test between a model assuming no difference between foreground and background branches versus a model allowing for positive selection in the foreground. We assessed the performance with simulations (Mendeley Data).

Convergent AA sites

We used two approaches to screen for convergent AA sites in *Callopanchax* and *Nothobranchius*. Our simple approach requires an otherwise conserved site to exclusively share the same derived AA in the two genera. The PCOC approach is more relaxed and allows for similar AA's to be considered as convergent. We assessed the performance of these approaches with simulations (Mendeley Data).

Consurf analysis

The Consurf package was used to assign a conservation score to each derived AA in *Nothobranchius* by estimating the evolutionary rate of the site in the homologs in other vertebrates. Higher scores suggest the site to be more conserved and more likely to be functionally constraint.

Similarly, Consurf scores were calculated using the longest protein-coding transcript for NCBI annotations of the human reference genome version GRCH37.

Codon bias

We calculated the codon adaptive index (CAI) (Sharp and Li, 1987) by comparing the synonymous codon compositions to the top 1% highly expressed genes.

Reanalysis of a longitudinal RNaseq dataset

We downloaded previous RNaseq data from a longitudinal study of *Nothobranchius furzeri*, consisted of 5 time points (5 – 39 weeks) and 3 organs (liver, brain, skin) (Reichwald et al., 2015). The raw reads were mapped to the NOR reference genome, and DESeq 2 was used to analyze genes significantly correlated with age. Genes are classified as upregulated in young (log(FoldChange) < 0, $p < 0.05$), upregulated in old (log(FoldChange) > 0, $p < 0.05$) and no-change ($p >= 0.05$).

We downloaded the normalized microarray counts from a human prefrontal cortex longitudinal expression dataset (Colantuoni et al., 2011), and directly applied a linear model Exp_level ~log(age) to determine the direction of expression.

Classification of single-copy and multi-copy genes

Based on the UPhO gene clustering results in 1.7, we found the orthologous groups where the same outgroup species gene has been repeatedly grouped with two different NOR genes. These are duplicated genes in NOR compared to the outgroup. Single-copied genes are those where the outgroup gene only appeared once.

Population resequencing

FastStructure analysis

We exported genotypes to faststructure format by treating one of the allele as missing for sites with < 4 reads, thinning markers to every 5kb, requiring MAF > 0.05 and excluded sites higher than twice the genome-wide coverage and covered in less than 20 individuals. In total, 2,842,829 sites were included. Faststructure was run with k = 2-6, with default parameters. Four putative hybrids were excluded from subsequent analyses.

Inference of population size changes

We mapped the reads of each individual to the LG version of the NOR reference with BWA-mem. ANGSD was used to infer polymorphism π and Watterson's θ from the alignment. For PSMC' (implemented in MSMC2) analysis, we mapped the highly covered individuals with BWA-MEM, followed by variant calling with bamCaller.py with recommended parameters without phasing. MSMC2 was run with default parameters with 40 EM iterations. Simulation was used to validate the method using the estimated demography with the split time inferred by smc++. For smc++ split analysis, we realigned the reads with GATK around INDELs, and called variants jointly for each population using samtools v1.6. We converted the v*cf.* file into smc++ format, masking variants with a variant quality < 60, and regions where the reference is N. Smc++ was run on all chromosomes with the following parameters: ''–time points 1e3,1e7–knots 16–hs 2–spline pchip–regularization-penalty 4.'' Split time was estimated by smc++ split, with the parameters: ''–regularization-penalty 4–nonseg-cutoff 2000.'' We performed 30 block bootstraps by chopping the chromosomes into 5Mb chunks and resampled the chunks with replacement to the original size, and reran smc++ with the same settings. We used an estimated mutational rate μ = 2.6321e-9/bp/generation to convert into real timescales (Mendeley Data).

McDonald-Kreitman α and DoS estimates

We use Package-GFE to estimate allele frequencies jointly with sequencing errors. Our simulations show that it is accurate when the minor allele frequency > = 0.05 (Mendeley Data). We used a previous method to compute asymptotic alpha for the four populations across different allele frequency bins (Messer and Petrov, 2013), pooling all protein-coding sites in the genome. We computed gene-wise direction of selection (DoS) statistics for correlations. Anavar was used to estimate the distribution of fitness effects (DFE). For human populations, we used the allele frequencies estimated from the downloaded VCF files polarized by a Gorilla outgroup.

Allele frequency of deleterious mutations

We estimated Consurf scores for 5,816,790 amino acids (17,450,370 nucleotides) in 14,360 protein-coding genes for NOR and matched them with the estimated derived allele frequencies (polarized by FTH) in each population. We binned the Consurf scores by their derived allele frequencies and compared between the dry and wet populations using Wilcoxon rank sum test at each bin. Rare alleles (AF < 0.05) were excluded. We also analyzed a full model where allele frequency and population (dry/wet) were specified as fixed effects. To quantify per-individual genetic load, we randomly sampled one read per site and summed up the Consurf score per killifish individual, and normalized to the amount of missing data. We chose an allele frequency cutoff of 0.075 when computing the total load for the low coverage data because this is where sequencing error drops below 5% in simulations (Mendeley Data).

Hominid genotype calls were downloaded from previous studies (Auton et al., 2015; de Manuel et al., 2016), and polarized by a single Gorilla sequence remapped to the human genome version GRCH37 (Gordon et al., 2016). The calculated allele frequencies for the non-synonymous variants in each *Pan* or *Homo* population were matched to the estimated Consurf scores of 5,534,398 amino acids (16,603,194 nucleotides) in 12,970 protein-coding genes. Because multiple subpopulations are available, we used linear mixed effect models to compare at each frequency bins and for the full model. The subpopulations were treated as random effects and the continent (among different human populations) or the genus (between human and chimpanzees) were specified as the fixed effect. We summed the Consurf scores for each individual (multiplied by 2 for homozygous sites) to calculate the total per-individual load. The variant callability mask downloaded from the chimpanzee dataset (ftp://ftp.cnag.cat/Chimpanzee/hg19/) was applied to both *Pan* and *Homo*, so that the same regions in the genome were compared.

We also contrasted Europeans and East Asians with Africans using the same approach. Because more than 50 individuals were available for these populations, we included an extra frequency bin of 1%–5%.

Identification of candidate sites under relaxation

To identify candidate sites under relaxed selection in different human populations, we filtered for sites with derived allele frequency between 0.01 and 0.4, because this is the range showing a significant enrichment of deleterious mutations in the European (EUR) and East Asian (EAS) populations compared to Africans (AFR). We then checked if the gene-wise average consurf scores between EUR or EAS and AFR are significantly different (FDR = 20%). For the genes passing this filter, we checked codon by codon and report codons that have a consurf score > = 7, and have a higher allele frequency in the focal population.

Recombination rate estimates

We used Reveel to impute missing genotypes and used LDHat to infer recombination rates. Simulations were done to assess performance of imputed genotypes for recombination inference (Mendeley Data).

Mitochondrial sequence assembly and tests

Novoplasty was used to assemble mtDNA from raw Illumina reads. Among 46 samples, Novoplasty failed to produce a complete assembly for 3 species, which were excluded. Genes were annotated by the MITOS webserver. We used the likelihood ratio test in BaseML on synonymous sites to test deviation from a molecular clock. The RELAX and CodeML tests followed the same procedure with nuclear genes.

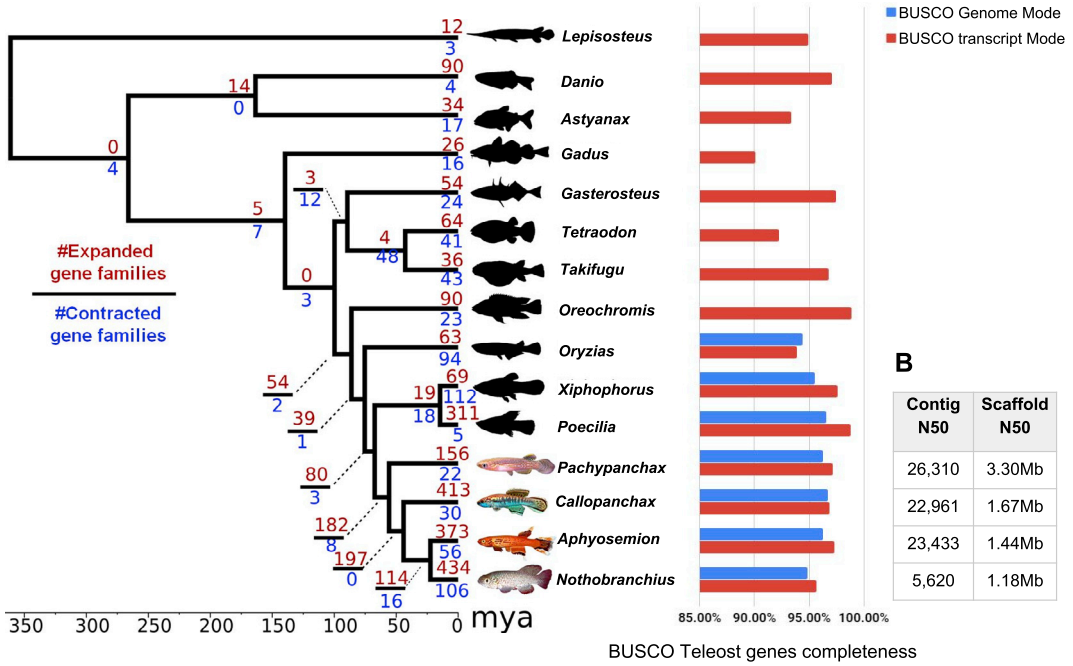
Assay for mutation frequency of mtDNA

Wilcoxon sign rank tests were used to compare the mutant strains with the wild-type, or to the strain carrying a long-lived fish mutation.

DATA AND CODE AVAILABILITY

The datasets and genome assemblies generated during the current study are available in the NCBI SRA repository (BioProject ID: PRJNA531796). Analysis and simulation results are available in Mendeley Data (https://doi.org/10.17632/f9phgbv4rs.1). All computer code is available from https://github.com/valenzano-lab/killigenomics.

A

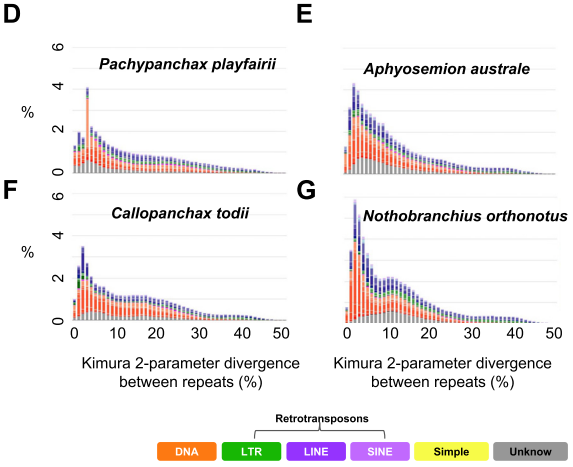


B

Contig N50	Scaffold N50
26,310	3.30Mb
22,961	1.67Mb
23,433	1.44Mb
5,620	1.18Mb

C

Branch	Gene birth rate λ	Gene death rate μ	Sum of rates $\lambda + \mu$	Net gene gain λ / μ	Branch-specific P value
Pachypanchax playfairii	0.00169	0.00086	0.00255	1.97	0.70
Callopanchax toddi	0.00397	0.00088	0.00485	**4.51**	**0.019**
Aphyosemion australe	0.00682	0.00285	0.00967	**2.39**	**0.017**
Nothobranchius orthonotus	**0.00970**	**0.00600**	**0.01570**	**1.61**	**0.014**
Background	0.00125	0.00116	0.00241	1.08	-



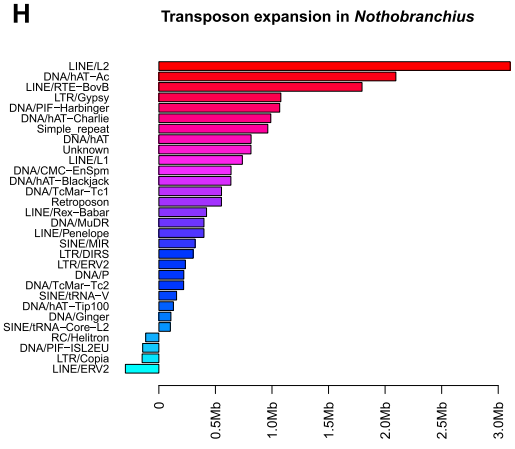
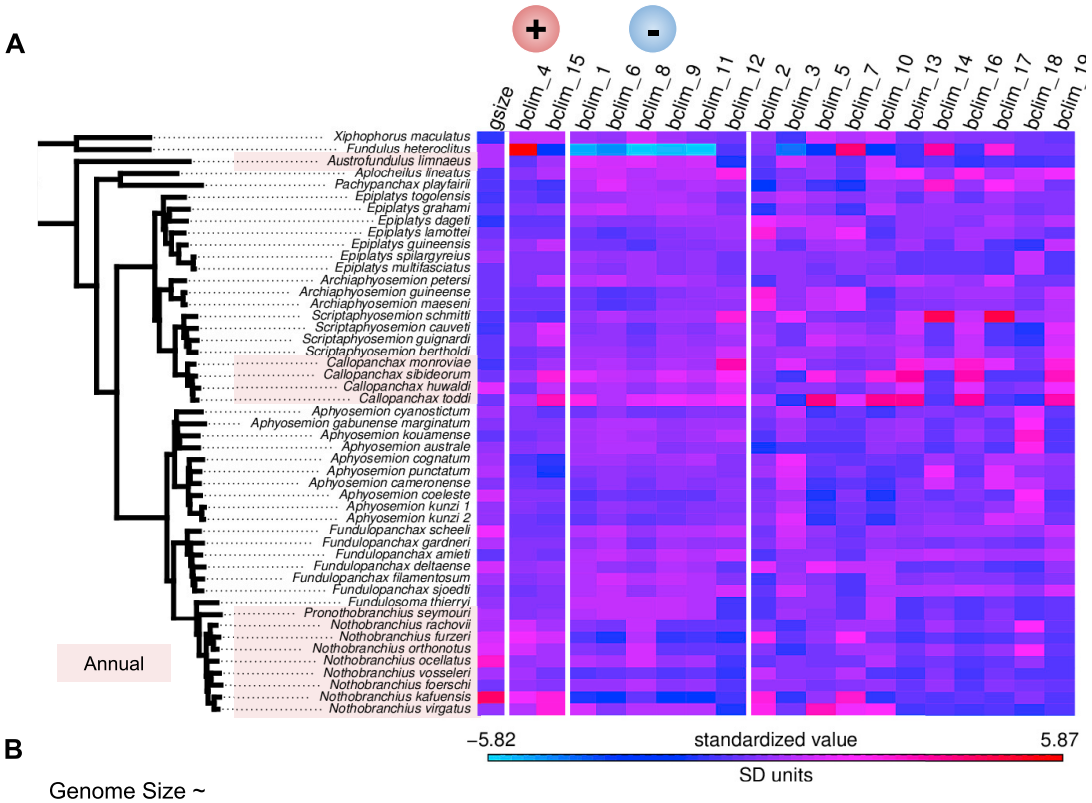


Figure S1. Related to Figure 1

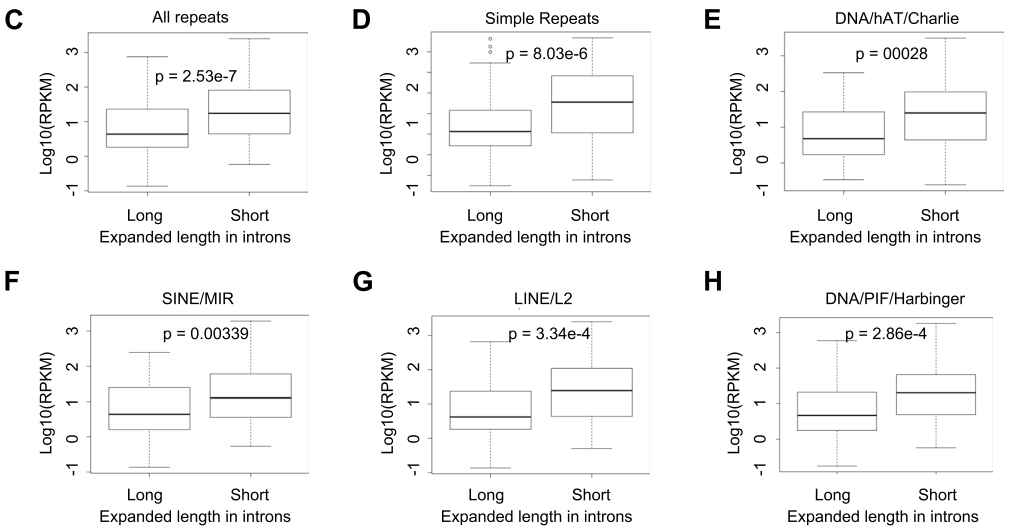
Genome and gene family evolution. (A) CAFe' gene family expansion/contraction analyses – number above line (red): expanded gene family, below line: contracted gene family. Barplot shows BUSCO statistics for Ensembl fish genomes and killfishes. (B) Assembly statistics for the *de novo* genomes (also see Mendeley Data). (C) Table shows gene gain/loss rate estimated for each branch. (D)-(G), Divergence of repeat elements in different *de novo* assemblies. (H) Expansion patterns of transposons within orthologous introns in *Nothobranchius* compared to *Aphyosemion*, estimated by repeatmasker.



B

Genome Size ~

BioClim	Meaning	PGLS w/o branch transformation			PGLS AIC branch transform		Raw linear regression	
		Slope	p	FDR	p	FDR	p	FDR
4	Temp. Seasonality	+	**0.0125**	**0.0297**	**0.0019**	**0.0349**	**0.0002**	**0.0015**
15	Precip. Seasonality	+	0.4583	0.4838	**0.0344**	0.0933	0.0582	0.0840
1	Annual Mean Temp.	-	**0.0001**	**0.0006**	**0.0394**	0.0936	**0.0112**	**0.0289**
6	Min Temp. of Coldest Month	-	**0.0002**	**0.0010**	**0.0062**	**0.0349**	**0.0024**	**0.0090**
8	Mean Temp. of Wettest Quarter	-	**0.0070**	**0.0190**	**0.0269**	0.0853	0.7050	0.7442
9	Mean Temp. of Driest Quarter	-	**0.0000**	**0.0000**	**0.0073**	**0.0349**	**0.0005**	**0.0031**
11	Mean Temp. of Coldest Quarter	-	**0.0000**	**0.0000**	**0.0040**	**0.0349**	**0.0001**	**0.0015**
12	Annual Precip.	-	0.2193	0.3169	**0.0202**	0.0768	**0.0008**	**0.0038**

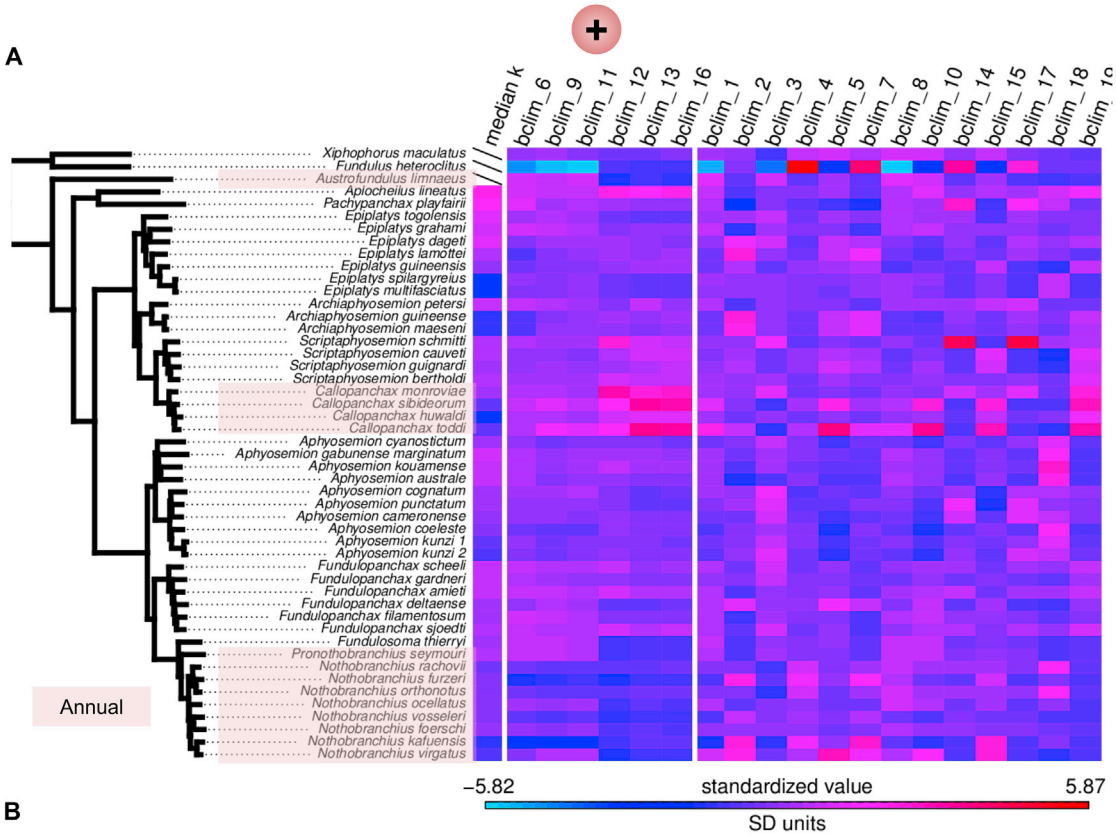


(legend on next page)

Figure S2. Related to Figure 1

Correlates with genome expansion. (A) Correlations between genome size and 15 bioclim variables. (B) Table shows correlation statistics from PGLS analyses (see also Mendeley Data). (C)-(H) Correlation between gene expression level and inserted repeat length in *Nothobranchius*. Edges of boxplots indicate the 1st and 3rd quartiles, whiskers are the data range excluding outliers.

A



B

Median RELAX k ~

BioClim	Meaning	PGLS w/o branch transformation			PGLS AIC branch transform		Raw linear regression	
		Slope	p	FDR	p	FDR	p	FDR
6	Min Temp. of Coldest Month	+	**0.0406**	0.1661	**0.0398**	0.1513	**0.0016**	**0.0148**
9	Mean Temp. Driest Quarter	+	**0.0312**	0.1661	**0.0155**	0.1513	**0.0289**	0.0698
11	Mean Temp. Coldest Quarter	+	0.0625	0.1696	**0.0389**	0.1513	**0.0294**	0.0698
12	Annual Precip.	+	**0.0239**	0.1661	**0.0250**	0.1513	0.1500	0.2375
13	Precip. Wettest Month	+	**0.0240**	0.1661	**0.0273**	0.1513	0.5520	0.6169
16	Precip. Wettest Quarter	+	**0.0437**	0.1661	0.0638	0.1994	0.5340	0.6169

Figure S3. Related to Figure 1

Correlates with relaxed selection k. (A) Correlations between gene-wise median k (estimated by the RELAX general descriptive model) and 15 bioclim variables (see also Mendeley Data). (B) Table shows correlation statistics from PGLS analyses.

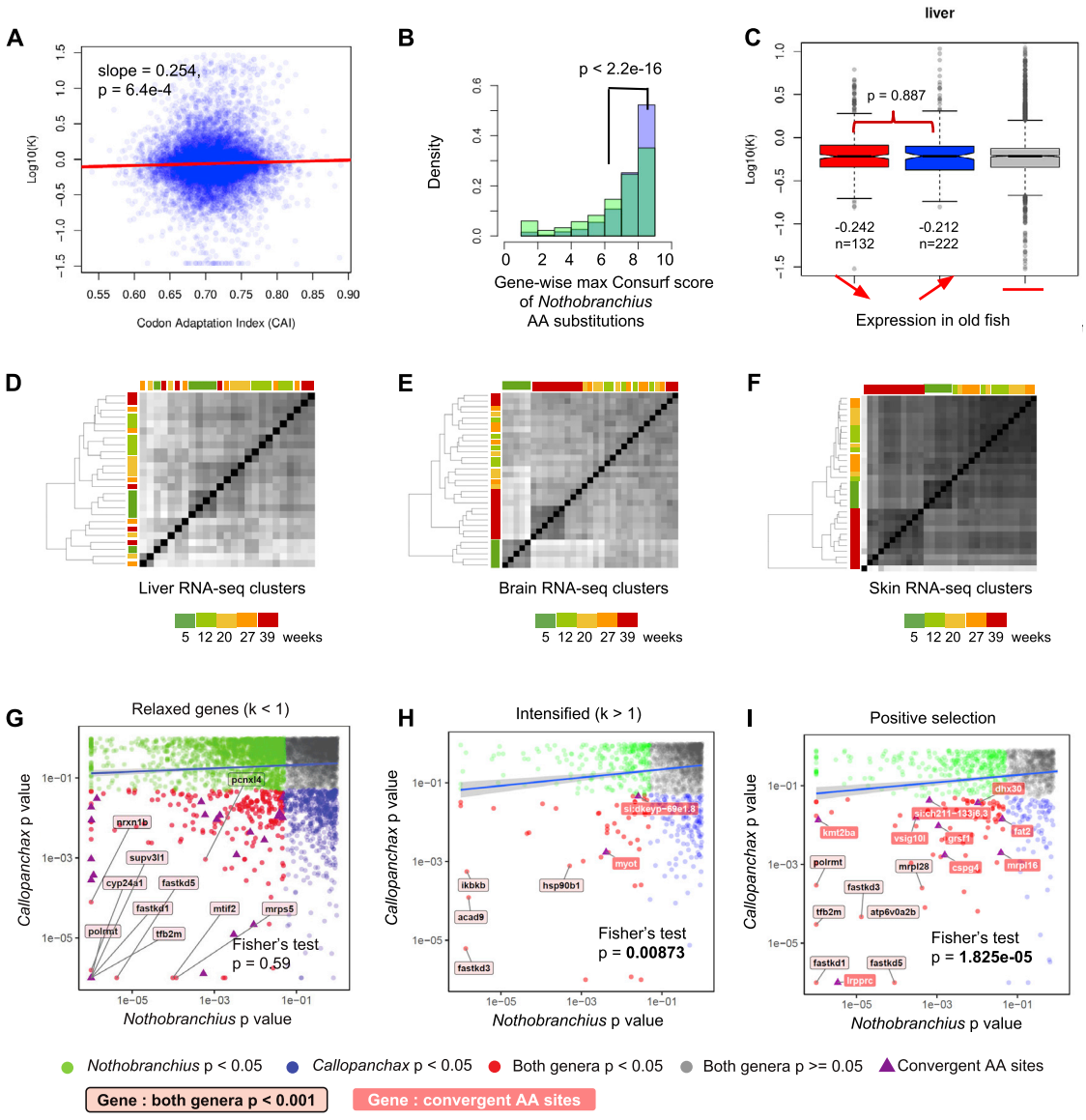


Figure S4. Related to Figure 2

Additional information on RELAX and CodeML tests. (A) The level of relaxation of selection in *Nothobranchius* is positively correlated with codon bias (measured by codon adaptive index CAI). (B) The maximal gene-wise Consurf scores in *Nothobranchius*-specific mutations, in relaxed (blue) and non-significant genes (green). Wilcoxon rank sum test of means. (C) Compared to Figures 2F and 2G, in liver the non-significant test is consistent with low power due to a small number of differentially expressed genes, based on 1000 downsampled pseudoreplicates of the skin (Prob(wilcox p > 0.05) = 0.529) and the brain (Prob(wilcox p > 0.05) = 0.594) datasets. Edges of boxplots indicate the 1st and 3rd quartiles, whiskers are the data range excluding outliers. (G) Numerous relaxed genes overlap between *Callopanchax* and *Nothobranchius*, but the overlap is not statistically significant. The overlapped relaxed genes are enriched for convergent amino acid substitutions (p = 0.0015, see Figures S5K–S5M for the analyses with CMDs excluded). (H) Intensified gene sets detected by RELAX significantly overlap between *Callopanchax* and *Nothobranchius* (but not so after CMD exclusion, Figure S4K). (I) Positively selected gene sets detected by CodeML in two annual genera.

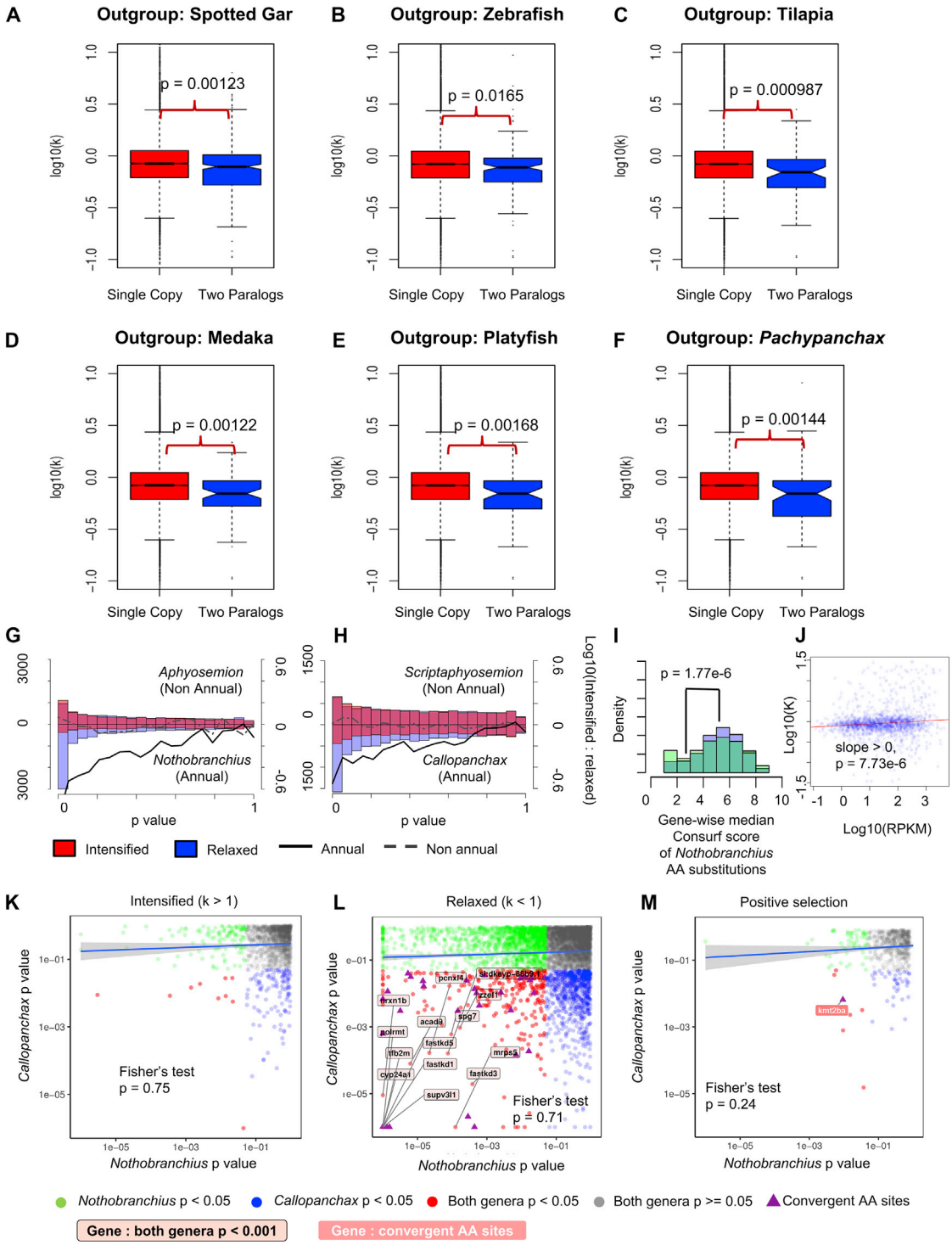
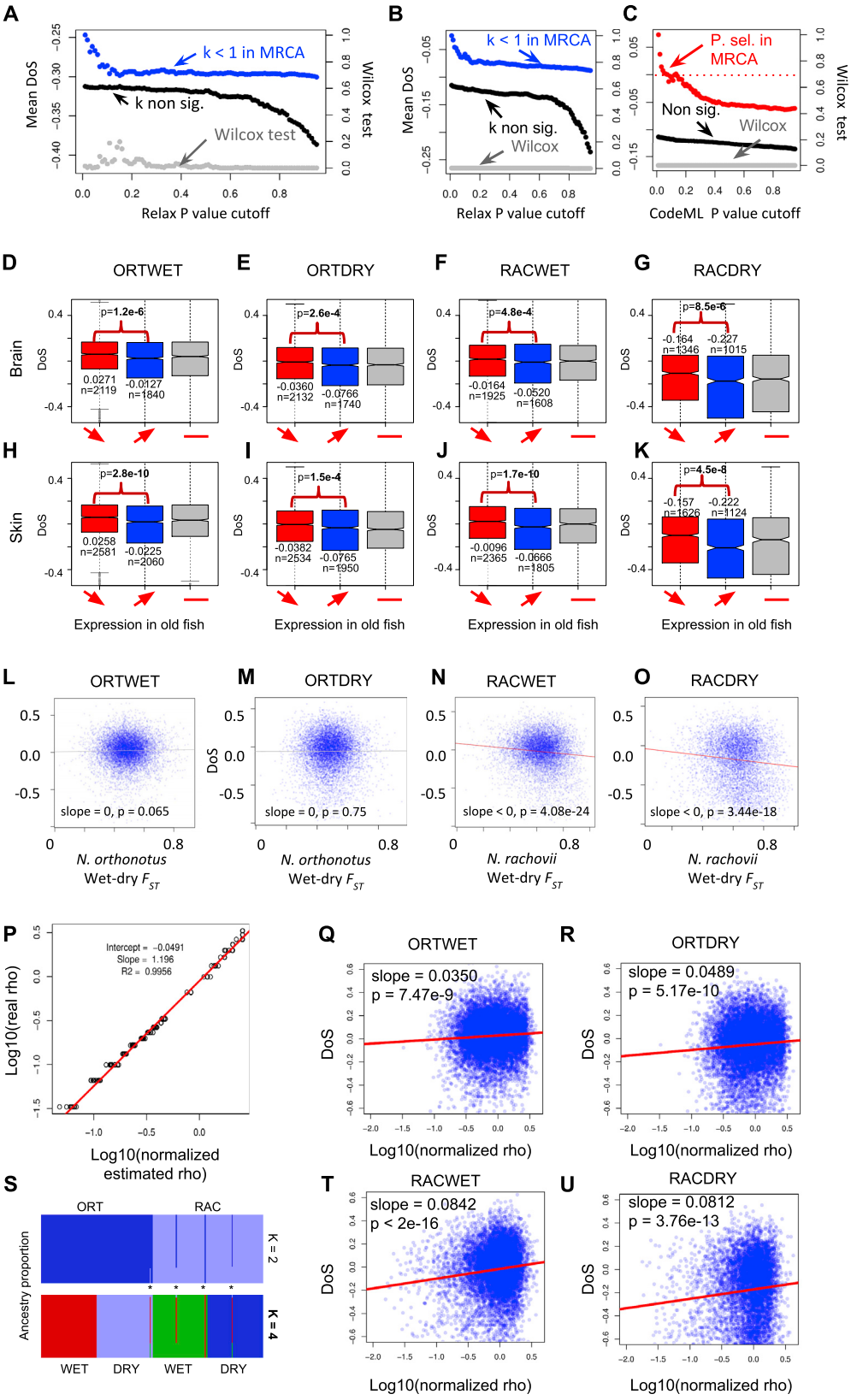


Figure S5. Related to Figure 2

Relaxation of paralogs and repetition of analyses with CMDs excluded. (A)-(F) Genes with two paralogs are on average more relaxed than single-copy genes. Different outgroup species were used. Edges of boxplots indicate the 1st and 3rd quartiles, whiskers are the data range excluding outliers. (G) and (H) CMD sites excluded, P value distributions of the RELAX tests of non-annual (top half) and annual (bottom half) sister genera. Blue bars – genes under relaxed selection (k < 1), red bars - intensified genes (k > 1). Dotted (non-annual) and solid (annual) lines are the log ratios of k > 1: k < 1 gene counts. (I) CMD sites excluded, distributions of the gene-wise median Consurf scores in relaxed (blue bars) and non-significant (green bars) genes. Higher Consurf scores suggest more conservation in other vertebrate homologs and a stronger deleterious effect upon mutation. Wilcoxon rank sum test. (J) Average gene expression level is positively correlated with the

(legend continued on next page)

relaxation parameter k when CMDs were excluded. (K) After CMD exclusion, intensified gene sets detected by RELAX no longer significantly overlap between *Callopanchax* and *Nothobranchius*. (L) Numerous relaxed genes overlap, but the overlap is not statistically significant after CMD exclusion. However, there is an enrichment for convergent amino acid substitutions (p = 4.067e-7). (M) Positively selected gene sets detected by CodeML in two annual genera *Callopanchax* and *Nothobranchius* no longer significantly overlap after CMD exclusion.



(legend on next page)

Figure S6. Related to Figure 3

Additional analyses of population resequencing data. (A) The genes detected to be relaxed in the most recent common ancestor (MRCA) branch of *Nothobranchius* have higher average DoS in the population MK tests (DAF = 0.20-0.95, outgroup NVG, DoS > 0 excluded, CMD sites excluded from RELAX and MK tests) compared to non-significant genes. (B) Genes detected to be relaxed in the most recent common ancestor (MRCA) branch of *Nothobranchius* are more likely to have higher DoS (DoS, derived allele freq. 0.20-0.95, outgroup NVG, CMD sites excluded from the RELAX and MK tests). (C) Genes detected as positively selected in the MRCA branch of *Nothobranchius* are more likely to have a positive DoS (same filters as (B)). (D)-(K), Genes highly expressed in young age (DESeq, $p < 0.05$) experienced less relaxed selection in brain and skin as measured by DoS (NVG outgroup, CMD sites and DAF < 0.20 excluded) in 4 populations. Edges of boxplots indicate the 1st and 3rd quartiles, whiskers are the data range excluding outliers. Wilcoxon rank sum test. (L) and (M) Direction of selection (DoS) is not correlated with F_{ST} between two *N. orthonotus* populations. (N) and (O) DoS (FTH outgroup, 5%–95% allele frequency, CMD included) is strongly negatively correlated with F_{ST} between two *N. rachovii* populations. Repeatable with NVG outgroup (20%–95% allele frequency, CMD excluded), (N) p = 4.84e-15, (O) p = 1.43e-10. (P) Correlation between LDhat-inferred normalized local recombination rate (rate / mean (chromosome rate) with simulated recombination rates on a log scale. Low coverage sequencing data (2.3X / individual) were simulated by scrm, and genotypes were imputed by Reveel. Despite a down-bias of the estimate due to incorrect heterozygosity calls, the inferred rates are tightly correlated with the true rate ($R^2 = 0.9956$) across two orders of magnitude. (Q), (R), (T) and (U) Correlation between local recombination rate with DoS in protein-coding genes. Derived allele frequency 5%–95%, outgroup FTH, CMD included. Repeatable with NVG outgroup (allele frequency 20%–95%, CMDs excluded): (Q) p = 2.08e-5 (R) p = 9.24e-8 (T) p < 2e-16 (U) p = 6.33e-11. (S) Ancestry proportions inferred by fastStructure. Optimal number of clusters k = 4, corresponding to the 4 populations. Cluster number does not increase beyond k = 4. All individuals were assigned with > 99% probability to their populations except for four individuals of putative hybrid origin (asteriks), which were excluded from further analyses.

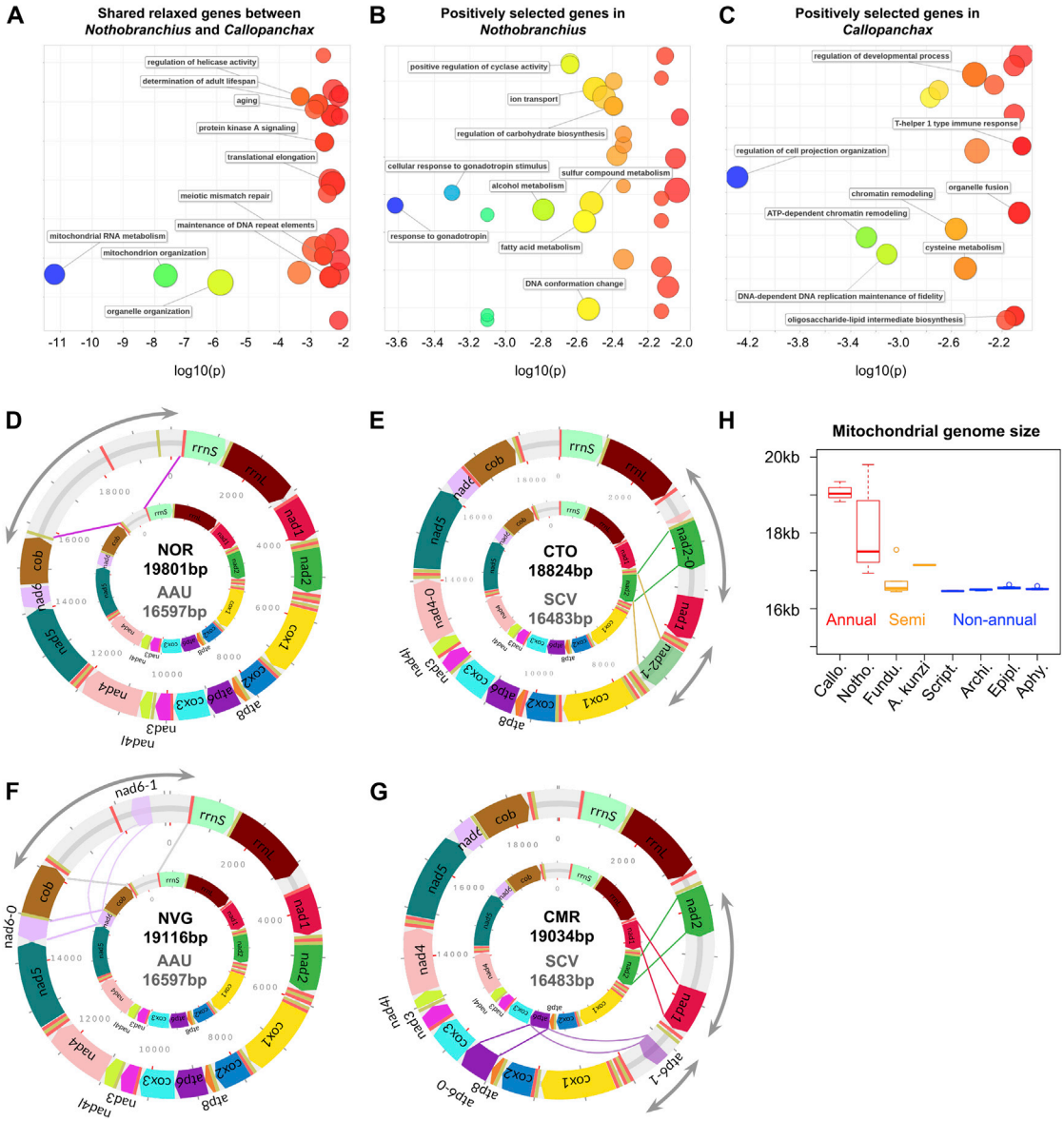


Figure S7. Related to Figure 6

GO analyses and mitochondrial genome expansion. (A) Enriched GO terms in relaxed genes shared between *Nothobranchius* and *Callopanchax* ($p < 0.05$, $k < 1$), (B) and (C) Enriched GO terms in positive selected genes in (B) *Nothobranchius* (CMDs excluded) and (C) *Callopanchax* (CMDs excluded, see Mendeley Data for gene lists and GO analyses). (D)-(G) Examples of mitochondrial expansion in (D) *N. orthonotus* (NOR), (F) *N. virgatus* (NVG), (E) *C. toddi* (CTO) and (G) *C. monroviae* (CMR). The mitochondrial genome of the closest related non-annual outgroups *A. australe* (AAU) and *S. cauveti* (SCV) are plotted inside, with solid lines delineating the putative expanded regions. In the genus *Nothobranchius*, the control region has been expanded. In *Callopanchax*, there has been genome re-arrangements between the 16 s rDNA and cox1. (H) Assembled mitochondrial genome size for each genus (clade). *Fundulopanchax* and *Aphyosemion kunzi* are Semi-annuals. Edges of boxplots indicate the 1st and 3rd quartiles, whiskers are the data range excluding outliers.