Intra-Species Differences in Population Size shape Life History and Genome Evolution

Authors: David Willemsen[1], Rongfeng Cui[1], Martin Reichard[2], Dario Riccardo Valenzano[1,3*]

Affiliations:

[1]Max Planck Institute for Biology of Ageing, Cologne, Germany.

[2]The Czech Academy of Sciences, Institute of Vertebrate Biology, Brno, Czech Republic.

[3]CECAD, University of Cologne, Cologne, Germany.

*Correspondence to: dvalenzano@age.mpg.de

Key words: life history, evolution, genome, population genetics, killifish, *Nothobranchius furzeri*, lifespan, sex chromosome, selection, genetic drift

Abstract

The evolutionary forces shaping life history trait divergence within species are largely unknown. Killifish (oviparous Cyprinodontiformes) evolved an annual life cycle as an exceptional adaptation to life in arid savannah environments characterized by seasonal water availability. The turquoise killifish (*Nothobranchius furzeri*) is the shortest-lived vertebrate known to science and displays differences in lifespan among wild populations, representing an ideal natural experiment in the evolution and diversification of life history. Here, by combining genome sequencing and population genetics, we investigate the evolutionary forces shaping lifespan among turquoise killifish populations. We generate an improved reference assembly for the turquoise killifish genome, trace the evolutionary origin of the sex chromosome, and identify genes under strong positive and purifying selection, as well as those evolving neutrally. We find that the shortest-lived turquoise killifish populations, which dwell in fragmented and isolated habitats at the outer margin of the geographical range of the species, are characterized by small effective population size and accumulate throughout the genome several small to large-effect deleterious mutations due to genetic drift. The genes most affected by drift in the shortest-lived turquoise killifish populations are involved in the WNT signalling pathway, neurodegenerative disorders, cancer and the mTOR pathway. As the populations under stronger genetic drift are the shortest-lived ones, we propose that limited population size due to habitat fragmentation and repeated population bottlenecks, by causing the genome-wide accumulation of deleterious mutations, cumulatively contribute to the short adult lifespan in turquoise killifish populations.

Main

The extent to which drift and selection shape life history trait evolution across species in nature is a fundamental question in evolutionary biology. Variations in population size among natural populations is expected to affect the rate of accumulation of advantageous and slightly deleterious gene variants, hence impacting the relative contribution of selection and drift to genetic polymorphisms[1]. Populations living in fragmented habitats, subjected to continuous and severe bottlenecks, are expected to undergo dramatic population size reduction and drift, which can significantly impact the accumulation of genetic polymorphisms in genes affecting important life history traits[2].

Among vertebrates, killifish represent a unique system, as they repeatedly and independently colonised highly fragmented habitats, characterized by cycles of rainfalls and drought[3]. While on the one hand intermittent precipitation and periodic drought pose strong selective pressures leading to the evolution of embryonic diapause, an adaptation that enables killifish to survive in absence of water[4,5], on the other hand they cause habitat and population fragmentation, promoting inbreeding and genetic drift. The co-occurrence of strong selective pressure for early-life and extensive drift characterizes life history evolution in African annual killifishes [6].

The turquoise killifish (*Nothobranchius furzeri*) is the shortest-lived vertebrate with a thoroughly documented post-embryonic life, which, in the shortest-lived strains, amounts to four months[4,5,7,8]. Turquoise killifish has recently emerged as a powerful new laboratory model to study experimental biology of aging due to its short lifespan and to its wide range of aging-related changes, which include neoplasias[9], decreased regenerative capacity[10], cellular senescence[11,12], and loss of microbial diversity[13]. At the same time, while sharing physiological adaptations that enable embryonic diapause and rapid sexual maturation, different wild turquoise

55 killifish populations display differences in lifespan, both in the wild and in captivity[14-16], making
56 this species an ideal evolutionary model to study the genetic basis underlying life history trait
57 divergence within species.

58 Characterisation of life history traits in wild-derived laboratory strains of turquoise killifish
59 revealed that while different populations have similar rates of sexual maturation[8], populations
60 from arid regions exhibit the shortest lifespans, while populations from more semi-arid regions
61 exhibit longer lifespans[8,14]. Hence, speed of sexual maturation and adult lifespan appear to be
62 independent in turquoise killifish populations. The evolutionary mechanisms responsible for the
63 lifespan differences among turquoise killifish populations are not yet clearly understood.
64 Mapping genetic loci associated with lifespan differences among turquoise killifish populations
65 showed that adult survival has a complex genetic architecture[15,17]. Here, combining genome
66 sequencing and population genetics, we investigate to what extent genomic divergence in natural
67 turquoise killifish populations that differ in lifespan is driven by adaptive or neutral evolution.

68 ***Genome assembly improvement and gene annotation***

69 To identify the genomic mechanism that led to the evolution of differences in lifespan between
70 natural populations of the turquoise killifish (*Nothobranchius furzeri*), we combined the currently
71 available reference genomes[15,18] into an improved reference turquoise killifish genome assembly.
72 Due to the high repeat content, assembly from short reads required a highly integrated and multi-
73 platform approach. We ran Allpaths-LG with all the available pair-end sequences, producing a
74 combined assembly with a contig N50 of 7.8kb, corresponding to a ~2kb improvement from the
75 previous versions. Two newly obtained 10X Genomics linked read libraries were used to correct
76 and link scaffolds, resulting in a scaffold N50 of 1.5Mb, i.e. a three-fold improvement from the
77 best previous assembly. With the improved continuity, we assigned 92.2% of assembled bases to

78 the 19 linkage groups using two RAD-tag maps[15]. Gene content assessment using the BUSCO
79 method improved "complete" BUSCOs from 91.43%[15] and 94.59%[18] to 95.20%. We mapped
80 Genbank *N. furzeri* RefSeq RNA to the new assembly to predict gene models. The predicted gene
81 model set is 96.1% for "complete" BUSCOs. The overall size of repeated regions (masked
82 regions) is 1.003 Gb, accounting for 66% of the entire genome, i.e. 20% higher than a previous
83 estimate[19].

84 ***Population genetics of natural turquoise killifish populations***

85 Natural populations of turquoise killifish occur along an aridity gradient in Zimbabwe and
86 Mozambique and populations from more arid regions are associated with shorter captive
87 lifespan[8,14]. A QTL study performed between short-lived and long-lived turquoise killifish
88 populations showed a complex genetic architecture of lifespan (measured as age at death), with
89 several genome-wide loci associated with lifespan differences among long-lived and short-lived
90 populations[15]. To further investigate the evolutionary forces shaping genetic differentiation in the
91 loci associated with lifespan among wild turquoise killifish populations, we performed pooled
92 whole-genome-sequencing (WGS) of killifish collected from four sampling sites within the
93 natural turquoise killifish species distribution, which vary in altitude, annual precipitation and
94 aridity (**Figure S1**, **Table S1**). Population GNP is located within the Gonarezhou National Park
95 at high altitude and in an arid climate (Koeppen-Geiger classification "BWh", **Figure S1**), in a
96 region at the outer edge of the turquoise killifish distribution (**Figure S1**)[20-22], which corresponds
97 to the place of origin of the "GRZ" laboratory strain, which has the shortest lifespan of all
98 laboratory strains of turquoise killifish[14,15]. Population NF414 (MZCS 414) is located in an arid
99 area in the center of the Chefu river drainage in Mozambique ("BWh", **Figure S1**)[20-22], and
100 population NF303 (MZCS 303) is located in a semi-arid area in transition to more humid climate

101 zones in the center of the Limpopo river drainage system (Koeppen-Geiger classification "BSh",
102 **Figure S1**)[20-22]. Altitude among localities ranges from 344 m (GNP) to 68 m (NF303, **Figure S1a**
103 and **Table S1**). The temporary habitat of turquoise killifish populations differs in terms of altitude
104 and aridity, as the ephemeral pools at higher altitude are drained earlier and persist for shorter
105 time, while water bodies in habitats at lower altitude last longer[14]. Population GNP is therefore
106 named "dry", population NF414 is named "intermediate" and population NF303 "wet"
107 throughout the manuscript.

108

109 ***High genetic differentiation and contrasting population demography between dry and wet***
110 ***populations***

111 We asked whether populations from dry, intermediate and wet areas, corresponding to shorter
112 and progressively longer lifespan, differ in genetic variability. We calculated genome-wide
113 estimates of average pairwise difference (π) and genetic diversity ($\theta_{\text{Watterson}}$) based on 50kb-non-
114 overlapping sliding windows using PoPoolation[23]. We found that π and $\theta_{\text{Watterson}}$ decrease from
115 wet to dry population ($\theta_{\text{Watterson GNP}}$: 0.0011, $\theta_{\text{Watterson NF414}}$: 0.0036, $\theta_{\text{Watterson NF303}}$: 0.0072; π_{GNP}:
116 0.0009, π_{NF414}: 0.0031, and π_{NF303}: 0.0054). To infer the genetic distance between the
117 populations, we computed the genome-wide pairwise genetic differentiation between populations
118 using F_{ST}[24]. Overall, the genetic differentiation between populations ranged between 0.14 and
119 0.26 and was the highest between population GNP (dry) and population NF303 (wet) (**Figure**
120 **1a**).

121 Next, we inferred the demographic history of the populations using pairwise sequentially
122 Markovian coalescent (PSMC) by resequencing at high-coverage single individuals for each
123 population[25]. The population GNP (dry) experienced a strong population decline starting

124 approximately 150k generations ago, a result consistent in both sequenced individuals from the
125 two sampling sites (GNP-G1-3 and GNP-G4, **Figure 1a**). In contrast to the demographic history
126 in GNP, we found indications for recent population expansions in populations from the center of
127 the Chefu and Limpopo basins clades. Analysis of population NF414 (intermediate) (**Figure 1a**,
128 NF414-Y and NF414-R) and NF303 (wet) (**Figure 1a**, blue line) shows population expansion
129 until recent time (~50k generations ago). To infer the effective population size (N_e) of the
130 populations, we used the published mutational rate of 2.6321e−9 per base pair per generation for
131 *Nothobranchius* computed via dated phylogeny and $\theta_{\text{Watterson}}$ [6]. In line with the decrease in genetic
132 diversity from wet to dry population, we found a decrease in N_e estimates (107221.8, 338849.48
133 and 683693.25 for GNP, NF414 and NF303, respectively; **Figure 1b**). Hence, our findings show
134 that dry populations from the outer edge of the species distribution show lower genetic diversity
135 and smaller effective population size compared to population from intermediate and more wet
136 regions.

137

138 ***Genetic differentiation among turquoise killifish populations***

139 To test whether regions underlying longevity QTL in turquoise killifish[15,17] display a genetic
140 signature for positive or purifying selection, we took advantage of the improved turquoise
141 killifish genome assembly and the newly sequenced wild turquoise killifish populations (**Figure**
142 **2**). The strongest QTL for lifespan differences among long-lived and short-lived populations
143 mapped on the sex chromosome[15,17], in proximity to the sex determining locus[15]. To identify a
144 genomic signature of strong selection, we performed an outlier approach based on the pairwise
145 genetic differentiation index (F_{ST}). To find highly differentiated regions that may underlie
146 positive selection in natural turquoise killifish populations, we scanned for regions with elevated

147 genetic differentiation between pairs of populations, i.e. exceeding the 0.995 quantile of Z-
148 transformed non-overlapping 50kb sliding windows of F_{ST}. To find regions under purifying
149 selection, we scanned for regions with lowered genetic differentiation among populations, i.e.
150 below the 0.005 quantile of Z-transformed non-overlapping 50kb sliding windows of F_{ST}
151 (**TableS7**). The outlier approach did not reveal clear signatures of positive or purifying selection
152 based on genetic differentiation in the four main clusters associated with lifespan in experimental
153 strains of turquoise killifish (**Figure 2**). We then analysed genomic regions carrying signatures of
154 positive and purifying selection in the natural turquoise killifish populations irrespective of the
155 QTL regions (**Figure 2**). The F_{ST} outlier approach led to the identification of several potential
156 regions under divergent selection between populations, in particular between the intermediate and
157 wet populations (**Table S4**) and only two between the dry and wet populations (**Table S5)**. Genes
158 significantly different and within regions of larger genetic differentiation based on Z-transformed
159 non-overlapping sliding windows of F_{ST} were located on chromosomes 6 and 10. The region on
160 chromosome 6 includes the gene *slc8a1*, which contains mutations with significant difference in
161 allele frequencies between the wet and intermediate population (Fisher's exact test implemented
162 in PoPoolation; adjusted p value < 0.001). The region on chromosome 10 contains four genes:
163 XM_015941868, XM_015941869, *lss* and *hibch.* All genes under the major F_{ST} peak on
164 chromosome 10 showed significant difference in allele frequencies between the intermediate and
165 wet population (Fisher's exact test; adjusted p value < 0.001) and additionally, *hibch* had
166 significantly different allele frequencies between the dry and wet population (Fisher's exact test;
167 adjusted p value < 0.001). Genes under F_{ST} peaks between populations that differ in lifespan, are
168 not necessarily causally involved in lifespan differences between populations, as sequence
169 differences could segregate in populations due to population structure and drift. However, to test
170 whether the genes located in genomic regions that are significantly divergent between

171 populations could be functionally involved in age-related phenotypes, we investigated whether
172 gene expression in these genes varied as a function of age. Analysing available turquoise killifish
173 longitudinal RNA-datasets generated in liver, brain and skin[26], we found that *hibch*, *lss* and
174 *slc8a1* are differentially expressed between adult and old killifish (**Table S10**, adjusted p value <
175 0.01). *hibch*, *lss and slc8a1* are involved in amino acid metabolism[27], biosynthesis of
176 cholesterol[28], and proton-mediated accelerated aging[29], respectively. Gene XM_015956265
177 (ZBTB14) is the only gene that is an F_{ST} outlier and that is differentially expressed in adult vs.
178 old individuals between at least two populations in all tissues (liver, brain and skin).
179 XM_015956265 encodes a transcriptional modulator with ubiquitous functions, ranging from
180 activation of dopamine transporter to repression of MYC, FMR1 and thymidine kinase
181 promoters[30]. However, although genomic regions that have sequence divergence between
182 turquoise killifish populations contain genes that are differentially expressed during ageing in
183 different tissues, whether any of these genes are causally involved in modulating ageing-related
184 changes between turquoise killifish wild populations still remains to be assessed. Based on the
185 outlier approach, we found two genomic regions with low genetic differentiation between all
186 pairs of populations, indicating strong purifying selection. The first region is located on the sex
187 chromosome and contains the putative sex determining gene *gdf6*[18], which is hence conserved
188 among these populations. This same region also contains *sybu*, a maternal-effect gene associated
189 with the establishment of embryo polarity[31]. The second region under low genetic differentiation
190 is located on chromosome 9 and harbours the genes XM_015965812 (*abi2-like*), *cnot11* and *lcp1*,
191 which are involved in phagocytosis[32], mRNA degradation[33] and cell motility[34], respectively.

192 ***Evolutionary origin of the sex chromosome***

193 Since we found reduced genetic differentiation among populations in the chromosomal region
194 containing the putative sex-determining gene in the sex chromosome, we used synteny analysis
195 and the new genome assembly to investigate the genomic events that led to evolution of this
196 chromosomal region (**Figure 3**). We found that the structure of the turquoise killifish sex
197 chromosome is compatible with a chromosomal translocation within an ancestral chromosome
198 and a fusion event between two chromosomes. The translocation event within an ancestral
199 chromosome corresponding to medaka´s chromosome 16 and platyfish´s linkage group 3 led to a
200 repositioning of a chromosomal region containing the putative sex-determining gene *gdf6*
201 (**Figure 3b**). The fusion of the translocated chromosome with a chromosome corresponding to
202 medaka chromosome 8 and platyfish linkage group 16, possibly led to the origin of turquoise
203 killifish sex chromosome. We could hence reconstruct a model for the origin of the turquoise
204 killifish sex chromosome (**Figure 3c**), which parsimoniously places a translocation event before a
205 fusion event. The occurrence of two major chromosomal rearrangements, namely a translocation
206 and a fusion, could have then contributed to suppressing recombination around the sex-
207 determining region[15,35].

208 ***Relaxed selection in turquoise killifish populations***

209 Since we could not identify specific signatures of genetic differentiation in the genomic regions
210 associated to longevity from previous QTL mapping, we asked whether other evolutionary forces
211 than directional selection may underlie differences in survival among wild turquoise killifish
212 populations. The difference in the recent and past demography between populations (**Figure 1**)
213 led us to ask whether demography could have led to evolutionary changes on genome-wide scale
214 between natural populations. For each population, we calculated the fraction of substitutions
215 driven to fixation by positive selection since divergence from the outgroup species

216 *Nothobranchius orthonotus* (NOR) using the asymptotic McDonald-Kreitman $\boldsymbol{\alpha}$[36]. Using NOR as
217 an outgroup, we infer the fraction of positive selection by pooling all coding sites (**Figure 4a**).
218 SNPs were called with the program SNAPE[37], which specifically deals with pooled sequencing.
219 We only included SNPs with a derived frequency between 0.05-0.95 and performed stringent
220 filtering. The asymptotic McDonald-Kreitman $\boldsymbol{\alpha}$ ranged from -0.21 to -0.01 in comparison to the
221 very closely related sister species *N. orthonotus*, confirming limited genome-wide positive
222 selection since divergence from *N. orthonotus* (**Figure 4a**). The population GNP, located in an
223 arid region at higher altitude and associated with the shortest recorded lifespan, shows the lowest
224 asymptotic McDonald-Kreitman $\boldsymbol{\alpha}$, as well as lower McDonald-Kreitman $\boldsymbol{\alpha}$ values throughout all
225 derived frequency bins, potentially suggesting a higher load of slightly deleterious mutations
226 segregating in this population (**Figure 4a**). Using as an outgroup species another annual killifish
227 species, *Nothobranchius rachovii* (NRC), we confirmed the lowest asymptotic McDonald-
228 Kreitman $\boldsymbol{\alpha}$ value in the dry population GNP (**Figure 4b**). Additionally, using *Nothobranchius*
229 *rachovii* (NRC) as outgroup species, the asymptotic McDonald-Kreitman $\boldsymbol{\alpha}$ ranged from -0.06 to
230 0.23 among populations, indicating that more alleles were driven to fixation by positive selection
231 in the ancestral lineage leading to *Nothobranchius furzeri* and *Nothobranchius orthonotus*. In
232 particular, the wet population NF303 had the highest asymptotic McDonald-Kreitman $\boldsymbol{\alpha}$ value
233 (**Figure 4b**). Using both *N. orthonotus* and *N. rachovii* as outgroups, we found that the dry GNP
234 population had the lowest McDonald-Kreitman $\boldsymbol{\alpha}$ values at the low derived frequency bins,
235 potentially consistent with a genome-wide accumulation of slightly deleterious mutations in these
236 isolated populations.

237 To directly estimate the fitness effect of gene variants associated with each population, we
238 analysed population-specific genetic polymorphisms to assign mutations as beneficial, neutral or
239 detrimental, and determine the distribution of fitness effect (DFE)[38]. Consistently with the overall

240 lower McDonald-Kreitman $\boldsymbol{\alpha}$ values throughout all derived frequency bins, we found more
241 mutations assigned as the slightly deleterious category in the dry GNP population, compared to
242 the other two populations (indicated by the higher number of deleterious SNPs in proximity to
243 $4N_eS \sim 0$ in the GNP population, **Figure 4d, TableS8**). To further infer the effect of the putative
244 deleterious mutations on protein function, we used the new turquoise killifish genome assembly
245 as a reference and adopted an approach that, by analysing sequence polymorphism among
246 populations, predicts functional consequences at the protein level[39]. We found that the proportion
247 of mutations causing a change in protein function is significantly larger in the GNP population
248 compared to populations NF414 and NF303 (Chi-square test: $P_{GNP\text{-}NF303}$<1.87e-119, $P_{GNP\text{-}NF414}$<
249 4.96e-57, $P_{NF303\text{-}NF414}$< 3.51e-35, **Figure 4e**). Additionally, the mutations with predicted
250 deleterious effects on protein function reached also higher frequencies in the dry population GNP
251 (**Figure 4c**). To further investigate the impact of mutations on protein function, we calculated the
252 Consurf [40-43] score, which determines the evolutionary constraint on an amino acid, based on
253 sequence conservation. Mutations at amino acid positions with high Consurf score (i.e. otherwise
254 highly conserved) are considered to be more deleterious. We found that the dry population GNP
255 had a significantly higher mean Consurf score for mutations at non-synonymous sites in
256 frequency bins from 5%-20% up to 40%-60%, compared to populations NF414 (intermediate)
257 and NF303 (wet) (**Figure S2**). The mutations in the dry GNP population had significantly higher
258 Consurf scores than the other populations using both outgroup species *N. orthonotus* and *N.*
259 *rachovii* (**Figure S2**). Upon exclusion of potential mutations at neighbouring sites (CMD: codons
260 with multiple differences), CpG hypermutation and genes containing mutations with highly
261 detrimental effect on protein function based on SnpEFF analysis, the dry population GNP had
262 higher mean Consurf score at the low frequency bin (**Figure S2**, **Table S11-12**). To note, we also
263 found a significantly higher average Consurf score at synonymous sites in GNP at low derived

264 frequencies (Figure S2, Supplementary Table S11 and S12), possibly suggestive of an overall
265 higher mutational rate in GNP.

266 ***Relaxation of selection in age-related disease pathways***

267 Computing the gene-wise direction of selection (DoS)[44] index, which enables to score the
268 strength of selection based on the count of mutations in non-synonymous and synonymous sites,
269 we found support to the hypothesis that the dry, short-lived population GNP has significantly
270 more slightly deleterious mutations segregating in the population, compared to the populations
271 NF414 and NF303 (**Figure 5a**, Median NOR: GNP: -0.17, NF414: -0.02, NF303: -0.01; Median
272 NRC: GNP: -0.14, NF414: 0.00, NF303: 0.00; Wilcoxon rank sum test: NOR: $P_{GNPNF303}$<2.21e-
273 105, $P_{GNP\text{-}NF414}$< 1.19e-76, $P_{NF303\text{-}NF414}$< 1.39e-06; NRC: $P_{GNP\text{-}NF303}$<4.61e-179, $P_{GNP\text{-}NF414}$< 1.42e-
274 100, $P_{NF303\text{-}NF414}$< 5.96e-22), indicating that purifying selection is relaxed in GNP. We calculated
275 DoS in all populations using independently as outgroup species *N. orthonotus* and *N. rachovii*
276 (**Figure 5a**).

277 To assess whether specific biological pathways were significantly more impacted by the
278 accumulation of slightly deleterious mutations, we performed pathway overrepresentation
279 analysis. We found a significant overrepresentation in the lower 2.5th DoS (i.e. genes under
280 relaxation of selection) in the GNP population for pathways associated with age-related diseases,
281 including *gastric cancer, breast cancer, neurodegenerative disease, mTOR signalling* and WNT
282 signalling (q-value <0.05, **Figure5b**, **Table S9**). Overall, relaxed selection in the dry GNP
283 population affected accumulation of deleterious mutations in age-related and in the WNT
284 pathway. Analysing the pathways affected by genes within the upper 2.5th DoS values –
285 corresponding to genes undergoing adaptive evolution – we found a significant enrichment for
286 mitochondrial pathways – potentially compensatory[6] – in population NF303 (**Figure5b**, **Table**

287 S9). Overall, our results show that differences in effective population size among wild turquoise
288 killifish are associated with an extensive relaxation of purifying selection, significantly affecting
289 genes involved in age-related diseases, and which could have cumulatively contributed to
290 reducing survival.

291 ## Discussion

292 The turquoise killifish (*Nothobranchius furzeri*) is the shortest-lived known vertebrate and while
293 its natural populations show similar timing for sexual maturation, exhibit differences in lifespan
294 along a cline of altitude and aridity in south-eastern Africa[8,14]. Here we generate an improved
295 genome assembly (NFZ v2.0) in turquoise killifish (*Nothobranchius furzeri*) and study the
296 evolutionary forces shaping genome evolution among natural populations.

297 Using the new turquoise killifish genome assembly and synteny analysis with medaka and
298 platyfish, we reconstructed the origin of the turquoise killifish sex chromosome, which appears to
299 have evolved through two independent chromosomal events, i.e. a fusion and a translocation
300 event.

301 Using the new genome assembly and pooled sequencing of natural turquoise killifish populations,
302 we found that genetic differentiation among populations of the short-lived turquoise killifish is
303 consistent with differences in demographic constraints. While we found that strong purifying
304 selection maintains low genetic diversity among populations at genomic regions underlying key
305 species-specific traits, such as in proximity to the sex-determining region, demography and
306 genetic drift largely shape genome evolution, leading to relaxation of selection and the
307 accumulation of deleterious mutations. We showed that isolated populations from an arid region,
308 dwelling at higher altitude and characterised by shorter lifespan, experienced extensive
309 population bottlenecking and a sharp decline in effective population size. We found that
310 relaxation of selection in highly drifted populations significantly affected the accumulation of
311 deleterious gene variants in pathways associated with neurodegenerative diseases and WNT-
312 signalling (**Figure 5**). While simple traits, such as male tail colour and sex have a simple genetic
313 architecture among turquoise killifish populations[15,35], we find that the complex genetic

314 architecture of lifespan differences among killifish populations[15] is entirely compatible with
315 genome-wide relaxation of selection. Additionally, the absence of genomic signature of positive
316 selection in genomic regions underlying survival QTL in killifish suggest that, rather than
317 directional selection, the neutral accumulation of deleterious mutations in short-lived populations
318 may be the evolutionary mechanism underlying survival differences among turquoise killifish
319 populations. The "antagonistic pleiotropy" evolutionary theory of ageing states that positive
320 selection could lead to the fixation of gene variants that, while overall beneficial for fitness, could
321 reduce survival and reproductive capacity in late life[45]. The lack of genomic signature of positive
322 selection at the genomic regions underlying survival QTL in turquoise killifish rather suggests
323 that accumulation of deleterious mutations may have played a key role in shaping genome and
324 phenotype differences among natural turquoise killifish populations. Historical fluctuations in the
325 size of natural turquoise killifish populations, especially in isolated and populations living in
326 more arid and elevated habitats, cause decreased efficiency of the strength of natural selection,
327 ultimately contributing to increased load of deleterious gene variants, preferentially in genes
328 associated with ageing-related diseases and in the WNT pathway.

329 Our findings highlight the role of demographic constraints in shaping life history within species.

330

331

332

333 ## Materials and Methods

334 Merging and Improvement of the Turquoise killifish genome assembly

335 ***10x Genomics read clouds***

336 A single GRZ male individual was sacrificed with MS222 (Sigma-Aldrich, Steinheim, Germany).
337 Blood was drawn from the heart and high molecular weight DNA was isolated with Qiagen
338 MagAttract kit following manufacturer's instructions. Gemcode v2 DNA library generation was
339 performed by Novogene (Beijing, China). Briefly, a proportion of the sample was run on a pulse
340 field agarose gel to confirm high molecularity > 100kb. Based on a genome size estimate of
341 1.54Gb (half of human genome), 0.6ng of DNA was used to construct 2 Gemcode libraries,
342 sequenced on two HiSeq X lanes to obtain a raw coverage of approximately 60X each. The
343 reported input molecular length by SuperNova[46] was 118kb for library 1 and 60.73kb for library
344 2. Both libraries were used to correct and scaffold the Allpath-LG assembly (see below), and
345 library 1 was also de novo assembled with the SuperNova assembler v.2 with default parameters.
346 The SuperNova assembly totaled 802.6Mb, with a contig N50 of 19.65kb, scaffolded into 6.78
347 thousand scaffolds with an N50 of 3.83Mb. Despite high continuity, however, the BUSCO[47]
348 metrics are much lower than the Allpath-LG assemblies.

349 ***Nanopore long reads***

350 DNA was extracted from a single GRZ male individual's muscle tissue by griding in liquid
351 nitrogen followed by phenol-chlorofom extraction (Sigma). The rapid sequencing kit (SQK-
352 RAD004) and the ligation kit (SQK-LSK108) were sued to prepare 6 libraries and were
353 sequenced on 6 MinION flow cells (R9.4.1). These runs yielded a total of 3.3 Gb of sequences

354 after trimming and correction by HALC[48]. For correction, Allpath-LG contigs (see below) and
355 short reads from the 10X genomic run were used.

356 ***Allpath-LG assembly***

357 Two independent short read datasets were previous collected for the GRZ strain of
358 *Nothobranchius furzeri*. Allpath-LG[49] was used on the pooled datasets. Together, 4 illumina short
359 read pair-end libraries with a fragment size distribution from 158bp to 179bp were used to
360 construct the contigs (sequence coverage 191.9X, physical coverage 153.5X), and 22 pair-end
361 and mate pair libraries distributed at 92bp, 135bp, 141bp, 176bp, 267bp, 2kb, 3kb 5kb and 10kb
362 were used for the scaffolding step (sequence coverage 135.7X, physical coverage 453.8X). The
363 published BAC library ends[18] with an insert size of 112kb were also included in the ALLPaths-
364 LG run (physical coverage 0.6X). The resulting assembly has a total contig length of
365 823,583,106bp distributed in 151,307 contigs > 1kb, with an N50 of 7.8kb. The total scaffold
366 length is 943,793,727bp distributed in 7830 scaffolds with an N50 of 421kb (with gaps). The
367 resulting assembly was further scaffolded by ARCS v1.0[50] + LINKS v1.8.5[51] with the following
368 parameters: arcs -e 50000 -c 3 -r 0.05 -s 98 and LINKS -m -d 4000 -k 20 -e 0.1 -l 3 -a 0.3 -t 2 -o
369 0 -z 500 -r -p 0.001 -x 0. This increased the scaffold N50 to 1.527 Mb. Next, scaffolds were
370 assigned to the RAD-tag linkage map[15] collected from a previous study with Allmaps [52], using
371 equal weight for the two independent mapping crosses. This procedure assigned 90.6% of the
372 assembled bases in 1131 scaffolds to 19 linkage groups, in which 76.6% can be oriented.
373 Misassemblies were corrected with the 10X genomic read cloud. Read clouds were mapped to the
374 preliminary assembly with longranger v2.1.6 using default parameters, and a custom script was
375 used to scan for sudden drops in barcode shares along the assembled linkage groups. The
376 scaffolds were broken at the nearest gap of the drop in 10x barcodes. The same ARCS + LINKS

377 pipeline was again run on the broken scaffolds, increasing the scaffold N50 to 1.823 Mb. Next,
378 BESST_RNA (https://github.com/ksahlin/BESST_RNA) was used to further scaffold the
379 assembly with RNASeq libraries, Allmaps was again used to assign the fixed scaffolds back to
380 linkage groups, increasing the assignable bases to 92.2% (879Mb) with 80.3% (765Mb) with
381 determined orientation. The assembly was again broken with longranger and reassigned to LG
382 with Allmaps, and the scaffolds were further partitioned to linkage groups due to linkage of some
383 left-over scaffolds with an assigned scaffold. Each partitioned scaffold groups were subjected to
384 the ARCS + LINKS pipeline again, to constraint the previously unassigned scaffolds onto the
385 same linkage group. Allmaps was run again on the improved scaffolds, resulting in 94.5%
386 (903.4Mb) of bases assigned and 89.1% (852Mb) of bases oriented. Longranger was run again,
387 visually checked and compared with the RADtag markers. Eleven mis-oriented positions were
388 identified and corrected. Gaps were further patched by GMCloser[53] with ~2X of nanopore long
389 reads corrected by HALC using BGI500 short PE reads with the following parameters: gmcloser
390 --blast --long_read --lr_cov 2 -l 100 -i 466 -d 13 --min_subcon 1 --min_gap_size 10 --iterate 2 --
391 mq 1 -c. The corrected long reads not mapped by GMCloser were assembled by CANU [54] into
392 7.9Mb of sequences, which are likely unassigned repeats.

393 ***Meta Assembly***

394 Five assemblies were integrated by MetAssembler[55] in the following order (ranked by BUSCO
395 scores) using a 20kb mate pair library: 1) The improved Allpaths-LG assembly assigned to
396 linkage groups produced in this study, 2) A previously published assembly with Allpaths-LG and
397 optical map[18] 3) A previously published assembly using SGA[15], 4) The SuperNova assembly
398 with only 10x Genomic reads and 5) Unassigned nanopore contigs from CANU. The final
399 assembly NFZ v2.0 has 911.5Mb of scaffolds assigned to linkage groups. Unassigned scaffolds

400 summed up to 142.2Mb, yielding a total assembly length of 1053.7Mb, approximately 2/3 of the
401 total genome size of 1.53Gb. The final assembly has 95.2% complete and 2.24% missing
402 BUSCOs.

403 ***Mapping of NCBI genbank gene annotations***

404 RefSeq mRNAs for the GRZ strain (PRJNA314891, PRJEB5837) were downloaded from
405 GenBank[56], and aligned to the assembly with Exonerate[57]. The RefSeq mRNAs have a BUSCO
406 score of 98.0% complete, 0.9% missing. The mapped gene models resulted in a BUSCO score of
407 96.1% complete, 2.1% missing.

408 ***Pseudogenome assembly generation***

409 The pseudogenomes for *Nothobranchius orthonotus* and *Nothobranchius rachovii* were generated
410 from sequencing data and the same method used in Cui *et al*. [6]. Briefly, the sequencing data were
411 mapped to the NFZ v2.0 reference genome by BWA-mem v0.7.12 in PE mode[58,59]. PCR
412 duplicates were marked with MarkDuplicates tool in the Picard (version 1.119,
413 http://broadinstitute.github.io/picard/) package. Reads were realigned around INDELs with the
414 IndelRealigner tool in GATK v3.4.46[60]. Variants were called with SAMTOOLS v1.2[61] mpileup
415 command, requiring a minimal mapping quality of 20 and a minimal base quality of 25. A
416 pseudogenome assembly was generated by substituting reference bases with the alternative base
417 in the reads. Uncovered regions, INDELs and sites with >2 alleles were masked as unknown "N".
418 The allele with more supporting reads was chosen at biallelic sites.

419 ***Mapping of longevity and sex quantitative trait loci***

420 The quantitative trait loci (QTL) markers published in Valenzano *et al.*[15] were directly provided
421 by Dario Riccardo Valenzano. In order to map the markers associated to longevity and sex, a

422 reference database was created using BLAST[62]. The nucleotide database was created with the
423 new reference genome of *N. furzeri* (NFZ v2.0). Subsequently, the QTL marker sequences were
424 mapped to the database. Only markers with full support for the total length of 95 bp were
425 considered as QTL markers.

426 ***Synteny analysis***

427 Synteny analysis was performed using orthologous information from Cui *et al.* [6] determined by
428 the UPhO pipeline[63]. For this, the 1-to-1 orthologous gene positions of the new turquoise killifish
429 reference genome (NFZ v2.0) were compared to two closely related teleost species, *Xiphophorus*
430 *maculatus* and *Oryzias latipes*. Result were visualized using Circos[64] for the genome-wide
431 comparison and the *genoPlotR* package[65] in R for the sex chromosome synteny analysis. Synteny
432 plots for orthologous chromosomes of *Xiphophorus maculatus* and *Oryzias latipes* were
433 generated with Synteny DB (http://syntenydb.uoregon.edu) [66].

434 ***Koeppen-Geiger index and bioclimatic variables***

435 The Koeppen-Geiger classification data was taken from Peel *et al.*[67] and the altitude, precipitation
436 per month, and the bioclimatic variables were obtained from the Worldclim database (v2.0[68]).
437 The monthly evapotranspiration was obtained from Trabucco and Zomer[69]. Aridity index was
438 calculated based on the sum of monthly precipitation divided by sum of monthly
439 evapotranspiration. Maps in FigureS1 were generated with QGIS version 2.18.20 combined with
440 GRASS version 7.4[70], the Koeppen-Geiger raster file, data from Natural Earth, and the river
441 systems database from Lehner *et al.*[71].

442 ***DNA isolation and pooled population sequencing***

443 The ethanol preserved fin tissue was washed with 1X PBS before extraction. Fin tissue was
444 digested with 10 µg/mL Proteinase K (Thermo Fisher) in 10 mM TRIS pH 8; 10mM EDTA; 0.5
445 SDS at 50°C overnight. DNA was extracted with phenol-chloroform-isoamylalcohol (Sigma)
446 followed by a washing step with chloroform (Sigma). Next, DNA was precipitated by adding 2.5
447 volume of chilled 100% ethanol and 0.26 volume of 7.5M Ammonium Acetate (Sigma) at -20°C
448 overnight. DNA was collected via centrifugation at 4°C at 12000rpm for 20 minutes. After a final
449 washing step with 70% ice-cold ethanol and air drying, DNA was eluted in 30µl of nuclease-free
450 water. DNA quality was checked on 1 agarose gels stained with RotiSafe (Roth) and a UV-VIS
451 spectrometer (Nanodrop2000c, Thermo Scientific). DNA concentration was measured with Qubit
452 fluorometer (BR dsDNA Assay Kit, Invitrogen). For each population, the DNA of the individuals
453 were pooled at equimolar contribution (GNP_G1_3, GNP_G4 N=29; NF414, NF303 N=30).
454 DNA pools were given to the Cologne Center of Genomic (CCG, Cologne, Germany) for library
455 preparation. The total amount of DNA provided to the sequencing facility was 3.2 µg per pooled
456 population sample. Libraries were sequenced with 150bp x 2 paired-ends on the HiSeq4000.
457 Sequencing of pooled samples resulted in a range of 419 - 517 million paired-end reads for each
458 population (**Table S2**).

459 ***Mapping of pooled sequencing reads***

460 Raw sequencing reads were trimmed using Trimmomatic-0.32 (ILLUMINACLIP:illumina-
461 adaptors.fa:3:7:7:1:true, LEADING:20, TRAILING:20, SLIDINGWINDOW:4:20,
462 MINLEN:50[72]. Data files were inspected with FastQC (version 0.11.22,
463 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmed reads were subsequently
464 mapped to the reference genome with BWA-MEM v0.7.12[58,59]. The SAM output was converted
465 into BAM format, sorted, and indexed via SAMTOOLS v1.3.1[61]. Filtering and realignment was

466 conducted with PICARD v1.119 (http://broadinstitute.github.io/picard/) and and GATK[60].
467 Briefly, the reads were relabelled, sorted, and indexed with AddOrReplaceReadGroups.
468 Duplicated reads were marked with the PICARD feature MarkDuplicates and reads were
469 realigned with first creating a target list with RealignerTargetCreator, second by IndelRealigner
470 from the GATK suite. Resulting reads were again sorted and indexed with SAMTOOLS. For
471 population genetic bioinformatics analyses the BAM files of the pooled populations were
472 converted into the required MPILEUP format via the SAMTOOLS mpileup command. Low
473 quality reads were excluded by setting a minimum mapping quality of 20 and a minimum base
474 quality of 20. Further, possible insertion and deletions (INDELs) were identified with
475 *identifygenomic-indel-regions.pl* script from the PoPoolation package[23] and were subsequently
476 removed via the *filter-pileup-by-gtf.pl* script[23]. Coding sequence positions that were identified to
477 be putative ambiguous were removed by providing the *filter-pileup-by-gtf.pl* script a custom
478 modified GTF file with the corresponding coordinates. After adapter and quality filtering,
479 mapping to the newly assembled reference genome resulted in mean genome coverage of 35x,
480 39x, and 47x for the population NF303, NF414, and GNP, respectively (**Table S2**).

481 ***Merging sequencing reads of populations from the Gonarezhou National Park***

482 Population GNP consists of two sampling sites (GNP-G1_3, GNP-G4) with very low genetic
483 differentiation (**Figure S1c**, **Table S3**). Sequencing reads of the two populations from the
484 Gonarezhou National Park (GNP) were combined used the SAMTOOLS 'merge' command. The
485 populations GNP-G1-3 and GNP-G4 were merged together and this population was subsequently
486 denoted as GNP.

487

488 ***Estimating genetic diversity***

489 Genetic diversity in the populations was estimated by calculating the nucleotide diversity π[73] and
490 Wattersons's estimator θ[74]. Calculation of π and θ was done with a sliding window approach by
491 using the *Variance-sliding.pl* script from the PoPoolation program[23]. Non-overlapping windows
492 with a length of 50 kb with a minimum count of two per SNP, minimum quality of 20 and the
493 population specific haploid pool size were used (GNP=116; NF414=60; NF303=60). Low
494 covered regions that fall below half the mean coverage of each population were excluded
495 (GNP=23; NF414=19; NF303=18), as well as regions that exceed a two times higher coverage
496 than the mean coverage (GNP=94; NF414=77; NF303=70). The upper threshold is set to avoid
497 regions with possible wrong assemblies. Mean coverage was estimated on filtered MPILEUP
498 files. Each window had to be at least covered to 30% to be included in the estimation.

499 ***Estimation of effective population size***

500 Wattersons's estimator of θ[74] is referred to as the population mutation rate. The estimate is a
501 compound parameter that is calculated as the product of the effective population size (N_e), the
502 ploidy (2p, with p is ploidy) and the mutational rate μ ($\theta = 2pN_e\mu$). Therefore, N_e can be obtained
503 when θ, the ploidy and the mutational rate μ are known. The turquoise killifish is a diploid
504 organism with a mutational rate of 2.6321e−9 per base pair per generation (assuming one
505 generation per year in killifish [6] and θ estimates were obtained with PoPoolation (see Section
506 2.1.2)[23].

507 ***Estimating population differentiation index F_{ST}***

508 The filtered and realigned BAM files of each population were merged into a single pileup file
509 with SAMTOOLS mpileup, with a minimum mapping quality and a minimum base quality of 20.
510 The pileup was synchronized using the *mpileup2sync.jar* script from the PoPoolation2 program
511 [24]. Insertions and deletions were identified and removed with the *identify-indel-regions.pl* and

512 *filter-sync-by-gtf.pl* scripts of PoPoolation2 [24]. Again, coding sequence positions that were
513 identified to be putative ambiguous were removed by providing the *filter-pileup-by-gtf.pl* script a
514 custom modified GTF file with the corresponding coordinates. Further a synchronized pileup file
515 for genes only were generated by providing a GTF file with genes coordinates to the *create-*
516 *genewise-sync.pl* from PoPoolation2[24]. F_{ST} was calculated for each pairwise comparison (GNP vs
517 NF303, GNP vs NF414, NF414 vs NF303) in a genome-wide approach using non-overlapping
518 sliding windows of 50kb with a minimum count of four per SNP, a minimum coverage of 20, a
519 maximum coverage of 94 for GNP, 77 for NF414, and 70 for NF303 and the corresponding pool
520 size of each population (N= 116; 60; 60). Each sliding window had to be at least covered to 30%
521 to be included in the estimation. The same thresholds, except the minimum covered fraction, with
522 different sliding window sizes were used to calculate the gene-wise F_{ST} for the complete gene
523 body (window-size of 2000000, step-size of 2000000) and single SNPs within genes (window-
524 size of 1, step-size of 1). The non-informative positions were excluded from the output.
525 Significance of allele differences per base-pair within the gene-coordinates were calculated with
526 the fisher´s exact test implemented in *the fisher-test.pl* script of PoPoolation2 [24]. Calculation of
527 unrooted neighbor joining tree based on the genome-wide pairwise F_{ST} averages was performed
528 with the ape package in R[75].

529 ***Detecting signatures of selection based on F_{ST} outliers***

530 For F_{ST}-outlier detection, the pairwise 50kb-window F_{ST}-values for each comparison were Z-
531 transformed (Z_{FST}). Next, regions potentially under strong selection were identified by applying
532 an outlier approach. Outliers were identified as non-overlapping windows of 50 kb within the
533 0.5% of lowest and highest genetic differentiation per comparison. To reduce the number of
534 false-positive results, the outlier threshold was chosen at 0.5% highest and lowest percentile of

535 each pairwise genetic differentiation[76,77]. To find candidate genes within windows of highest
536 differentiation, a total of three selection criteria were used. First, the window-based Z_{FST} value
537 had to be above the 99.5th percentile of pairwise genetic differentiation. Second, the gene F_{ST}
538 value had to be above the 99.5th percentile of pairwise genetic differentiation and last, the gene
539 needed to include at least one SNP with significant differentiation based on Fisher's exact test
540 (calculated with PoPoolation2[24]; P<0.001, Benjamini-Hochberg corrected P-values [78]).

541 ***Identifying polymorphic sites***

542 SNP calling was performed with Snape[37]. The program requires information of the prior
543 nucleotide diversity θ. Hence, the initial values of nucleotide diversity obtained with PoPoolation
544 were used. Snape was run with folded spectrum and prior type informative. As Snape requires the
545 MPILEUP format, the previously generated MPILEUP files were used. SNP calling was
546 separately performed on coding and non-coding parts of the genome. Therefore, each population
547 MPILEUP file was filtered by coding sequence position with the *filter-pileup-by-gtf.pl* script of
548 PoPoolation. For coding sequences the --keep-mode was set to retain all coding sequences. The
549 non-coding sequences were obtained by using the default option and thus discarding the coding
550 sequences from the MPILEUP file. Snape produces a posterior probability of segregation for
551 each position. The posterior probability of segregation was used to filter low-confidence SNPs
552 and indicated in the specific section.

553 ***Divergence and polymorphisms in 0-fold and 4-fold sites***

554 Polarization of synonymous sites (four-fold degenerated sites) and non-synonymous sites (zero-
555 fold degenerated sites) was done using the pseudogenomes of outgroups *Nothobranchius*
556 *orthonotus* and *Nothobranchius rachovii*. For each population the genomic information of the
557 respective pseudogenome was extracted with bedtools getfasta command [79,80] and the derived

558 allele frequency of every position was inferred with a custom R script. Briefly, only sites with the
559 bases A, G, T or C in the outgroup pseudogenome were included and checked whether the
560 position has an alternative allele in each of the investigated populations. Positions with an
561 alternative allele present in the population data were treated as possible divergent or polymorphic
562 sites. The derived frequency was determined as frequency of the allele not present in the
563 outgroup. Occasions with an alternate allele present in the population data were treated as
564 possible divergent or polymorphic sites. Divergent sites are positions in the genome were the
565 outgroup allele is different from the allele present in the population. Polymorphic sites are sites in
566 the genome that have more than one allele segregating in the population. Only biallelic
567 polymorphic sites were used in this analysis. The DAF was determined as frequency of the allele
568 not shared with the respective outgroups. In general, positions with only one supporting read for
569 an allele were treated as monomorphic sites. SNPs with a DAF < 5% or > 95% were treated as
570 fixed mutations. Further filtering was done based on the threshold of the posterior probability of
571 >0.9 calculated with Snape (see previous subsection), combined with a minimum and maximum
572 coverage threshold per population (GNP: 24, 94; NF414:19, 77; NF303: 18, 70).

573

574 ***Asymptotic McDonald-Kreitman α***

575 The rate of substitutions that were driven to fixation by positive selection was evaluated with an
576 improved method based on the McDonald-Kreitman test[81]. The test assumes that the proportion
577 of non-synonymous mutations that are neutral has the same fixation rate as synonymous
578 mutations. Therefore, under neutrality the ratio between non-synonymous to synonymous
579 substitutions (Dn/Ds) between species is equal to the ratio of non-synonymous to synonymous
580 polymorphisms within species (Pn/Ps). If positive selection takes place, the ratio between non-

581 synonymous to synonymous substitutions between species is larger than the ratio of non-
582 synonymous to synonymous polymorphism within species[81]. The concept behind this is that the
583 selected variant reaches fixation in a shorter time than by random drift. Therefore, the selected
584 variant increases Dn, not Pn. The proportion of non-synonymous substitutions that were fixed by
585 positive selection (α) was estimated with an extension of the McDonald-Kreitman test[82]. Due to
586 the presence of slightly deleterious mutations the estimate of α can be underestimated. For this
587 reason, the method used by Messer and Petrov was implemented to calculate α as a function of
588 the derived allele frequency x[36,83]. With this method the true value of α can be inferred as the
589 asymptote of the function of α. Additionally, the value of α(x) for low derived frequencies should
590 give an estimate of the number of slightly deleterious mutations that segregate in the population.

591 ***Direction of selection (DoS)***

592 To further investigate the signature of selection, the direction of selection (DoS) index for every
593 gene was calculated[44]. DoS standardizes α to a value between -1 and 1. A positive value of DoS
594 indicates adaptive evolution (positive selection) and a negative value indicates the segregation of
595 slightly deleterious alleles, therefore weaker purifying selection[44]. This ratio is undefined for
596 genes without any information about polymorphic or substituted sites. Therefore, only genes with
597 at least one polymorphic and one substituted site were included.

598 ***Inference of distribution of fitness effects***

599 The distribution of fitness effects (DFE) was inferred using the program polyDFE2.0[38]. For this
600 analysis the unfolded site frequency spectra (SFS) of non-synonymous (0-fold) and synonymous
601 sites (4-fold) were projected into 10 chromosomes for each population. Information about the
602 fixed derived sites was included in this analysis (using *Nothobranchius orthonotus*). PolyDFE2.0
603 estimates either the full DFE, containing deleterious, neutral and beneficial mutations, or only the

604 deleterious DFE. The best model for each population was obtained using a model testing
605 approach with three different models implemented in PolyDFE2.0 (Model A, B, C). Due to
606 possible biases from erroneous polarization or unknown demography, runs accounting for
607 polarization errors and demography (+eps, +r) were included. Initial parameters were
608 automatically estimated with the –e option, as recommended. To ensure that the parameter space
609 is explored thoroughly, the basin hopping option was applied with a maximum of 500 iterations
610 (-b). The best model for each population was chosen based on the Akaike Information Criterion
611 (AIC). Confidence intervals were generated by running 200 bootstrap datasets with the same
612 parameters used to infer the best model.
613

614 ***Variant annotation***

615 Classification of changes in the coding-sequence (CDS) was done with the variant annotator
616 SnpEFF [39]. The new genome of *Nothobranchius furzeri* (NFZ v2.0) was implemented to the
617 SnpEFF pipeline. Subsequently, a database for variant annotation with the genome NFZ v2.0
618 FASTA file and the annotation GTF file was generated. For variant annotation the population
619 specific synonymous and non-synonymous sites with a change in respect to the reference genome
620 NFZ v2.0 were used to infer the impact of these sites. The possible annotation impact classes
621 were low, moderate, and high. SNPs with a frequency below 5% or above 95% were excluded for
622 this analysis. To be consistent with the analysis of the distribution of fitness effects, only
623 positions also found to be present in the *N. orthonotus* pseudogenome were considered. Positions
624 with warnings in the variant annotation were removed.

625 **Consurf analysis**

626 The Consurf score was calculated accordingly to the method used in Cui *et al.* [6]. We used the
627 Consurf [40-43] package to assign each AA a conservation score based on the evolutionary rate in
628 homologs of other vertebrates. Consurf scores were estimated for 12575 genes of *N. furzeri* and
629 synonymous and non-synonymous genomic positions were matched with the derived allele
630 frequency of *N. orthonotus* and *N. rachovii*, respectively. The derived frequencies were binned in
631 five bins and we used pairwise Wilcoxon rank sum test to assess significance after correcting for
632 multiple testing (Benjamini & Hochberg adjustment) between each subsequent bin per population
633 and matching bins between populations.

634 ***Over-representation analysis***

635 Gene ontology (GO) and pathway overrepresentation analysis was performed with the online tool
636 ConsensusPathDB (http://cpdb.molgen.mpg.de;version34)[84] using "KEGG" and "REACTOME"
637 databases. Briefly, each gene present in the outlier list was provided with an ENSEMBL human
638 gene identifier[85], if available, and entered as the target list into the user interface. All genes
639 included in the analysis and with available human ENSEMBL identifier were used as the
640 background gene list. ConsensusPathDB maps the entries to the databases and calculates the
641 enrichment score for each entity by comparing the proportion of target genes in the entity over
642 the proportion of background genes in the entity. For each of the enrichment a P-value is
643 calculated based on a hyper geometric model and is corrected for multiple testing using the false
644 discovery rate (FDR). Only GO terms and pathways with more than two genes were included.
645 Overrepresentation analysis was performed on genes falling below the 2.5th percentile or above
646 the 97.5th percentile thresholds. The percentiles for either F_{ST} or DoS values were calculated with
647 the quantile() function in R.

648 **Statistical analysis and data processing**

649 Statistical analyses were performed using R studio version 1.0.136 (R version 3.3.2[86]) on a local
650 computer and R studio version 1.1.456 (R version 3.5.1) in a cluster environment at the Max-
651 Planck-Institute for Biology of Ageing (Cologne). Unless otherwise stated, the functions t.test()
652 and wilcox.test() in R have been used to evaluate statistical significance. To generate a pipeline
653 for data processing we used Snakemake[87]. Figure style was modified using Inkscape version
654 0.92.4. For circular visualization of genomic data we used Circos[64].

655 **Inference of demographic population history with individual resequencing data**

656 To infer the demographic history, we performed whole genome re-sequencing of single
657 individuals from all populations resulting in mean genome coverage between 13-21x (**Table S2**).
658 Demographic history was inferred from single individual sequencing data using Pairwise
659 Sequential Markovian Coalescence (PSMC' mode from MSMC2[25]). Re-sequencing of single
660 individuals was performed with the DNA of single individuals extracted for the pooled
661 sequencing for each examined population. The Illumina short-insert library was constructed
662 based on a published protocol[88]. Extracted DNA (500ng) was digested with fragmentase (New
663 England Biolabs) for 20min at 37°C, followed by end-repair and A-tailing (1.0µl NEB End-repair
664 buffer, 0.5µl Klenow fragment, 0.5µl Taq.Polymerase, 0.2µl T4 polynucleotide kinase, 10µl
665 reaction volume, 30min at 25°C, 30min at 75°C) and adapter ligation (NEB Quick ligase buffer
666 12.5µl, Quick ligase 0.5µl, 1µl adapter P1 (D50X), 1µl adapter P2 (universal), 5µM each; 20min
667 at 20°C, 25µl reaction volume). Next, ligation mix was diluted to 50µl and used 0.583:1 volume
668 of home-brewed SPRI beads (SPRI binding buffer: 2.5M NaCL, 20mM PEG 8000, 10mM Tric-
669 HCL, 1mM EDTA,oh=8, 1mL TE-washed SpeedMag beads GE Healthcare, 65152105050250
670 per 100mL buffer) for purification. The ligation products were amplified with 9 PCR cycles using
671 KAPA Hifi kit (Roche, P5 universal primer and P7 indexed primer D7XX). The samples were

672 pooled and sequenced on Hiseq X. Raw sequencing reads were trimmed using Trimmomatic-
673 0.32 (ILLUMINACLIP:illumina-adaptors.fa:3:7:7:1:true, LEADING:20, TRAILING:20,
674 SLIDINGWINDOW:4:20, MINLEN:50)[72]. Data files were inspected with FastQC v0.11.22.
675 Trimmed reads were subsequently mapped to the reference genome with BWA-MEM (version
676 0.7.12). The SAM output was converted into BAM format, sorted, and indexed via SAMTOOLS
677 v1.3.1[61]. Filtering and realignment was conducted with PICARD v1.119 and GATK[60]. Briefly,
678 the reads were relabeled, sorted, and indexed with AddOrReplaceReadGroups. Duplicated reads
679 were marked with the PICARD feature MarkDuplicates and reads were realigned with first
680 creating a target list with RealignerTargetCreator, second by IndelRealigner from the GATK
681 suite. Resulting reads were again sorted and indexed with SAMTOOLS. Next, the guidance for
682 PSMC' (https://github.com/stschiff/msmc/blob/master/guide.md) was followed; VCF-files and
683 masked files were generated with the *bamCaller.py* script (MSMC-tools package). This step
684 requires the chromosome coverage information to mask regions with too low or too high
685 coverage. As recommended in the guidelines, the average coverage per chromosome was
686 calculated using SAMTOOLS. In addition, this step was performed using a coverage threshold of
687 18 as recommended by Nadachowska-Brzyska *et al*.[89]. Final input data was generated using the
688 *generate_multihetsep.py* script (MSMC-tools package). Subsequently, for each sample PSMC'
689 was run independently. Bootstrapping was performed for 30 samples per individual and input
690 files were generated with the *multihetsep_bootstrap.py* script (MSMCtools package).

691 ***Analysis of differential expressed genes with age***

692 We downloaded the previously published RNaseq data from a longitudinal study of
693 *Nothobranchius furzeri*[18]. The data set contains five time points (5w, 12w, 20w, 27w, 39w) in
694 three different tissues (liver, brain, skin). The raw reads were mapped to the NFZ v2.0 reference

genome and subsequently counted using STAR (version 2.6.0.c)[90] and FeatureCounts (version 1.6.2)[91]. We performed statistical analysis of differential expression with age using DESeq2[92] and age as factor. Genes are classified as upregulated in young (log(FoldChange) < 0, adjusted p < 0.01), upregulated in old (log(FoldChange) > 0,adjusted p < 0.01).

700 ## Figures

701 **Figure 1. Demography and natural occurrence of turquoise killifish populations.** a) Inferred
702 ancestral effective population size (N_e) (using PSMC') on y-axis and past generations on x-axis in
703 GNP (red, orange), NF414 (black, grey) and NF303 (blue). Inset: unrooted neighbour joining tree
704 based on pairwise genetic differentiation (F_{ST}) values. b) Geographical locations of sampled
705 natural population of turquoise killifish (*Nothobranchius furzeri)*. The area of the coloured circles
706 represents the estimated effective population size (N_e) based on $\theta_{\text{Watterson}}$. c) Natural environment
707 of turquoise killifish and schematic of the annual life cycle. Figure 1 was partly made with
708 Biorender®.

709 **Figure 2. Genomic regions of high and low genetic divergence between pairs turquoise**
710 **killifish populations.** Left) Genomic regions with high or low genetic differentiation between
711 turquoise killifish populations identified with an F_{ST} outlier approach. Z-transformed F_{ST} values
712 of all pairwise comparison in solid lines, with "NF303vsNF414" in yellow, "NF303vsGNP" in
713 blue, and "NF414vsGNP" in green. The significance thresholds of upper and lower 5‰ are
714 shown in dotted lines with same colour coding. Center) Circos plot of Z-transformed F_{ST} values
715 between all pairwise comparisons with "NF303vsNF414" in the inner circle (yellow),
716 "NF414vsGNP" in the middle circle (green), and "NF303vsGNP" in the outer circle (blue).
717 Right) Pairwise genetic differentiation based on F_{ST} in the four main clusters associated with
718 lifespan (QTL from Valenzano *et al.*[15]).

719 **Figure 3. Synteny and sex chromosome evolution in turquoise killifish.** a) Synteny circos
720 plots based on 1-to-1 orthologous gene location between the new turquoise killifish assembly
721 (black chromosomes) and platyfish (*Xiphophorus maculatus*, coloured chromosomes, left circos
722 plot) and between the new turquoise killifish assembly (black chromosomes) and medaka

723 (*Oryzias latipes*, coloured chromosomes, right circos plot). Orthologous genes in concordant
724 order are visualized as one syntenic block. Synteny regions are connected via colour-coded
725 ribbons, based on their chromosomal location in platyfish or medaka. If the direction of the
726 syntenic sequence is inverted compared to the compared species, the ribbon is twisted. Outer data
727 plot shows –log(q-value) of survival quantitative trait loci (QTL, ordinate value between 0 and
728 3.5, every value above 3.5 is visualized at 3.5[15]) and the inner data plot shows –log(q-value) of
729 the sex QTL (ordinate value between 0 and 3.5, every value above 3.5 is visualized at 3.5). Boxes
730 between the two circos plots show genes within the peak regions of the four highest –log(q-value)
731 of survival QTL on independent chromosomes (red box) and the highest association to sex (black
732 box). b) High resolution synteny map between the sex-chromosome of the turquoise killifish
733 (Chr3) with platyfish chromosome 16 and 3 in the upper plot, and between the turquoise killifish
734 and medaka chromosome 8 and 16 (lower plot). The middle plot shows the QTLs for survival and
735 sex along the turquoise killifish sex chromosome. c) Model of sex chromosome evolution in the
736 turquoise killifish. A translocation event within one ancestral autosome led to the emergence of a
737 chromosomal region harbouring a new sex-determining-gene (SDG). The fusion of a second
738 autosome led to the formation of the current structure of the turquoise killifish sex chromosome.

739 **Figure 4. Genome-wide signatures of natural and relaxed selection in turquoise killifish**
740 **populations.** Asymptotic McDonald-Kreitman alpha (MK $\boldsymbol{\alpha}$) analysis based on derived
741 frequency bins using as outgroups a) *Nothobranchius orthonotus* and b) *Nothonbranchius*
742 *rachovii*. Population GNP is shown in red, NF414 in black, and NF303 in blue. c) Proportion of
743 non-synonymous SNPs binned in allele frequencies of non-reference (alternative) alleles for GNP
744 (red), NF414 (black) and NF303 (blue). d) Negative distribution of fitness effects of populations
745 GNP (red), NF414 (black) and NF303 (blue) with cumulative proportion of deleterious SNPs on
746 y-axis and the compound measure of $4N_es$ on x-axis. e) Proportion of different effect types of

747 SNPs in coding sequences of all populations. The effect on amino acid sequence for each genetic
748 variant is represented by colours (legend). Significance is based on ratio between synonymous
749 effects to non-synonymous effects (significance based on Chi-square test).

750 **Figure 5. Pathway enrichment in genes under adaptive and neutral evolution in turquoise**
751 **killifish populations.** a) Distribution of direction of selection (DoS) represented with median of
752 distribution for population GNP (red), NF414 (grey) and NF303 (blue). Left panel shows DoS
753 distribution computed using *Nothobranchius orthonotus* as outgroup and right panel shows DoS
754 distribution computed using *Nothobranchius rachovii* as outgroup. Significance based on
755 Wilcoxon-Rank-Sum test. b) Pathway over-representation analysis of genes below the 2.5% level
756 of gene-wise DoS values are shown with red background and above the 97.5% level of gene-wise
757 DoS values are shown with green background. Only pathway terms with significance level of
758 FDR corrected q-value < 0.05 are shown (in -log(q-value)). Terms enriched in population GNP
759 have red dots, enriched in population NF414 have black dots, and enriched in population NF303
760 have blue dots, respectively.

761 ## Acknowledgments

762 We would like to thank Patience and Edson Gandiwa for their administrative support, Tamuka
763 Nhiwatiwa for helping with logistics and samples handling; Evious Mpofu, Hugo and Elsabe van
764 der Westhuizen and all the rangers of the Gonarezhou National Park for their support in the field.
765 We are thankful to Zimbabwe National Parks for allowing our team to conduct research in the
766 Gonarezhou National Park; Itamar Harel, Matej Polacik and Radim Blazek for hands-on
767 contribution with the field work. We further thank all members of the Valenzano lab for their
768 continuous scientific input and support. The Czech Science Foundation provided financial
769 support to MR for sampling Mozambican populations (19-01789S). This project was funded by

770 the Max Planck Institute for Biology of Ageing, the Max Planck Society and the CECAD at the
771 University of Cologne.

772 ## References

773 1 Lanfear, R., Kokko, H. & Eyre-Walker, A. Population size and the rate of evolution. *Trends Ecol Evol*
774 **29**, 33-41, doi:10.1016/j.tree.2013.09.009 (2014).
775 2 Nonaka, E. *et al.* Scaling up the effects of inbreeding depression from individuals to
776 metapopulations. *J Anim Ecol* **88**, 1202-1214, doi:10.1111/1365-2656.13011 (2019).
777 3 Furness, A. I. The evolution of an annual life cycle in killifish: adaptation to ephemeral aquatic
778 environments through embryonic diapause. *Biol. Rev. Camb. Philos. Soc.* **91**, 796-812,
779 doi:10.1111/brv.12194 (2016).
780 4 Cellerino, A., Valenzano, D. R. & Reichard, M. From the bush to the bench: the annual
781 *Nothobranchius* fishes as a new model system in biology. *Biol Rev Camb Philos Soc* **91**, 511-533,
782 doi:10.1111/brv.12183 (2016).
783 5 Hu, C. K. & Brunet, A. The African turquoise killifish: A research organism to study vertebrate aging
784 and diapause. *Aging Cell* **17**, e12757, doi:10.1111/acel.12757 (2018).
785 6 Cui, R. *et al.* Relaxed Selection Limits Lifespan by Increasing Mutation Load. *Cell* **178**, 385-399 e320,
786 doi:10.1016/j.cell.2019.06.004 (2019).
787 7 Kim, Y., Nam, H. G. & Valenzano, D. R. The short-lived African turquoise killifish: an emerging
788 experimental model for ageing. *Dis Model Mech* **9**, 115-129, doi:10.1242/dmm.023226 (2016).
789 8 Blazek, R. *et al.* Repeated intraspecific divergence in life span and aging of African annual fishes
790 along an aridity gradient. *Evolution* **71**, 386-402, doi:10.1111/evo.13127 (2017).
791 9 Di Cicco, E., Tozzini, E. T., Rossi, G. & Cellerino, A. The short-lived annual fish *Nothobranchius furzeri*
792 shows a typical teleost aging process reinforced by high incidence of age-dependent neoplasias.
793 *Exp Gerontol* **46**, 249-256, doi:10.1016/j.exger.2010.10.011 (2011).
794 10 Wendler, S., Hartmann, N., Hoppe, B. & Englert, C. Age-dependent decline in fin regenerative
795 capacity in the short-lived fish *Nothobranchius furzeri*. *Aging Cell* **14**, 857-866,
796 doi:10.1111/acel.12367 (2015).
797 11 Ahuja, G. *et al.* Loss of genomic integrity induced by lysosphingolipid imbalance drives ageing in
798 the heart. *EMBO Rep* **20**, doi:10.15252/embr.201847407 (2019).
799 12 Valenzano, D. R., Terzibasi, E., Cattaneo, A., Domenici, L. & Cellerino, A. Temperature affects
800 longevity and age-related locomotor and cognitive decay in the short-lived fish *Nothobranchius*
801 *furzeri*. *Aging Cell* **5**, 275-278, doi:10.1111/j.1474-9726.2006.00212.x (2006).
802 13 Smith, P. *et al.* Regulation of life span by the gut microbiota in the short-lived African turquoise
803 killifish. *Elife* **6**, doi:10.7554/eLife.27014 (2017).
804 14 Terzibasi, E. *et al.* Large differences in aging phenotype between strains of the short-lived annual
805 fish Nothobranchius furzeri. *PLoS One* **3**, e3866, doi:10.1371/journal.pone.0003866 (2008).
806 15 Valenzano, D. R. *et al.* The African Turquoise Killifish Genome Provides Insights into Evolution and
807 Genetic Architecture of Lifespan. *Cell* **163**, 1539-1554, doi:10.1016/j.cell.2015.11.008 (2015).
808 16 Vrtilek, M., Zak, J., Polacik, M., Blazek, R. & Reichard, M. Longitudinal demographic study of wild
809 populations of African annual killifish. *Sci Rep* **8**, 4774, doi:10.1038/s41598-018-22878-6 (2018).
810 17 Kirschner, J. *et al.* Mapping of quantitative trait loci controlling lifespan in the short-lived fish
811 *Nothobranchius furzeri*--a new vertebrate model for age research. *Aging Cell* **11**, 252-261,
812 doi:10.1111/j.1474-9726.2011.00780.x (2012).
813 18 Reichwald, K. *et al.* Insights into Sex Chromosome Evolution and Aging from the Genome of a
814 Short-Lived Fish. *Cell* **163**, 1527-1538, doi:10.1016/j.cell.2015.10.071 (2015).
815 19 Reichwald, K. *et al.* High tandem repeat content in the genome of the short-lived annual fish
816 *Nothobranchius furzeri*: a new vertebrate model for aging research. *Genome Biol* **10**, R16,
817 doi:10.1186/gb-2009-10-2-r16 (2009).

20 Dorn, A. *et al.* Phylogeny, genetic variability and colour polymorphism of an emerging animal model: the short-lived annual *Nothobranchius* fishes from southern Mozambique. *Mol Phylogenet Evol* **61**, 739-749, doi:10.1016/j.ympev.2011.06.010 (2011).

21 Bartakova, V. *et al.* Strong population genetic structuring in an annual fish, *Nothobranchius furzeri*, suggests multiple savannah refugia in southern Mozambique. *BMC Evol Biol* **13**, 196, doi:10.1186/1471-2148-13-196 (2013).

22 Bartáková, V., Reichard, M., Blažek, R., Polačik, M. & Bryja, J. Terrestrial fishes: rivers are barriers to gene flow in annual fishes from the African savanna. *Journal of Biogeography* **42**, 1832-1844, doi:10.1111/jbi.12567 (2015).

23 Kofler, R. *et al.* PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* **6**, e15925, doi:10.1371/journal.pone.0015925 (2011).

24 Kofler, R., Pandey, R. V. & Schlotterer, C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435-3436, doi:10.1093/bioinformatics/btr589 (2011).

25 Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**, 919-925, doi:10.1038/ng.3015 (2014).

26 Baumgart, M. *et al.* A miRNA catalogue and ncRNA annotation of the short-living fish *Nothobranchius furzeri*. *BMC Genomics* **18**, 693, doi:10.1186/s12864-017-3951-8 (2017).

27 Ferdinandusse, S. *et al.* HIBCH mutations can cause Leigh-like disease with combined deficiency of multiple mitochondrial respiratory chain enzymes and pyruvate dehydrogenase. *Orphanet J Rare Dis* **8**, 188, doi:10.1186/1750-1172-8-188 (2013).

28 Huff, M. W. & Telford, D. E. Lord of the rings--the mechanism for oxidosqualene:lanosterol cyclase becomes crystal clear. *Trends Pharmacol Sci* **26**, 335-340, doi:10.1016/j.tips.2005.05.004 (2005).

29 Osanai, T. *et al.* Novel anti-aging gene NM_026333 contributes to proton-induced aging via NCX1-pathway. *J Mol Cell Cardiol* **125**, 174-184, doi:10.1016/j.yjmcc.2018.10.021 (2018).

30 Orlov, S. V. *et al.* Novel repressor of the human FMR1 gene - identification of p56 human (GCC)(n)-binding protein as a Kruppel-like transcription factor ZF5. *FEBS J* **274**, 4848-4862, doi:10.1111/j.1742-4658.2007.06006.x (2007).

31 Nojima, H. *et al.* Syntabulin, a motor protein linker, controls dorsal determination. *Development* **137**, 923-933, doi:10.1242/dev.046425 (2010).

32 Ulvila, J. *et al.* Cofilin regulator 14-3-3zeta is an evolutionarily conserved protein required for phagocytosis and microbial resistance. *J Leukoc Biol* **89**, 649-659, doi:10.1189/jlb.0410195 (2011).

33 Mauxion, F., Preve, B. & Seraphin, B. C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. *RNA Biol* **10**, 267-276, doi:10.4161/rna.23065 (2013).

34 Kell, M. J. *et al.* Targeted deletion of the zebrafish actin-bundling protein L-plastin (lcp1). *PLoS One* **13**, e0190353, doi:10.1371/journal.pone.0190353 (2018).

35 Valenzano, D. R. *et al.* Mapping loci associated with tail color and sex determination in the short-lived fish *Nothobranchius furzeri*. *Genetics* **183**, 1385-1395, doi:10.1534/genetics.109.108670 (2009).

36 Messer, P. W. & Petrov, D. A. Frequent adaptation and the McDonald-Kreitman test. *Proc Natl Acad Sci U S A* **110**, 8615-8620, doi:10.1073/pnas.1220835110 (2013).

37 Raineri, E. *et al.* SNP calling by sequencing pooled samples. *BMC Bioinformatics* **13**, 239, doi:10.1186/1471-2105-13-239 (2012).

38 Tataru, P., Mollion, M., Glemin, S. & Bataillon, T. Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data. *Genetics* **207**, 1103-1119, doi:10.1534/genetics.117.300323 (2017).

39 Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92, doi:10.4161/fly.19695 (2012).

40 Pupko, T., Bell, R. E., Mayrose, I., Glaser, F. & Ben-Tal, N. 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**, S71-77, doi:10.1093/bioinformatics/18.suppl_1.s71 (2002).

41 Mayrose, I., Graur, D., Ben-Tal, N. & Pupko, T. Comparison of site-specific rate-inference methods for protein sequences: Empirical Bayesian methods are superior. *Molecular Biology and Evolution* **21**, 1781-1791, doi:10.1093/molbev/msh194 (2004).

42 Glaser, F. *et al.* ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163-164, doi:10.1093/bioinformatics/19.1.163 (2003).

43 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-350, doi:10.1093/nar/gkw408 (2016).

44 Stoletzki, N. & Eyre-Walker, A. Estimation of the neutrality index. *Mol Biol Evol* **28**, 63-70, doi:10.1093/molbev/msq249 (2011).

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

46 Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res* **27**, 757-767, doi:10.1101/gr.214874.116 (2017).

47 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).

48 Bao, E. & Lan, L. HALC: High throughput algorithm for long read error correction. *BMC Bioinformatics* **18**, 204, doi:10.1186/s12859-017-1610-3 (2017).

49 Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* **108**, 1513-1518, doi:10.1073/pnas.1017351108 (2011).

50 Coombe, L. *et al.* ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinformatics* **19**, 234, doi:10.1186/s12859-018-2243-x (2018).

51 Warren, R. L. *et al.* LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience* **4**, 35, doi:10.1186/s13742-015-0076-3 (2015).

52 Tang, H. *et al.* ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol* **16**, 3, doi:10.1186/s13059-014-0573-1 (2015).

53 Kosugi, S., Hirakawa, H. & Tabata, S. GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. *Bioinformatics* **31**, 3733-3741, doi:10.1093/bioinformatics/btv465 (2015).

54 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).

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

56 Sayers, E. W. *et al.* GenBank. *Nucleic Acids Res* **47**, D94-D99, doi:10.1093/nar/gky989 (2019).

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

58 Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv preprint*, doi:eprint arXiv:1303.3997 (2013).

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

914 60 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
915 generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).
916 61 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079,
917 doi:10.1093/bioinformatics/btp352 (2009).
918 62 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool.
919 *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
920 63 Ballesteros, J. A. & Hormiga, G. A New Orthology Assessment Method for Phylogenomic Data:
921 Unrooted Phylogenetic Orthology. *Mol Biol Evol* **33**, 2481, doi:10.1093/molbev/msw153 (2016).
922 64 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**,
923 1639-1645, doi:10.1101/gr.092759.109 (2009).
924 65 Guy, L., Kultima, J. R. & Andersson, S. G. genoPlotR: comparative gene and genome visualization
925 in R. *Bioinformatics* **26**, 2334-2335, doi:10.1093/bioinformatics/btq413 (2010).
926 66 Catchen, J. M., Conery, J. S. & Postlethwait, J. H. Automated identification of conserved synteny
927 after whole-genome duplication. *Genome Res* **19**, 1497-1505, doi:10.1101/gr.090480.108 (2009).
928 67 Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Koppen-Geiger climate
929 classification. *Hydrol Earth Syst Sc* **11**, 1633-1644, doi:DOI 10.5194/hess-11-1633-2007 (2007).
930 68 Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land
931 areas. *Int. J. Climatol.* **37**, 4302-4315, doi:10.1002/joc.5086 (2017).
932 69 Trabucco, A. & Zomer, R. Global Aridity Index and Potential Evapotranspiration Climate Database
933 v2. *CGIAR Consortium for Spatial Information* **available at:**
934 **https://cgiarcsi.community/2019/01/24/global-aridity-index-andpotential-**
935 **evapotranspiration-climate-database-v2/** (2019).
936 70 Neteler, M., Bowman, M. H., Landa, M. & Metz, M. GRASS GIS: A multi-purpose open source GIS.
937 *Environ Modell Softw* **31**, 124-130, doi:10.1016/j.envsoft.2011.11.014 (2012).
938 71 Lehner, B., Verdin, K. & A., J. HydroSHEDS Technical Documentation. *World Wildlife Fund US,*
939 *Washington* **available at http://hydrosheds.cr.usgs.gov** (2006).
940 72 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
941 *Bioinformatics* **30**, 2114-2120 (2014).
942 73 Nei, M. & Li, W. H. Mathematical-Model for Studying Genetic-Variation in Terms of Restriction
943 Endonucleases. *P Natl Acad Sci USA* **76**, 5269-5273, doi:DOI 10.1073/pnas.76.10.5269 (1979).
944 74 Watterson, G. A. Number of Segregating Sites in Genetic Models without Recombination.
945 *Theoretical Population Biology* **7**, 256-276, doi:Doi 10.1016/0040-5809(75)90020-9 (1975).
946 75 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language.
947 *Bioinformatics* **20**, 289-290, doi:10.1093/bioinformatics/btg412 (2004).
948 76 Pruisscher, P., Nylin, S., Gotthard, K. & Wheat, C. W. Genetic variation underlying local adaptation
949 of diapause induction along a cline in a butterfly. *Mol Ecol*, doi:10.1111/mec.14829 (2018).
950 77 Guo, B., Li, Z. & Merila, J. Population genomic evidence for adaptive differentiation in the Baltic
951 Sea herring. *Mol Ecol* **25**, 2833-2852, doi:10.1111/mec.13657 (2016).
952 78 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful
953 Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300 (1995).
954 79 Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc*
955 *Bioinformatics* **47**, 11 12 11-34, doi:10.1002/0471250953.bi1112s47 (2014).
956 80 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.
957 *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
958 81 McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*
959 **351**, 652-654, doi:10.1038/351652a0 (1991).
960 82 Smith, N. G. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022-1024,
961 doi:10.1038/4151022a (2002).

962 83 Haller, B. C. & Messer, P. W. asymptoticMK: A Web-Based Tool for the Asymptotic McDonald-
963 Kreitman Test. *G3 (Bethesda)* **7**, 1569-1575, doi:10.1534/g3.117.039693 (2017).
964 84 Herwig, R., Hardt, C., Lienhard, M. & Kamburov, A. Analyzing and interpreting genome data at the
965 network level with ConsensusPathDB. *Nat Protoc* **11**, 1889-1907, doi:10.1038/nprot.2016.117
966 (2016).
967 85 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res* **46**, D754-D761, doi:10.1093/nar/gkx1098
968 (2018).
969 86 RStudio: Integrated Development for R (RStudio, Inc. , Boston, MA, 2015).
970 87 Koster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics*
971 **28**, 2520-2522, doi:10.1093/bioinformatics/bts480 (2012).
972 88 Rowan, B. A., Patel, V., Weigel, D. & Schneeberger, K. Rapid and inexpensive whole-genome
973 genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3*
974 *(Bethesda)* **5**, 385-398, doi:10.1534/g3.114.016501 (2015).
975 89 Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. PSMC analysis of effective population
976 sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Mol Ecol* **25**,
977 1058-1072, doi:10.1111/mec.13540 (2016).
978 90 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
979 91 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning
980 sequence reads to genomic features. *Bioinformatics* **30**, 923-930,
981 doi:10.1093/bioinformatics/btt656 (2014).
982 92 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
983 seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

984

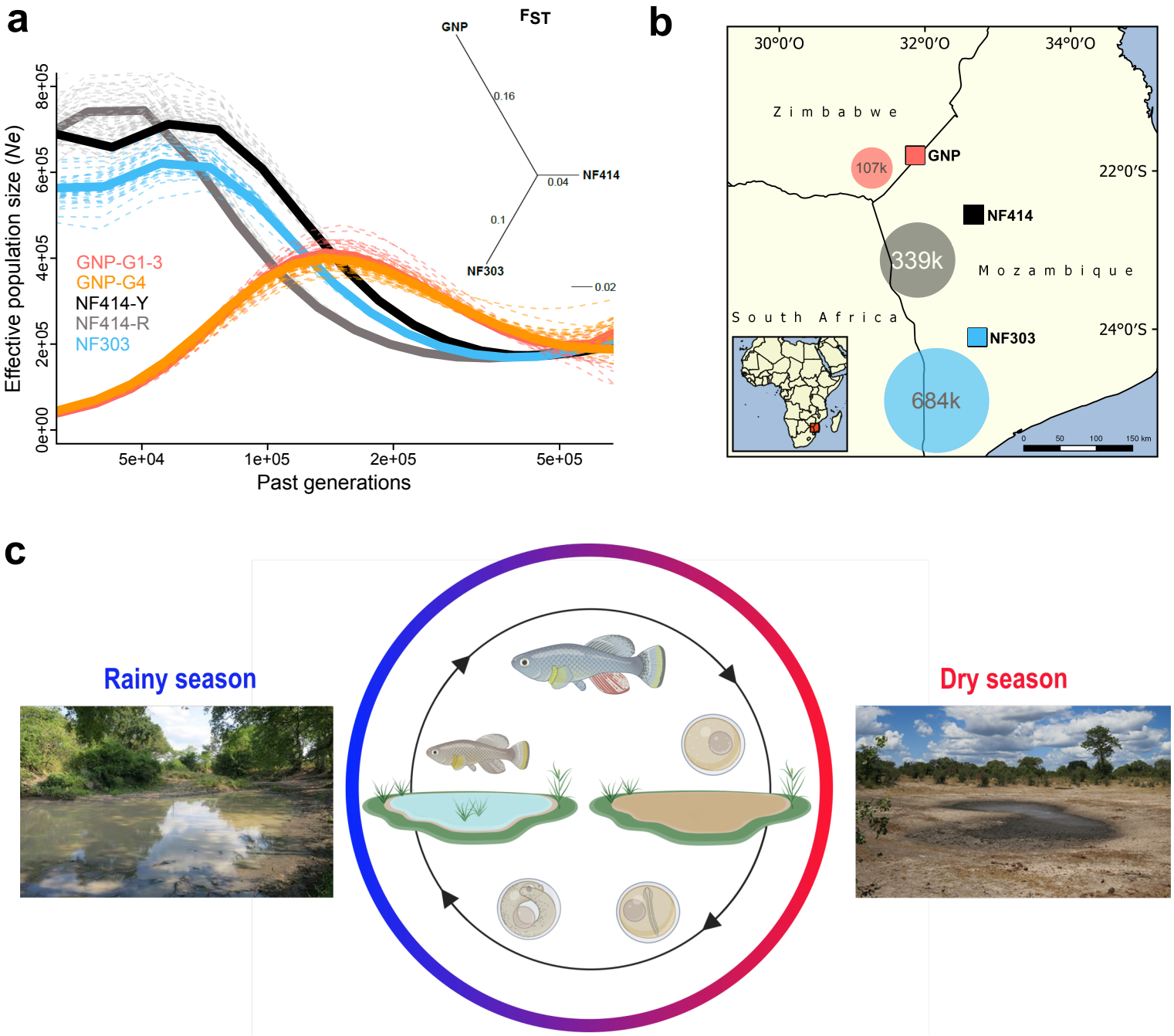

Figure 1. Demography and natural occurence of turquoise killifish populations. a) Inferred ancestral effective population size (N_e) (using PSMC') on y-axis and past generations on x-axis in GNP (red, orange), NF414 (black, grey) and NF303 (blue). Inset: unrooted neighbour joining tree based on pairwise genetic differentiation (F_{ST}) values. b) Geographical locations of sampled natural population of turquoise killifish (*Nothobranchius furzeri*). The area of coloured circles represent the estimated effective population size (N_e) based on $\theta_{Watterson}$. c) Natural environment of turquoise killifish and schematic of the annual life cycle. Figure 1 partly made with Biorender®.

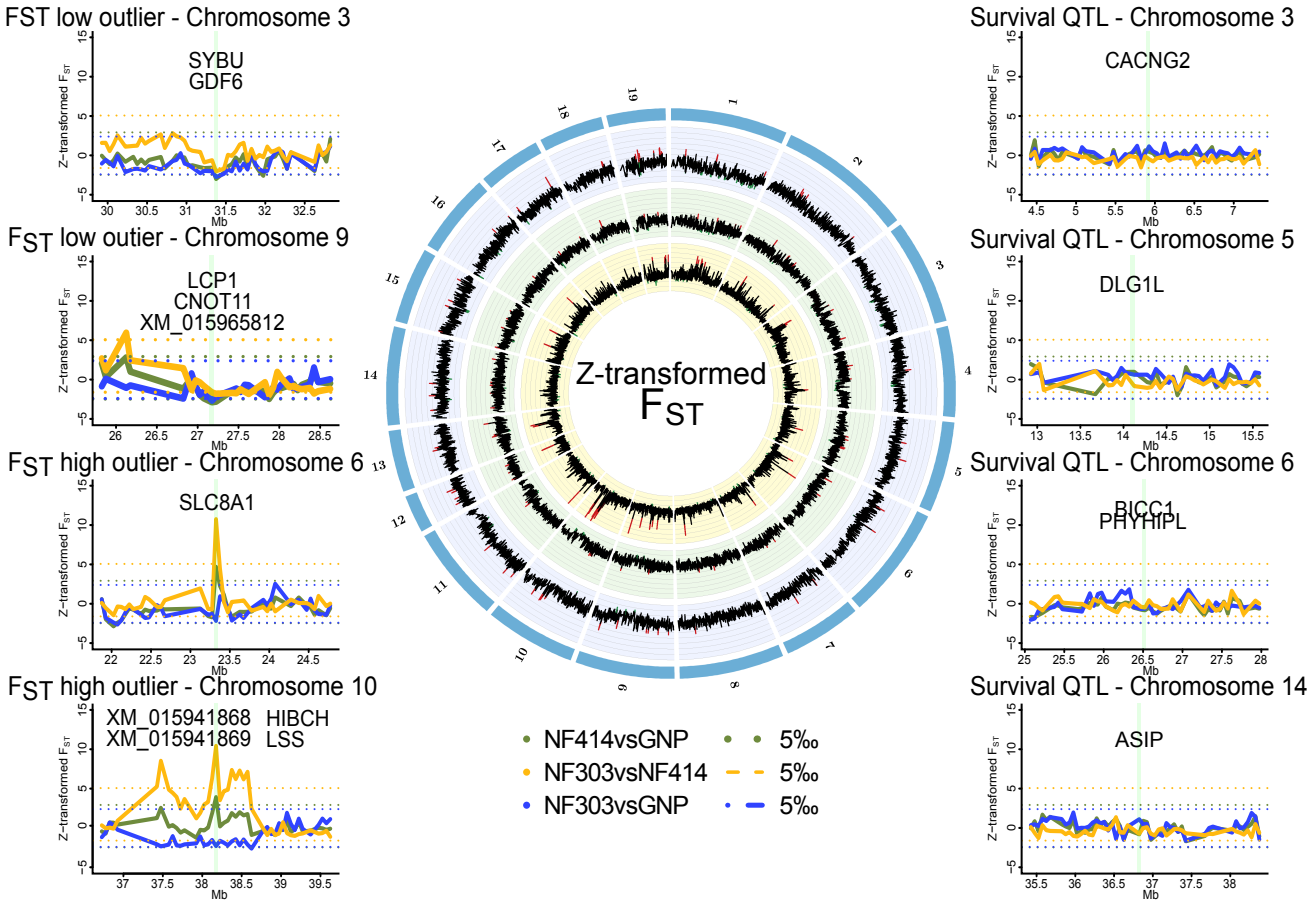

Figure 2. Genomic regions of high and low genetic divergence between pairs of turquoise killifish populations. Left) Genomic regions with high or low genetic differentiation between turquoise killifish populations identified with an F_{ST} outlier approach. Z-transformed F_{ST} values of all pairwise comparison in solid lines, with "NF303vsNF414" in yellow, "NF303vsGNP" in blue, and "NF414vsGNP" in green. The significance thresholds of upper and lower 5‰ are shown in spotted lines with same colour coding. Center) Circos plot of Z-transformed F_{ST} values between all pairwise comparisons with "NF303vsNF414" in the inner circle (yellow), "NF414vsGNP" in the middle circle (green), and "NF303vsGNP" in the outer circle (blue). Right) Pairwise genetic differentiation based on FST in the four main clusters associated with lifespan (QTL from Valenzano et al.[13]).

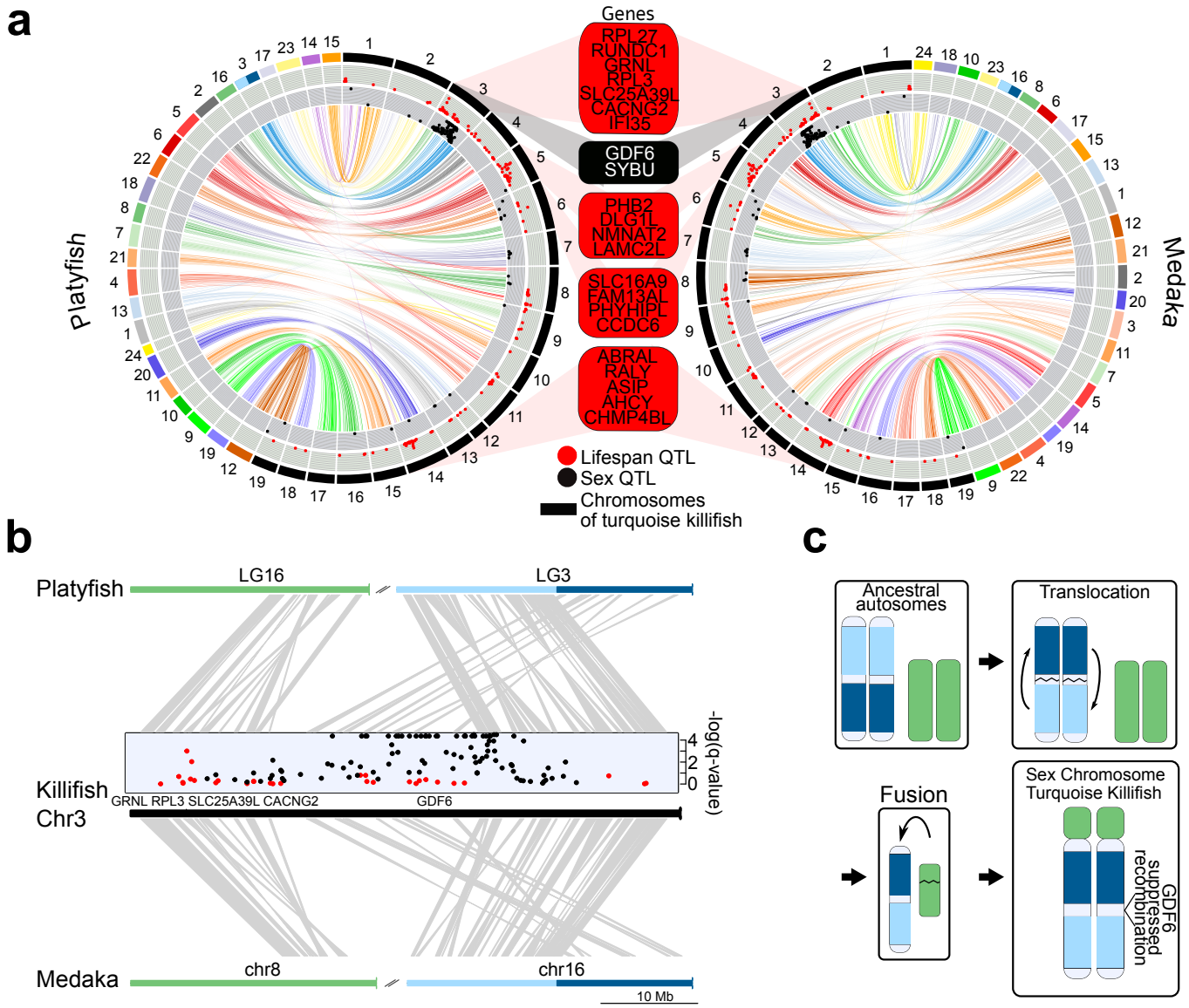

Figure 3. Synteny and sex chromosome evolution in turquoise killifish. a) Synteny circos plots based on 1-to-1 orthologous gene location between the new turquoise killifish assembly (black chromosomes) and platyfish (*Xiphophorus maculatus*, coloured chromosomes, left circos plot) and between the new turquoise killifish assembly (black chromosomes) and medaka (*Oryzias latipes*, coloured chromosomes, right circos plot). Orthologous genes in concordant order are visualized as one syntenic block. Synteny regions are connected via colour-coded ribbons, based on their chromosomal location in platyfish or medaka. If the direction of the syntenic sequence is inverted compared to the compared species, the ribbon is twisted. Outer data plot shows –log(q-value) of survival quantitative trait loci (QTL, ordinate value between 0 and 3.5, every value above 3.5 is visualized at 3.5[13]) and the inner data plot shows –log(q-value) of the sex QTL (ordinate value between 0 and 3.5, every value above 3.5 is visualized at 3.5). Boxes between the two circos plots show genes within the peak regions of the four highest –log(q-value) of survival QTL on independent chromosomes (red box) and the highest association to sex (black box) . b) High resolution synteny map between the sex-chromosome of the turquoise killifish (Chr3) with platyfish chromosome 16 and 3 in the upper plot, and between the turquoise killifish and medaka chromosome 8 and 16 (lower plot). The middle plot shows the QTLs for survival and sex along the turquoise killifish sex chromosome. c) Model of sex chromosome evolution in the turquoise killifish. A translocation event within one ancestral autosome led to the emergence of a chromosomal region harbouring a new sex-determining-gene (SDG). The fusion of a second autosome led to the formation of the current structure of the turquoise killifish sex chromosome.

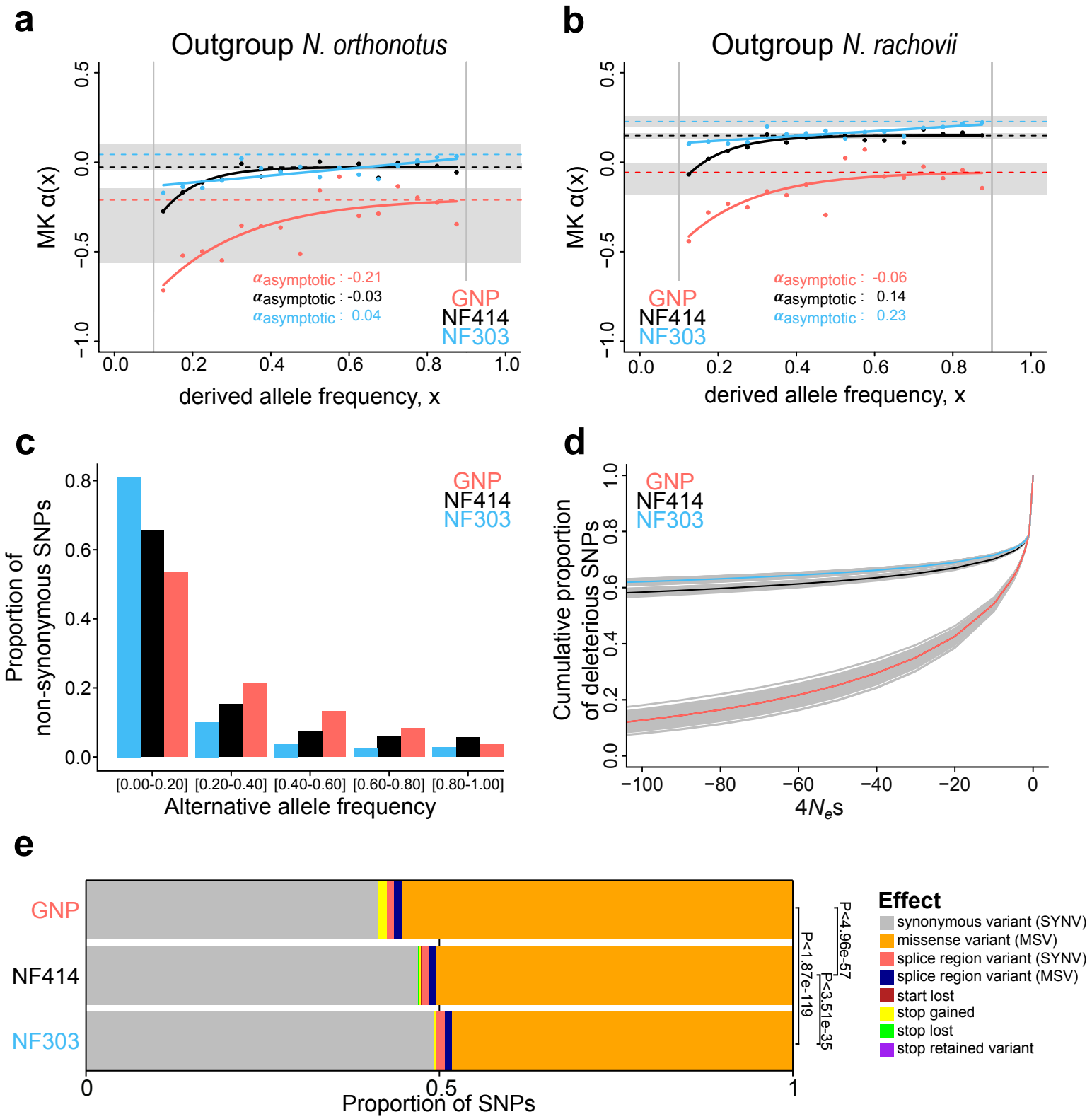

Figure 4. Genome-wide signatures of natural and relaxed selection in turquoise killifish. Asymptotic McDonald-Kreitman alpha (MK $\boldsymbol{\alpha}$) analysis based on derived frequency bins using as outgroups a) *Nothobranchius orthonotus* and b) *Nothonbranchius rachovii*. Population GNP is shown in red, NF414 in black, and NF303 in blue. c) Proportion of non-synonymous SNPs binned in allele frequencies of non-reference (alternative) alleles for GNP (red), NF414 (black) and NF303 (blue). d) Negative distribution of fitness effects of populations GNP (red), NF414 (black) and NF303 (blue) with cumulative proportion of deleterious SNPs on y-axis and the compound measure of $4N_es$ on x-axis. e) Proportion of different effect types of SNPs in coding sequences of all populations. The effect on amino acid sequence for each genetic variant is represented by colours (legend). Significance is based on ratio between synonymous effects to non-synonymous effects (significance based on Chi-square test).

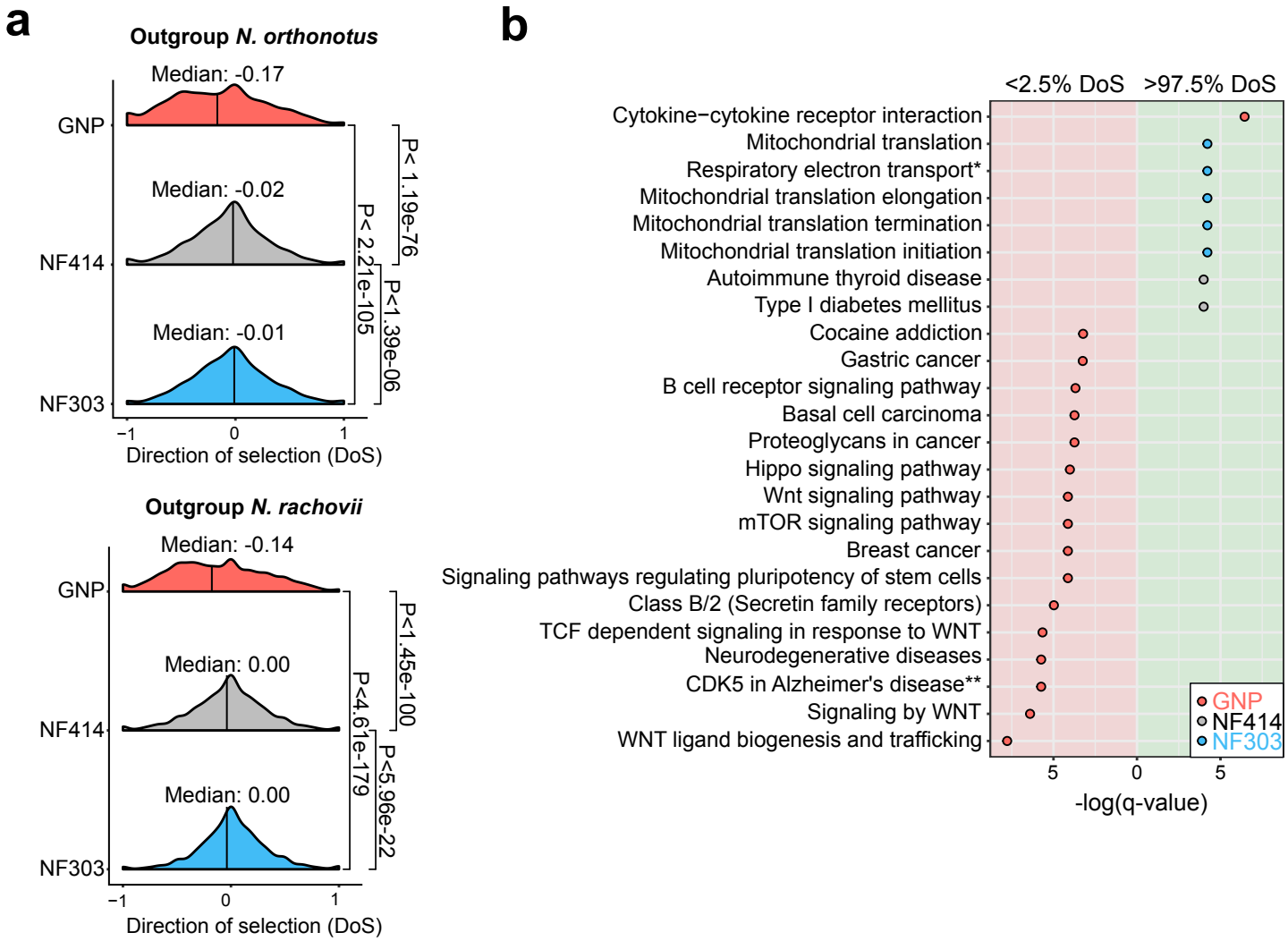

Figure 5. Pathway enrichment in genes under adaptive and neutral evolution in turquoise killifish populations. a) Distribution of direction of selection (DoS) represented with median of distribution for population GNP (red), NF414 (grey) and NF303 (blue). Left panel shows DoS distribution computed using *Nothobranchius orthonotus* as outgroup and right panel shows DoS distribution computed using *Nothobranchius rachovii* as outgroup. Significance based on Wilcoxon-Rank-Sum test. b) Pathway over-representation analysis of genes below the 2.5% level of gene-wise DoS values are shown with red background and above the 97.5% level of gene-wise DoS values are shown with green background. Only pathway terms with significance level of FDR corrected q-value < 0.05 are shown (in -log(q-value)). Terms enriched in population GNP have red dots, enriched in population NF414 have black dots, and enriched in population NF303 have blue dots, respectively. *ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins. ** Deregulated CDK5 triggers multiple neurodegenerative pathways in Alzheimer's disease models.

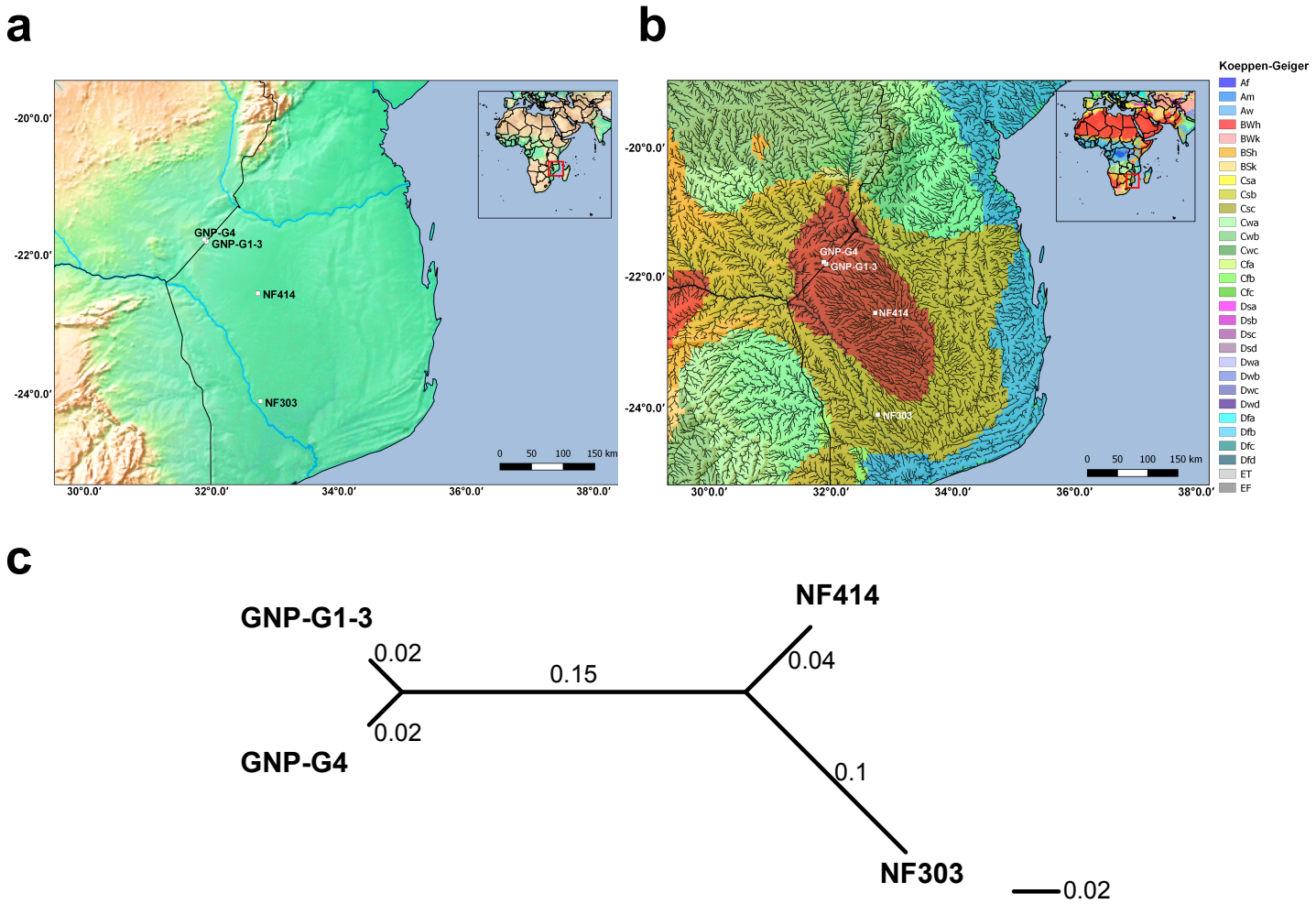

Figure S1. Altitude, climate classification and genetic differentiation of studied samples. a) Altitude elevation map with studied samples. b) Climate classification based on Koeppen-Geiger index combined with a high resolution river map. c) Unrooted neighbor joining tree based on pairwise genetic differentiation (F_{ST}) between all sample sites.

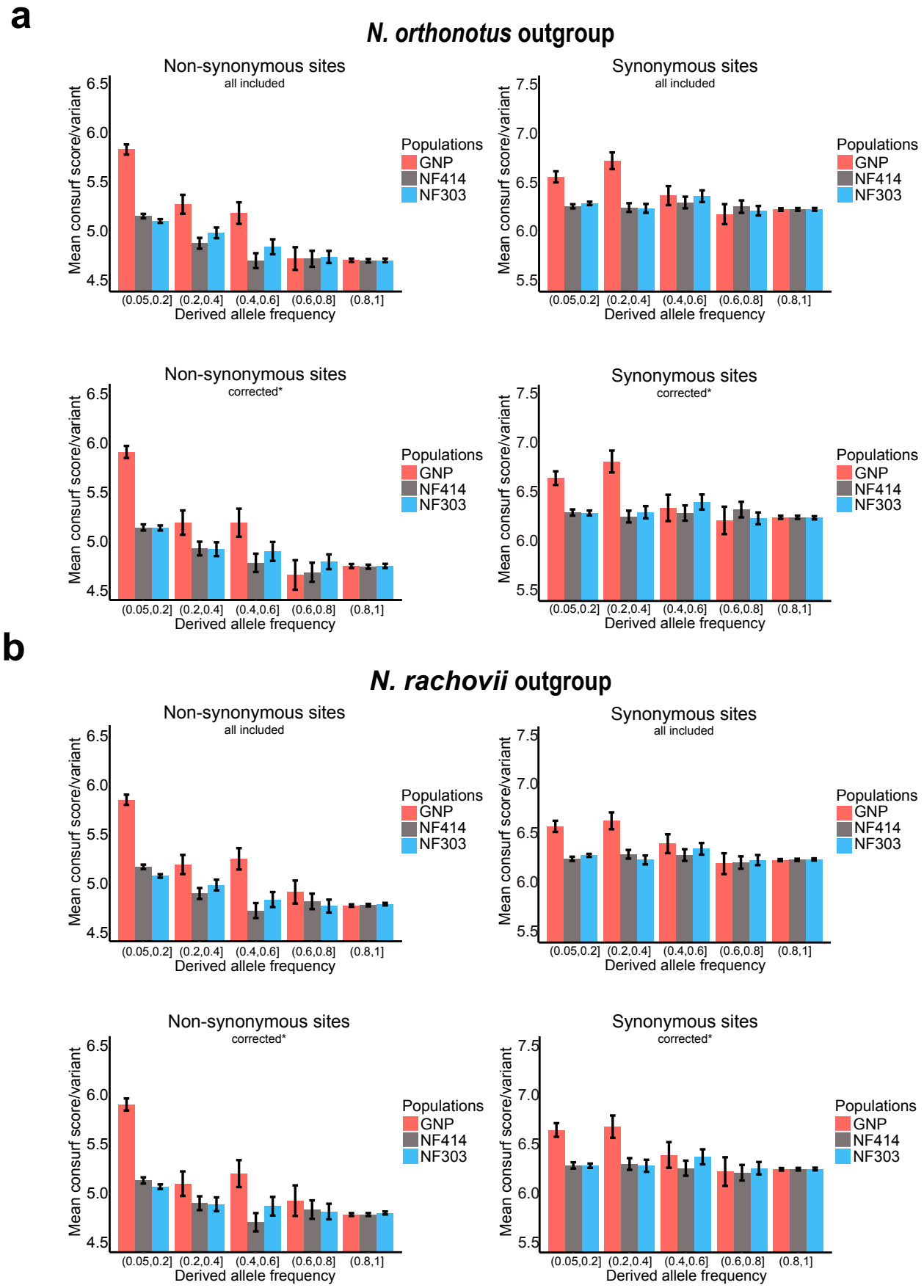

Figure S2. Mean Consurf score per variant based on derived frequency bins. a) Mean Consurf score based on derived frequency bins in non-synonymous (left) and synonymous (right) sites using *Nothobranchius orthonotus* as outgroup, including all available sites (upper panel) or only sites corrected for CMD, CpG hypermutation and highly detrimental effect based on SnpEFF analysis (lower panel). **b)** Mean Consurf score based on derived frequency bins in non-synonymous (left) and synonymous (right) sites using *Nothobranchius rachovii* as outgroup, including all available sites (upper panel) or only sites corrected for CMD, CpG hypermutation and highly detrimental effect based on SnpEFF analysis (lower panel). Mean consurf scores per variant are shown with SEM.