



HAL
open science

Age-specific survivorship and fecundity shape genetic diversity in marine fishes

Pierre Barry, Thomas Broquet, Pierre-Alexandre Gagnaire

► **To cite this version:**

Pierre Barry, Thomas Broquet, Pierre-Alexandre Gagnaire. Age-specific survivorship and fecundity shape genetic diversity in marine fishes. 2021. hal-03443785v1

HAL Id: hal-03443785

<https://hal.science/hal-03443785v1>

Preprint submitted on 23 Nov 2021 (v1), last revised 30 Aug 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Age-specific survivorship and fecundity shape genetic diversity in marine fishes

Pierre Barry¹, Thomas Broquet², Pierre-Alexandre Gagnaire¹

¹ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.

²CNRS & Sorbonne Université, UMR 7144, Station Biologique de Roscoff, 29680 Roscoff, France.

Abstract

1

2 Genetic diversity varies among species due to a range of eco-evolutionary processes that
3 are not fully understood. The neutral theory predicts that the amount of variation in the
4 genome sequence between different individuals of the same species should increase with its
5 effective population size (N_e). In real populations, multiple factors that modulate the variance
6 in reproductive success among individuals cause N_e to differ from the total number of individuals
7 (N). Among these, age-specific mortality and fecundity rates are known to have a direct impact
8 on the $\frac{N_e}{N}$ ratio. However, the extent to which vital rates account for differences in genetic
9 diversity among species remains unknown. Here, we addressed this question by comparing
10 genome-wide genetic diversity across 16 marine fish species with similar geographic distributions
11 but contrasted lifespan and age-specific survivorship and fecundity curves. We sequenced the
12 whole genome of 300 individuals to high coverage and assessed their genome-wide heterozygosity
13 with a reference-free approach. Genetic diversity varied from 0.2 to 1.4% among species, and
14 showed a negative correlation with adult lifespan, with a large negative effect ($slope = -0.089$
15 per additional year of lifespan) that was further increased when brooding species providing
16 intense parental care were removed from the dataset ($slope = -0.129$ per additional year of
17 lifespan). Using published vital rates for each species, we showed that the $\frac{N_e}{N}$ ratio resulting
18 simply from life tables parameters can predict the observed differences in genetic diversity
19 among species. Using simulations, we further found that the extent of reduction in $\frac{N_e}{N}$ with
20 increasing adult lifespan is particularly strong under Type III survivorship curves (high juvenile
21 and low adult mortality) and increasing fecundity with age, a typical characteristic of marine
22 fishes. Our study highlights the importance of vital rates as key determinants of species genetic
23 diversity levels in nature.

24 **Key words:** genetic diversity, life tables, adult lifespan, variance in reproductive success,
25 marine fishes

Author Summary

27 Understanding how and why genetic diversity varies across species has important implica-
28 tions for evolutionary and conservation biology. Although genomics has vastly improved our
29 ability to document intraspecific DNA sequence variation at the genome level, the range and
30 determinants of genetic diversity remain partially understood. At a broad taxonomic scale in
31 eukaryotes, the main determinants of diversity are reproductive strategies distributed along a
32 trade-off between the quantity and the size of offspring, which likely affect the long-term effective
33 population size. Long-lived species also tend to show lower genetic diversity, a result which
34 has however not been reported by comparative studies of genetic diversity at lower taxonomic
35 scales. Here, we compared genetic diversity across 16 European marine fish species showing
36 marked differences in longevity. Adult lifespan was the best predictor of genetic diversity, with
37 genome-wide average heterozygosity ranging from 0.2% in the black anglerfish (*L. budegassa*)
38 to 1.4% in the European pilchard (*S. pilchardus*). Using life tables summarizing age-specific
39 mortality and fecundity rates for each species, we showed that the variance in lifetime reproduc-
40 tive success resulting from age structure, iteroparity and overlapping generations can predict
41 the range of observed differences in genetic diversity among marine fish species. We then used
42 computer simulations to explore how combinations of vital rates characterizing different life
43 histories affect the relationship between adult lifespan and genetic diversity. We found that
44 marine fishes that display high juvenile but low adult mortality, and increasing fecundity with
45 age, are typically expected to show reduced genetic diversity with increased adult lifespan.
46 However, the impact of adult lifespan vanished using bird and mammal-like vital rates. Our
47 study shows that variance in lifetime reproductive success can have a major impact on species
48 genetic diversity and explains why this effect varies widely across taxonomic groups.

49 Introduction

50 Genetic diversity, the substrate for evolutionary change, is a key parameter for species adapt-
51 ability and vulnerability in conservation and management strategies (Frankham, 1995; Lande,
52 1995; DeWoody et al., 2021). Understanding the determinants of species' genetic diversity
53 has been, however, a long-standing puzzle in evolutionary biology (Lewontin, 1974). Advances
54 in DNA sequencing technologies have allowed to describe the range of genetic diversity levels
55 across eukaryote species and identify the main evolutionary processes governing that variation
56 (Leffler et al., 2012; Romiguier et al., 2014). Yet, the extent and reasons for which life his-
57 tory traits, and in particular reproductive strategies, influence genetic diversity remain to be
58 clarified (Ellegren and Galtier, 2016).

59 The neutral theory provides a quantitative prediction for the amount of genetic variation
60 at neutral sites (Kimura, 1983). Assuming equilibrium between the introduction of new vari-
61 ants by mutations occurring at rate μ , and their removal by genetic drift at a rate inversely
62 proportional to the effective population size N_e , the amount of genetic diversity (θ) of a stable
63 randomly mating population is equal to $4N_e\mu$ (Kimura and Crow, 1964). This quantity should
64 basically determine the mean genome-wide heterozygosity expected at neutral sites for any
65 given individual in that population. However, since the neutral mutation-drift balance can be
66 slow to achieve, contemporary genetic diversity often keeps the signature of past demographic
67 fluctuations rather than being entirely determined by the current population size. Therefore,
68 genetic diversity should be well predicted by estimates of N_e that integrate the long-term effect
69 of drift over the coalescent time. Unfortunately, such estimates are very difficult to produce
70 using demographic data only.

71 Demographic variations set aside, the most proximate determinant of N_e is the actual num-
72 ber of individuals (N), also called the census population size. Comparative genomic studies in
73 mammals and birds have showed that current species abundance correlates with the long-term
74 coalescence N_e , despite a potential deviation from long-term population stability in several of
75 the species studied (Díez-Del-Molino et al., 2018; Leroy et al., 2020; Peart et al., 2020). Gen-
76 eral laws in ecology, such as the negative relationship between species abundance and body
77 size (White et al., 2007) have also been used to predict the long-term N_e . Higher genetic di-
78 versity in small body size species was found in butterflies and Darwin's finches (Mackintosh
79 et al., 2019; Brüniche-Olsen et al., 2019), while in the latter genetic diversity also positively
80 correlated with island size, another potential proxy for the long-term N_e (Brüniche-Olsen et al.,
81 2019). Surprisingly, however, genetic diversity variation across Metazoans is much better ex-
82 plained by fecundity and propagule size than classical predictors of species abundance such as
83 body size and geographic range (Romiguier et al., 2014). This result has been attributed to
84 differences in extinction risk for species that have contrasted reproductive strategies. Under
85 this hypothesis, species with low fecundity and large propagule size (K -strategists) would be
86 more resilient to low population size episodes compared to species with high fecundity and
87 small propagule size (r -strategists) which would go extinct if they reach such population sizes
88 (Romiguier et al., 2014). By contrast, Mackintosh et al. (2019) found no effect of propagule
89 size on genetic diversity within Papilionoidea, a superfamily showing little variation in repro-
90 ductive strategy. Therefore, the major effect of the r/K gradient on genetic diversity variation
91 across Metazoa probably hides other determinants that act within smaller branches of the tree
92 of life. In particular, how demography and evolutionary processes influence genetic variation
93 in different taxa remains unclear.

94 Other factors than fluctuations in population size are known to reduce the value of N_e rela-
95 tive to the census population size, impacting the $\frac{N_e}{N}$ ratio to a different extent from one species
96 to another. These factors include unbalanced sex-ratios, variance in lifetime reproductive suc-
97 cess among individuals, age structure, kinship-correlated survival and some metapopulation

98 configuration (Wright, 1969; Falconer, 1989; Lande and Barrowclough, 1987). A potentially
99 strong effect comes from variance in the number of offspring per parent (V_k), which reduces N_e
100 compared to N following $N_e = \frac{4N-4}{V_k+2}$ (Crow and Kimura, 1970). Variance in reproductive suc-
101 cess can naturally emerge from particular age-specific demographic characteristics summarized
102 in life tables that contain age- (or stage-) specific survival and fecundity rates (Ricklefs and
103 Miller, 1999). The impact of life tables characteristics on expected $\frac{N_e}{N}$ ratio has been the focus
104 of a large body of theoretical and empirical works (Nunney, 1991, 1996; Waples, 2002, 2016b,a;
105 Waples et al., 2018). Accounting for iteroparity and overlapping generations, a meta-analysis
106 of vital rates in 63 species of plants and animals revealed that half of the variance in $\frac{N_e}{N}$ among
107 species can be explained by just two life history traits: adult lifespan and age at maturity
108 (Waples et al., 2013). Interestingly, longevity was the second most important factor explaining
109 differences in genetic diversity across Metazoans (Romiguier et al., 2014). However, there is still
110 no attempt to evaluate the extent to which lifetime variance in reproductive success explains
111 differences in genetic diversity between species with different life table components.

112 Marine fishes are good candidates to address this issue. They are expected to show a partic-
113 ularly high variance in reproductive success as a result of high abundance, type III survivorship
114 curves (i.e. high juvenile mortality and low adult mortality) and increasing fecundity with age.
115 Consequently, it has been suggested that marine fish species show a marked discrepancy be-
116 tween adult census size and effective population size, resulting in $\frac{N_e}{N}$ ratios potentially smaller
117 than 10^{-3} . The disproportionate contribution of a few lucky winners to the offspring of the next
118 generation is sometimes referred as the "big old fat fecund female fish" (BOFFFF) effect, a
119 variant of the "sweepstakes reproductive success" hypothesis (Hedgecock, 1994; Hedrick, 2005;
120 Hedgecock and Pudovkin, 2011) that is often put forward to explain low empirical estimates
121 of effective population sizes from genetic data (Hauser and Carvalho, 2008). However, subse-
122 quent theoretical work showed that low values of $\frac{N_e}{N}$ below 0.01 can only be generated with
123 extreme age-structure characteristics (Waples, 2016b). The real impact of lifetime variance
124 in reproductive success on genetic diversity thus remains unclear, even in species like fish in
125 which its impact is supposed to be strong. Contrasting results have been obtained by com-
126 parative studies in marine fishes, including negative relationship between diversity and body
127 size (Pinsky and Palumbi, 2014; Waples, 1991), fecundity (Martinez et al., 2018) and overfish-
128 ing (Pinsky and Palumbi, 2014). However, these studies relied on few nuclear markers, that
129 could provide inaccurate or biased estimates of genetic diversity (Väli et al., 2008). They also
130 compared species sampled from different locations, thus, likely having different demographic
131 histories, which could blur the relationship between species characteristics and genetic diversity
132 (Ellegren and Galtier, 2016).

133 Here, we compared the genome-average heterozygosity to the life history traits and life
134 table characteristics of 16 marine teleostean species sharing similar Atlantic and Mediterranean
135 distributions. We estimated genetic diversity from unassembled whole-genome reads using
136 **GenomeScope** (Vurture et al., 2017) and checked the validity of these estimates with those
137 obtained using a high-standard reference-based variant calling approach. Using this data, we
138 related species genetic diversity to eight simple quantitative and qualitative life history traits.
139 Then, we built species life tables and determined if the lifetime variance in reproductive success
140 induced by these tables could explain observed differences in genetic diversity using an analytical
141 and a forward-in-time simulation approach. Finally, we generalized our findings by exploring
142 the influence of age-specific survival and fecundity rates on the variance in reproductive success
143 and ultimately genetic diversity via simulated lifetimes tables.

144 Material and Methods

145 Sampling, DNA extraction and whole-genome sequencing

146 We sampled 16 marine teleostean fish species presenting a wide diversity of life history strate-
147 gies expected to affect genetic diversity (Table 1). All these species share broadly overlapping
148 distributions across the North-eastern Atlantic and Mediterranean regions. Sampling was per-
149 formed at the same four locations for all species: two in the Atlantic (the Bay of Biscay in
150 South-western France or North-western Spain and the Algarve in Portugal), and two in the
151 Western Mediterranean Sea (the Costa Calida region around Mar Menor in Spain and the Gulf
152 of Lion in France see Fig 1A). Individual whole-genome sequencing libraries were prepared fol-
153 lowing the Illumina TruSeq DNA PCR-Free Protocol and sequenced to an average depth of 20X
154 on an Illumina NovaSeq 6000 platform by Genewiz Inc (USA). Raw reads were preprocessed
155 with `fastp` v.0.20.0 (Chen et al., 2018) using default parameters (see Supplementary Material).

156 Estimation of genetic diversity

157 We used `GenomeScope` v.1.0 to estimate individual genome-wide heterozygosity (Vurture et al.,
158 2017). Briefly, this method uses a k -mers based statistical approach to infer overall genome
159 characteristics, including total haploid genome size, percentage of repeat content and genetic
160 diversity from unassembled short-read sequencing data. We used `jellyfish` v.2.2.10 to com-
161 pute the k -mer profile of each individual (Marçais and Kingsford, 2011). The genetic diversity
162 of each species was determined as the median of the individual genome-wide heterozygosity val-
163 ues. We chose the median instead of the mean diversity since it is less sensitive to the possible
164 presence of individuals with non-representative genetic diversity values (e.g. inbred or hybrid
165 individuals) in our samples.

166 In order to assess the reliability of `GenomeScope` and detect potential systematic bias, we
167 compared our results with high-standard estimates of genetic diversity obtained after read
168 alignment against available reference genomes (see details in Supplementary Material). To
169 perform this test, we used the sea bass (*D. labrax*) and the European pilchard (*S. pilchardus*),
170 two species that represent the lower and upper limits of the range of genetic diversity in our
171 dataset (Table 1, Fig 1D).

172 Life history traits database

173 We collected seven simple quantitative variables describing various aspects of the biology and
174 ecology of the 16 species: body size, trophic level, fecundity, propagule size, age at maturity,
175 lifespan and adult lifespan (Table 1, Table S4 for detailed informations on bibliographic refer-
176 ences). We used the most representative values for each species and each trait when reported
177 traits varied among studies due to plasticity, selection or methodology. In addition, we collected
178 two qualitative variables describing the presence/absence of hermaphroditism and brooding be-
179 haviour, as revealed by males carrying the eggs in a brood pouch (*H. guttulatus* and *S. typhle*)
180 or nest-guarding (*C. galerita*, *S. cinereus* and *S. cantharus*). Detailed information on data
181 collection is available in Supplementary Material.

182 Construction of life tables

183 Life tables summarize survival rates and fecundities at each age during lifetime (Ricklefs and
184 Miller, 1999). Thus, they provide detailed information on vital rates that influence the variance
185 in lifetime reproductive success among individuals. This tool is well designed to describe popu-
186 lation structure from the probability of survival to a specific age at which a specific number of

187 offspring are produced. Ideally, age-specific survival is estimated by direct demographic mea-
 188 sures, such as mark-recapture. Unfortunately, direct estimates of survival were not available
 189 for the 16 studied species. We thus followed [Benvenuto et al. \(2017\)](#) to construct species life
 190 tables. Age-specific mortality of species sp , $m_{sp,a}$, is a function of species body length at age
 191 a , $L_{sp,a}$, species asymptotic Von Bertalanffy length $L_{sp,inf}$, and species Von Bertalanffy growth
 192 coefficient, K_{sp} :

$$m_{sp,a} = \left[\left(\frac{L_{sp,a}}{L_{sp,inf}} \right) \right]^{-\frac{1}{5}} \times K_{sp} \quad (1)$$

193 Age-specific survival rates, $s_{sp,a}$ were then estimated as:

$$s_{sp,a} = e^{-m_{sp,a}} \quad (2)$$

194 We collected age-specific length from empirical data and estimated L_{inf} and K values from
 195 age-length data as explained in the appendix, setting survival probability to zero at the maxi-
 196 mum age (Appendix 1). When differences in age-specific lengths between sexes were apparent
 197 in the literature, we estimated a different age-specific survival curve for each sex. The relation-
 198 ship between absolute fecundity and individual length is usually well fitted with the power-law
 199 function ($F = \alpha L^\beta$), although some studies also used an exponential function ($F = \alpha e^{\beta L}$) or
 200 a linear function ($F = \alpha + L\beta$). We collected empirical estimates of α and β and determined
 201 age-specific fecundity from the age-specific length and the fecundity-length function reported
 202 in the literature for each species. Fecundity was set to zero before the age at first maturity.

203 Effect of the variance in reproductive success on the N_e/N ratio

204 To understand how differences in life tables drive differences in genetic diversity between species,
 205 we estimated the variance in lifetime reproductive success, V_k and the ensuing ratio $\frac{N_e}{N}$ using the
 206 analytic framework developed in [AgeNe \(Waples et al., 2011\)](#). [AgeNe](#) infers V_k using informations
 207 from life tables only. Hence, the estimated variance in reproductive success estimated is only
 208 generated by inter-individual differences in fecundity and survival. [AgeNe](#) assumes constant
 209 population size, stable age structure, and no heritability of survival and fecundity. We used the
 210 life tables constructed as described above and set the number of new offspring to 1000 per year.
 211 This setting is an arbitrary value which has no influence on the estimation of either V_k nor $\frac{N_e}{N}$
 212 by [AgeNe](#). For all species, we set an initial sex ratio of 0.5 and equal contribution of individuals
 213 of the same age (i.e. no sweepstake reproductive success among same-age individuals). We ran
 214 [AgeNe](#) and estimated $\frac{N_e}{N}$ for each species.

215 Four life tables components can generate differences in $\frac{N_e}{N}$ between species: age at matu-
 216 rity, age-specific survival rates, age-fecundity relationships and sex-related differences in these
 217 components. To determine the role that each parameter plays in shaping levels of genetic diver-
 218 sity among species, we built 16 alternative life tables where the effect of each component was
 219 added one after the other, while the others were kept constant across species. Thus, in our null
 220 model, age at maturity was set at 1 year old for all species, fecundity and survival did not vary
 221 with age (constant survival chosen to have 0.01% of individuals remaining at maximum age,
 222 following [Waples \(2016b\)](#)), and there were no differences between sexes. Next, the effect of each
 223 component was tested by replacing these constant values with their biological values in species'
 224 life tables. For each of the 16 life tables thus constructed, we tested whether variation in $\frac{N_e}{N}$
 225 explained the variation in observed genetic diversity after scaling these two variables by their
 226 maximum value. With this scaling, the correlation between $\frac{N_e}{N}$ and genetic diversity should
 227 overlap with the $y = x$ function in cases where a decrease in $\frac{N_e}{N}$ predicts an equal decrease in
 228 genetic diversity, indicating a strong predictive power of the components induced in life tables.

Forward simulations

A complementary analysis of the contribution of life table properties on genetic diversity was performed using forward simulations in SLiM v.3.3.1 (Haller and Messer, 2017). Compared to the deterministic model implemented in AgeNe, the forward simulations include the stochastic variation inherent to the coalescent process and directly predict genetic diversity. Thus, they provide another approach to the problem and can lead to a more intuitive understanding of why vital rates affect N_e over the long-term, and ultimately genetic diversity. We simulated populations with overlapping generations, sex-specific lifespan, and age- and sex-specific fecundity and survival. We used life tables estimated as previously, and sex-specific lifespan estimates were collected in the literature as described above. However, age at maturity was not taken into account in these forward simulations for technical reasons. Age and species-specific fecundity were determined as previously and scaled between 0 (age 0) and 100 (maximum age) within each species. In the simulations, each individual first reproduces and then either survives to the next year or dies following a probability determined by its age and the corresponding life table. We kept population size constant and estimated the mean genetic diversity (i.e., the proportion of heterozygous sites along the locus) over the last 10000 years of the simulation after the mutation-drift equilibrium was reached and using 50 replicates (see Supplementary Material for further informations).

As previously, we evaluated the contribution of each component among 8 alternative life tables by comparing scaled observed and simulated genetic diversity.

Evaluating the impact of life tables beyond marine fish

To generalize our understanding of the influence of life tables on genetic diversity beyond the species analyzed in this study, we simulated a wide range of age-specific survival and fecundity curves and explored their effect on the relationship between adult lifespan and variance in reproductive success. To this end, we defined 16 theoretical species with age at first maturity and lifespan equal to that of our real species and then introduced variation in survival and fecundity curves. First, age-specific mortality was simulated following Pinder et al. (1978):

$$M(\text{Age}, \text{Age} + 1) = 1 - \exp\left(\frac{\text{Age}}{b}\right)^c - \left(\frac{\text{Age}+1}{b}\right)^c \quad (3)$$

where c defines the form of the survivorship curve, with $c > 1$, $c = 1$ and $c < 1$ defining respectively a *Type I* (e.g. mammals), *Type II* (e.g. birds) and *Type III* (e.g. fish) survival curves. We took values of c from 0.01 to 30 (Fig 4A). Parameter b was equal to $-\frac{\text{Lifespan}}{\log(0.01)^{1/c}}$ to scale survivorship curves in such a way that 1% of the initial population remains at maximum age.

Second, age-specific fecundity was simulated with two models: constant and exponential. In the first model, fecundity is constant for all ages since maturity. In the second model, fecundity increases or decreases exponentially with age following $F_{\text{Age}} = \exp^{f \times \text{Age}}$, as it is often observed in marine fishes (Curtis and Vincent, 2006). We first set $f = 0.142$ as the median of the f values for the 16 species. Secondly, we took values of f ranging from -1 to 1 (Fig 4A). We scaled maximum fecundity to 1 for all simulations.

For each combination of c and f , and for each fecundity model, we simulated all species life tables given age at maturity and lifespan. Then, we ran AgeNe and estimated $\frac{N_e}{N}$ for each simulated species and estimated the slope of the regression between adult lifespan and $\frac{N_e}{N}$ across all 16 species. We explored the impact of alternative fecundity-age models on the relationship between adult lifespan and $\frac{N_e}{N}$ (see details in Supplementary Material).

272 **Intraspecific variation in genetic diversity**

273 We addressed the potential effects of population structure, demography and historical contin-
274 gencies on genetic diversity by examining the extent of spatial variation in genetic diversity
275 between the four populations within each species. First, we evaluated the relative amount of
276 intraspecific compared to interspecific variation in genetic diversity. Then, we applied a z -
277 transformation of individual genetic diversity within each species to put spatial differences in
278 within-species diversity on the same scale. In order to detect similar spatial patterns of genetic
279 diversity among species, we finally performed a hierarchical clustering analysis of the matrix of
280 z -transformed genetic diversity values with `pheatmap` function available in `pheatmap v1.0.12`
281 R package.

282 **Statistical analyses**

283 All statistical analyses were carried out using R-3.6.1 (R Core Team, 2018). We fitted beta
284 regression models between genetic diversity and any covariate with the R-package `betareg`
285 v.3.1-3 (Cribari-Neto and Zeileis, 2010). We tested statistical interactions between any quanti-
286 tative and qualitative covariates using likelihood tests with the `lmttest v.0.9-37` package (Zeileis
287 and Hothorn, 2002).

Results

Whole-genome resequencing data set

We resequenced 300 individual genomes from 16 marine teleostean species, with high read quality scores (mean Q30 rate = 92.4%) and moderate duplication rates (10.8%) (Fig S2). GC content was moderately variable among species and highly consistent among individuals of the same species, except for three individuals that showed a marked discrepancy with the overall GC content of their species (Fig S2). These three individuals were thus removed from downstream analyses to avoid potential issues due to contamination or poor sequencing quality.

Estimation of genetic diversity with GenomeScope

The GenomeScope model successfully converged for all of the 297 individual genomes retained (Fig S6E). The average depth of sequencing coverage per diploid genome exceeded 20X in most individuals. Estimated genome sizes were very consistent within species (Fig S6A-C). Estimated levels of genetic diversity were also homogeneous among individuals of the same species with some few exceptions (e.g. *S. cinereus* and *S. typhle*) and most of the variability in genetic diversity was observed between species (Fig 1D). Two individuals (one *D. puntazzo* and one *P. erythrinus*) showed a surprisingly high genetic diversity (more than twice the average level of their species), indicating possible issues in the estimation of genome-wide heterozygosity. Therefore we removed these individuals from subsequent analysis, although their estimated genome size and GC content matched their average species values (therefore excluding contamination as a cause of genetic diversity estimation failures).

Observed values of genetic diversity ranged from 0.225% for *L. budegassa* to 1.415% for *S. pilchardus*. We found no correlation between species genetic diversity and genome size (p - value = 0.983). The estimation of genetic diversity was robust to the choice for k -mer lengths ranging from 21 to 25, suggesting a low sensitivity of GenomeScope regarding this parameter (Fig S4). The fraction of reads mapped against reference genomes ranged between 96.72 and 98.50% for *D. labrax* and between 87.45 and 96.42 % for *S. pilchardus* (Table S2; Fig S3). We found similar species genetic diversity estimates between GenomeScope and the GATK reference-based variant calling approach for the two control species, representing extreme values within the range of genetic diversity in our dataset (Fig 1B).

Adult lifespan is the best predictor of genetic diversity

318 We evaluated the effect of several key life history traits that potentially affect species genetic
319 diversity (Table S1).

320 Two widely used predictors of population size, body size and trophic level, were not sig-
321 nificantly correlated to genetic diversity (p -value = 0.119 and 0.676 respectively, Fig S8A-B).
322 Although we detected a significant negative relationship between the logarithm of fecundity
323 and propagule size (p -value = 0.00131, slope = -0.4385 ± 0.1076) as in Romiguier et al. (2014),
324 we found no significant correlation between either propagule size (p -value = 0.561), or the
325 logarithm of fecundity (p -value = 0.785) and genetic diversity (Fig S8C-D).

326 By contrast, both lifespan (p -value = 0.011) and adult lifespan (p -value = 0.007) were
327 significantly negatively correlated with genetic diversity (Table S1, Fig 2). The percentage of
328 variance explained by each variable reached 43.8 and 42.9 %, respectively. Repeating the same
329 statistical analyses with genetic diversity estimates obtained either only from mediterranean
330 or atlantic individuals led to the same results, revealing no effect of within-species population
331 structure on the relationship between genetic diversity and life history traits (Fig 1C, Fig S9,
332 Table S3).

333 We found no significant interaction between hermaphroditism and any of the previous vari-
334 ables on genetic diversity. By contrast, parental care showed a significant interaction with
335 lifespan (p -value = 0.0011), adult lifespan (p -value = 0.0008) and body size (p -value = 0.0035)
336 on genetic diversity. Brooding species (nest protection by males for *C. galerita*, *S. cinereus* and
337 *S. cantharus* and male abdominal brood-pouch for *H. guttulatus* and *S. typhle*) had systemat-
338 ically lower genetic diversity than non-brooding species with similar adult lifespan.

339 When considering only non-brooding species, we found steeper negative correlations and
340 higher percentages of between-species variance in genetic diversity explained by lifespan (p -value =
341 $1.017e^{-7}$, pseudo- $R^2 = 0.851$) and adult lifespan (p -value = $1.645e^{-7}$, pseudo- $R^2 = 0.829$, Fig
342 2, Table S1). To test the relevance of considering this sub-dataset, we estimated the slope of the
343 regression and the pseudo- R^2 for all combinations of 11 out of 16 species and compared the dis-
344 tribution of these values to the estimated slope and pseudo- R^2 obtained for the 11 non-brooding
345 species (Fig S13). The estimated slope for non-brooders lied outside of the 95% confidence in-
346 terval of the distribution of estimated slopes (slope = -0.129 , 95% CI = $[-0.122, -0.049]$) and
347 the same was found for pseudo- R^2 (pseudo- $R^2 = 0.829$, 95% CI = $[0.073, 0.727]$). Furthermore,
348 considering non-brooding species only, there was still no significant correlation between genetic
349 diversity and trophic level (p -value = 0.259), propagule size (p -value = 0.170), and fecundity
350 (p -value = 0.390), but genetic diversity appeared significantly negatively correlated to body
351 size (p -value = $6.602e^{-5}$, pseudo- $R^2 = 0.616$). We did not detect any significant correlation
352 between any trait variable and genetic diversity within the sub-dataset of brooding species.
353 However, this should be taken with caution given the very low number of brooding species
354 ($n = 5$) in our dataset.

355 Body size and lifespan were highly positively correlated traits in our dataset (p -value =
356 0.0013, $R^2 = 0.536$, Fig S7). Thus, using empirical observations only, it was not possible to
357 fully disentangle the impact of each of these traits among the possible determinants of genetic
358 diversity in marine fishes. However, we found important differences in effect sizes for body size
359 (slope = -0.014), lifespan (-0.095) and adult lifespan (-0.129), which rule out body size as a
360 major determinant of diversity in our dataset.

Table 1 - Life history traits and observed genetic diversity of the 16 teleostean marine species. - For each species, number of individuals used for the estimation of genetic diversity; observed median genetic diversity among all individuals (\pm standard deviation); body size (in centimeters); trophic level; fecundity (in eggs/day); propagule size (in millimeters); age at first maturity (in years), lifespan (in years), adult lifespan (in years, defined as the difference between lifespan and age at maturity), parental care behaviour (- = no egg protection; NG = nest-guarders; MP = male brood-pouch) and hermaphroditism (- = no hermaphroditism; PG = protogynous; PA = protandrous, RUD = rudimentary). Detailed bibliographic references are provided in supplementary material.

Species	Vernacular name	N.	Genetic diversity (%)	Body size (cm)	Trophic level	Fecundity	Propagule Size (mm)	Maturity (years)	Lifespan (years)	Adult lifespan (years)	Parental care	Herma.
<i>Coryphoblennius galerita</i>	Montagu's blenny	16	0.607(\pm 0.014)	7	2.28	NA	3.3	1.5	6	4.5	NG	-
<i>Coris julis</i>	Rainbow wrasse	20	1.172(\pm 0.056)	27.2	3.24	169.81	0.63	1	7	6	-	PG
<i>Dicentrarchus labrax</i>	European sea bass	20	0.375(\pm 0.031)	102.15	3.47	12436.52	1.15	3	15	12	-	-
<i>Diplodus puntazzo</i>	Sharp-snout seabream	19	0.533(\pm 0.074)	49.69	3.07	277.87	0.87	2	10	8	-	RUD
<i>Hippocampus guttulatus</i>	Long-snouted seahorse	12	0.313(\pm 0.090)	19.8	3.5	1.21	12	0.5	5	4.5	MP	-
<i>Lophius budegassa</i>	Black anglerfish	20	0.225(\pm 0.015)	103	4.23	2304.03	1.88	7.5	21	13.5	-	-
<i>Lithognathus mormyrus</i>	Striped seabream	20	0.553(\pm 0.027)	37.85	3.42	214.09	0.75	2	12	10	-	PA
<i>Merluccius merluccius</i>	European hake	20	0.844(\pm 0.025)	88.9	4.43	2294.54	1.07	3	11	8	-	-
<i>Mullus surmuletus</i>	Striped red mullet	19	1.135(\pm 0.048)	30.18	3.46	2569.32	0.86	1.5	6	4.5	-	-
<i>Pagellus erythrinus</i>	Common pandora	19	1.100(\pm 0.020)	36	3.46	2280.46	0.77	2	8	6	-	PG
<i>Serranus cabrilla</i>	Comber	19	1.205(\pm 0.055)	30.8	3.68	37.97	0.91	2	6	4	-	-
<i>Spondyliosoma cantharus</i>	Black seabream	19	0.478(\pm 0.034)	35.7	3.27	425.62	2.1	3	10	7	NG	PG
<i>Symphodus cinereus</i>	Grey wrasse	10	0.660(\pm 0.125)	14.1	3.3	13.20	2.87	1.5	6	4.5	NG	-
<i>Sardina pilchardus</i>	European pilchard	20	1.415(\pm 0.182)	20.35	2.94	22.89	1.64	1	5	4	-	-
<i>Syngnathus typhle</i>	Broadnosed pipefish	20	0.859(\pm 0.047)	26.2	3.75	0.38	20	1	3	2	MP	-
<i>Sarda sarda</i>	Atlantic bonito	20	0.896(\pm 0.208)	68.9	4.34	15647.73	1.3	1	4	3	-	-

Variance in reproductive success explains levels of observed genetic diversity

To understand the mechanisms by which adult lifespan affects genetic diversity and test if it can alone explain our results, we built life tables for each of the 16 species by gradually incorporating age-specific fecundity and survival, age at first maturity, lifespan and sex-specific differences in these parameters.

Non-genetic estimates of $\frac{N_e}{N}$ ratio obtained with **AgeNe** ranged from 0.104 in *L. budegassa* to 0.671 for *S. cinereus*. When considering the 16 species together, the $\frac{N_e}{N}$ ratio was not significantly correlated with genetic diversity (p -value = 0.0935). However, four out of five brooding species had low genetic diversity despite high $\frac{N_e}{N}$ ratios (Fig 3A). As previously observed, removing the 5 brooders increased the slope and the percentage of variance of genetic diversity explained by the $\frac{N_e}{N}$ ratio above null expectations obtained by removing groups of 5 species at random ($slope = 1.849$, 95% CI = [0.048, 1.582], $pseudo-R^2 = 0.55$, 95% CI = [0.004, 0.533], Fig S14). Thus, the $\frac{N_e}{N}$ ratio predicted by life tables was positively correlated to genetic diversity when considering non-brooding species only (Fig 3A, p -value = 0.000966).

Our next step was to determine the impact of each component of life tables as well as their combinations on genetic diversity (Fig 3C-G). Starting from a null model (Fig 3C), in which species life tables differed only in lifespan, we found that the $\frac{N_e}{N}$ ratio ranged from 0.558 to 0.733, a variance much lower than that of observed genetic diversities. Then, adding separately age at maturity (Fig 3D) or age-specific survival (Fig 3E) did not better predict the range of observed genetic diversities. However, combining age at maturity and age-specific survival (Fig 3F) or adding only age-specific fecundity (Fig 3G) enable us to explain the range of observed diversity values. Finally, combining these three parameters together (age at maturity, age-specific survival, and fecundity, model 8, Fig S10H) resulted in the best fitted slope for both non-brooding species and the whole data set. Adding sex-specific differences in life tables did not improve the fit, however (models 9 to 16, Fig S10I-P).

Our final step was to further explore the role of the variance in reproductive success on genetic diversity by simulating genetic diversity at mutation-drift equilibrium with the age-specific vital rates of the 16 species.

We simulated a population of 2000 individuals with age-specific survival and fecundity. As expected, including age-specific vital rates decreased the equilibrium level of genetic diversity compared to expectations under the classical Wright-Fisher model ($\theta = 4N_e\mu = 0.08\%$). It was reduced to 0.070% in the species with the least effect of age-specific vital rates (*C. galerita*), and down to 0.010% in the species with the greatest effect (*L. budegassa*). Again, simulated genetic diversity was not correlated to genetic diversity considering all 16 species (p -value = 0.297, Fig 3B), but significantly positively correlated within the sub-sample of the 11 non-brooding species (p -value = 0.0115).

Life tables drive correlation between lifespan and the N_e/N ratio

In order to determine the general effect of life table properties on the relation between adult lifespan and $\frac{N_e}{N}$ beyond the case of marine fish, we modeled 16 life tables with age at maturity and lifespan similar to those observed in our species but with simulated age-specific survival and fecundity (Fig 4A).

Considering models including constant fecundity with age, we found a significant relationship between adult lifespan and $\frac{N_e}{N}$ for species with type III survivorship curves ($c < 1$) but not for species having an age-specific survivorship curve constant, c , superior to 2, including type I species (Fig 4B). The slope between adult lifespan and $\frac{N_e}{N}$ was steepest for type III species, reaching -0.053 for $c = 0.1$. For $c < 2$, the percentage of variation in $\frac{N_e}{N}$ explained by adult lifespan was higher than 60%. Interestingly, it reached a maximum for $c = 1.03$ at 89% and abruptly dropped down around $c = 2$ (Fig 4B).

Then, we added an exponential increase in fecundity with age, first taking $f = 0.142$, which is close to the mean empirical estimation across our 16 species (Fig 4B). The slope between adult lifespan and $\frac{N_e}{N}$ became steeper for type I and type II species and reached -0.074 for extreme type III species ($c = 0.01$). When we included this exponential increase of fecundity with age, the percentage of variation explained was superior for approximately all values of c , and the abrupt drop of the percentage of variation explained shifted toward higher c values, around $c = 3$. Interestingly, we found significant positive relationships associated with low slope values when c became superior to 10 (type I species).

Then, we compared values of slope and R^2 for all c values and for f ranging from -1 to 1 (Fig 4C-D). The steepest slope between adult lifespan and $\frac{N_e}{N}$ that we obtained reached -0.076 for extreme type III species (c around 0.1), and exponential constant, f , between 0.18 and 0.31. For type III and type II species ($c < 1$), both the slope and the percentage of variation explained first increased with increasing exponential constant and then decreased. Significant negative relationships were found for $c < 1$ for any values of f , except some extreme values near -1, whereas no significant relationship was found for $c > 1$ when f is negative except for values of c near 1 and values of f near 0. The steepest slope and the highest percentage of variation explained were obtained for type III species with intermediate values of f ($0.1 < f < 0.5$) and for type II species ($1 < c < 5$) for positive values of f . For type I species, as c values increased, higher values of f are needed to obtain a significant negative relationship between adult lifespan and the $\frac{N_e}{N}$ ratio. Above $c > 20$, no significant negative relationship was found for any values of f . Again, we found significant positive relationships and low slopes for $c > 15$ and intermediate positive values of f .

We found similar results considering a power-law relationship between age and fecundity, with slightly flatter slopes between $\frac{N_e}{N}$ and adult lifespan, and no significant correlations for extreme positive values of f and extreme low values of c (Fig S16C). In contrast, we found limited or no impact of f on the relationship between $\frac{N_e}{N}$ and adult lifespan, respectively, for the linear and the polynomial age-fecundity model (Fig S16A-B).

Discussion

In this study, we used whole-genome high-coverage sequencing data to estimate the genetic diversity of 16 marine teleost fish with similar geographic distribution ranges. We found that adult lifespan was the best predictor of genetic diversity, species with long reproductive lifespans generally having lower genetic diversities (Fig 2). Longevity was already identified as one of the most important determinants of genetic diversity across Metazoans and plants, in which it also correlates with the efficacy of purifying selection (Romiguier et al., 2014; Chen et al., 2017). A positive correlation between longevity and the ratio of nonsynonymous to synonymous substitutions (dN/dS) was also found in teleost fishes (Rolland et al., 2020), thus suggesting lower N_e in long-lived species. However, the mechanisms by which lifespan impacts genetic diversity remain poorly understood and may differ among taxonomic groups. Here we showed that age-specific fecundity and survival (i.e. vital rates), summarized in life tables, naturally predict the empirical correlation between adult lifespan and genetic diversity in marine fishes.

Impact of life tables on genetic diversity

On a broad taxonomic scale including plants and animals, Waples et al. (2013) showed that almost half of the variance in $\frac{N_e}{N}$ estimated from life tables can be explained with only two life history traits: age at maturity and adult lifespan. Therefore, the effect of adult lifespan on genetic diversity should reflect variations in age-specific fecundity and survival across species. If the species vital rates used to derive $\frac{N_e}{N}$ ratios are relatively stable over time, the reduction in N_e due to lifetime variance in reproductive success should not only apply to contemporary time scales but more generally throughout the coalescent time. Thus, a direct impact of life tables on genetic diversity can be expected for iteroparous species with overlapping generations.

Using both an analytical (with AgeNe) and a simulation-based (with SLiM) approach, we showed that age-specific survival and fecundity rates alone can explain a significant fraction of the variance in genetic diversity among species (Fig 3A-B). This may appear surprising at first sight, considering that we did not account for among species variation in population census sizes, which vary by several orders of magnitude in marine fishes (Hauser and Carvalho, 2008). Our results thus support that intrinsic vital rates are crucial demographic components of the neutral model to understand differences in levels of genetic diversity in marine fishes. But how generalizable is this finding to other taxa?

Age-specific survivorship curves are one of the main biological components of life tables. Three main types of survivorship curves are classically distinguished: type I curves are characterized by low juvenile and adult mortality combined with an abrupt decrease of survival when approaching the maximum age (e.g. mammals); in type II curves, survival is relatively constant during lifetime (e.g. birds) while type III curves are characterized by high juvenile mortality followed by low adult mortality (e.g. fishes and marine invertebrates). Type III survivorship curves favor the disproportionate contribution of a few lucky winners that survive to old age, compared to type I survivorship curves, where individuals have more equal contributions to reproduction, generating a lower variance in reproductive success. Thus, in type III species, higher lifetime variance in reproductive success is expected as the lifespan of a species increases. By simulating extreme type III survivorship curves ($c = 0.1$) for our 16 species while keeping their true adult lifespans, we found that $\frac{N_e}{N}$ can decrease by at most 0.05 per year of lifespan (Fig 4B, extreme left). This can theoretically induce up to 60% difference in genetic diversity between the species with the shortest and the longest lifespans of our dataset. In contrast, we found no correlation between adult lifespan and $\frac{N_e}{N}$ when simulating type I survivorship curves with the true lifespan values of the 16 species studied here (Fig 4B, $c > 2$), meaning than lifespan and variance in reproductive success may have limited influence in other taxonomic

485 groups, such as birds or mammals.

486 Another important component of life tables is age-specific fecundity. In marine fishes,
487 fecundity is positively correlated to female ovary size, and the relationship between fecundity
488 and age is usually well approximated with an exponential ($F = a \exp^{Ab}$) or power-law ($F = aA^b$)
489 function. By adding an exponential increase in fecundity with age to our simulations, we found
490 that $\frac{N_e}{N}$ decreases even more strongly with increasing adult lifespan ($\frac{N_e}{N}$ decreases by up to
491 0.07 per extra year of reproductive life). Using both type III survivorship and exponentially
492 increasing fecundity with age, we could thus predict up to 84% of the variance in genetic
493 diversity between species with the shortest and longest lifespans.

494 We found that $\frac{N_e}{N}$ predicted from fecundity alone or age at maturity combined with age-
495 specific survival, explained as much variation in genetic diversity as life tables with both of
496 these components (Fig 3). This is because both of these two scenarios create sharp differences
497 in fitness between young and old age classes. By contrast, variation in age at maturity alone (all
498 other parameters being held constant across species) introduces some variation in $\frac{N_e}{N}$ because
499 the onset of reproduction varies from 1 to 7 years old depending on the species, but this effect
500 is buffered by the long subsequent period during which adults will reproduce equally. Similarly,
501 the effect of survival alone is insufficient if individuals of all species start reproducing early
502 enough at 1 year old.

503 Although these predicted relationships were pretty close to our empirical findings, genome-
504 wide heterozygosity decreased by about 0.09 per additional year of lifespan in our real dataset
505 (Fig 2), which seems to be a stronger effect compared to theoretical predictions based on vital
506 rates alone. It is thus likely that other correlates of adult lifespan and unaccounted factors also
507 contribute to observed differences in genetic diversity among species.

508 Correlated effects

509 When relating measures of diversity with the estimates of $\frac{N_e}{N}$ derived from life tables, we
510 did not take into account differences in census size (N) between species. Population census
511 sizes can be huge and are notoriously difficult to estimate in marine fishes. For that reason,
512 abundance data remain largely unavailable for the 16 species of this study. We nevertheless
513 expect long-lived species to have lower abundance compared to short-lived species because in
514 marine fishes N is generally negatively correlated to body size (White et al., 2007), which
515 is itself positively correlated to adult lifespan in our dataset (Fig S7). Hence, while we have
516 demonstrated here that variation in vital rates has a direct effect on long-term genetic diversity,
517 the slope between adult lifespan and genetic diversity may be inflated by uncontrolled variation
518 in N . Recent genome-wide comparative studies found negative correlations between $\frac{N_e}{N}$ and N
519 in Pinnipeds (Peart et al., 2020) as well as between genetic diversity and body size in butterflies
520 and birds (Mackintosh et al., 2019; Brüniche-Olsen et al., 2019). Here, a highly significant
521 negative correlation was found between genetic diversity and body size and the strength of that
522 correlation was comparable to that found in a meta-analysis of microsatellite diversity using
523 catch data and body size as proxies for fish abundance (McCusker and Bentzen, 2010). We note,
524 however, that body size was not as good a predictor of genetic diversity as lifespan and adult
525 lifespan for the 11 non-brooding species and it was even not significant in the whole dataset of
526 the 16 species (Table S1).

527 Another potentially confounding effect is the impact of r/K strategies which are the main
528 determinant of genetic diversity across Metazoans (Romiguier et al., 2014). In our dataset,
529 fecundity and propagule size (proxies for the r/K gradient) showed only little variance compared
530 to their range of variation across Metazoans, and none of them were correlated to adult lifespan.
531 However, we found that the 5 brooding species of our dataset, which are typical K-strategists,
532 displayed lower genetic diversities with respect to their adult lifespan (Fig 2). Most interestingly,

533 when these species were removed from the analysis, the effect of adult lifespan on genetic
534 diversity was amplified, indicating a potentially confounding effect of parental care in marine
535 fishes. Alternatively, low levels of genetic diversity in brooding species can also be explained by
536 underestimated lifetime variance in reproductive success by AgeNe due to unaccounted variance
537 in reproductive success within age-classes. This may be particularly important in males as
538 the age-fecundity relationship is empirically estimated for females only. This effect could be
539 high for species with strong sexual selection and mate choice (Hastings, 1988; Naud et al.,
540 2009). Moreover, most of these species inhabit lagoons and coastal habitats, corresponding to
541 smaller and more instable ecological niches compared to species with no parental care, thus
542 potentially resulting in lower long-term abundances. The discrepancy introduced by brooders
543 in the relationship that we observed here between adult lifespan and genetic diversity may thus
544 involve a variety of effects that remain to be elucidated.

545 Temporal fluctuations of effective population size may also have impacted observed levels
546 of genetic diversity (Nei et al., 1975). All studied species possibly went through a bottleneck
547 during the Last Glacial Maximum (Jenkins et al., 2018), which may have simultaneously de-
548 creased their genetic diversities. As the time of return to mutation-drift equilibrium is positively
549 correlated to generation time, which is itself directly linked to adult lifespan, we may expect
550 long-lived species to have recovered less genetic variation than short-lived species following their
551 latest bottleneck. Moreover, long-lived species may not have recovered their pre-bottleneck pop-
552 ulation sizes as rapidly as short-lived species. If true, the negative relationship between adult
553 lifespan and genetic diversity may be inflated compared to the sole effect of life tables.

554 Variation in mutation rates between species could not be accounted for due to a lack of
555 estimates. However, if species-specific mutation rates were correlated with adult lifespan, we
556 would expect mutation rate variation to have a direct effect on genetic diversity. Mutation
557 rate could be linked with species life history traits through three possible mechanisms. First,
558 the drift-barrier hypothesis predicts a negative correlation between species effective population
559 size and the per-generation mutation rate (Sung et al., 2012). However, this hypothesis can
560 not explain our results since species with the highest effective population sizes have the highest
561 genetic diversity. Second, species with larger genome size tend to have more germline cell
562 divisions, hence possibly higher mutation rates. But we did not find any correlation between
563 genome size and genetic diversity or any other qualitative and quantitative life history traits.
564 Third, species with longer generation time, which is positively correlated to lifespan and age
565 at maturity, may have higher per-generation mutation rate as older individuals accumulate
566 more germinal mutations throughout their lives. Again, under this assumption, we would
567 expect species with longer lifespan to have higher mutation rate and genetic diversity, which
568 goes against our observations. In summary, variation in mutation rates among species due
569 to differences in lifespan is unlikely to explain the negative lifespan-diversity relationship we
570 observed. If anything, variation in mutation rates should theoretically oppose this relationship.

571 Using one of the few direct estimates of the per-generation mutation rate in fish, Feng et al.
572 (2017) explained the surprisingly low nucleotide diversity found in the Atlantic herring *Clupea*
573 *harengus* ($\pi = 0.3\%$) by a very low mutation rate of 2×10^9 per base per generation estimated
574 from pedigree analysis. Although the herring is one of the most abundant and fecund pelagic
575 species in the North Atlantic Ocean, its genetic diversity appears approximately 80% lower
576 than that of the European pilchard *S. pilchardus*, another member of the *Clupeidae* family
577 that shows the highest diversity in our study. Even if *C. harengus* has a larger body size
578 (approximately 30 cm, compared to 20 cm for *S. pilchardus*, Froese et al. (2000)), it has above
579 all a much longer lifespan (between 12 and 25 years) and a later age at maturity (between 2
580 and 6.5 years) (Jennings and Beverton, 1991). Considering even the lowest estimate of adult
581 lifespan reported for the herring (10 years), the corresponding genetic diversity predicted by
582 our model linking adult lifespan to genetic diversity would be around 0.5 %, which is pretty

583 close to the empirical estimate.

584 Finally, we did not take into account the erosion of neutral diversity through linked se-
585 lection. Addressing that issue would need to generate local estimates of nucleotide diversity
586 and population recombination rate along the genome of each species using resequencing data
587 aligned to a reference assembly, which was out of the scope of this study. The predicted effect
588 of linked selection could be, however, to remove more diversity in species with large compared
589 to small N_e . It is therefore likely that linked selection would rather attenuate the negative
590 relationship between adult lifespan and genetic diversity compared to neutral predictions.

591 Conclusion

592 Here we used a simple approach to generate reference-free genome-wide estimates of diversity
593 with k -mer analyses. Tested on two species with genetic diversities ranging from 0.22 to 1.42%
594 the k -mer approach performed close to the level of a high-standard reference-based method in
595 capturing fine-scale variation in diversity between evolutionary lineages and even populations
596 of the same species. This opens the possibility to address the determinants of genetic diversity
597 in other groups of taxa at limited costs without relying on existing genomics resources. Across
598 Metazoans, the level of genetic diversity showed no significant relationship with the species'
599 conservation status (Romiguier et al., 2014). Studies performed at lower phylogenetical scales
600 such as in Darwin's finches and Pinnipeds, however, found reduced contemporary genetic di-
601 versity in threatened compared to non-threatened species (Brüniche-Olsen et al., 2019; Peart
602 et al., 2020). Our results complement and extend this literature by showing the importance of
603 taking into account life tables in comparisons of genetic diversity between species.

604 Acknowledgments

605 The data used in this work were partly produced with the support of the GenSeq genotyp-
606 ing and sequencing platform, and bioinformatics data analysis benefited from the Montpellier
607 Bioinformatics Biodiversity MBB platform, both platforms being supported by ANR program
608 "Investissements d'avenir" (ANR-10-LABX-04-01). We would like to thank Rémy Dernas and
609 Khalid Belkhir for their invaluable assistance in data storage, management and processing. We
610 are grateful to the colleagues who provided us with samples as well as to those who facilitated or
611 participated in sampling: F. Schlichta, T. Pastor, R. Castilho, R. Cunha, R. Lechuga, D. Pilo,
612 C. Mena, J. Charton, T. Robinet, A. Darnaude, S. Vaz, M. Duranton, N. Bierne, S. Villéger,
613 S. Blouet, as well as the fishermen and employees of fish markets and fish auctions. This work
614 was supported by the ANR grant CoGeDiv ANR-17-CE02-0006-01. The authors declare no
615 conflicts of interest.

616 Author contributions

617 P.B., T.B. and P.-A.G. wrote the manuscript. P.B. and P.-A.G. performed fieldwork. P.B.
618 performed molecular experiments, and all bioinformatics and evolutionary genomics analyses
619 with inputs from T.B. and P.-A.G. P.-A.G. conceived the project and managed financial support
620 and genome sequencing.

621 Data archiving

622 Data and scripts used in this study are freely available in the GitHub repository [https://](https://github.com/pierrebarry/life_tables_genetic_diversity_marine_fishes)
623 github.com/pierrebarry/life_tables_genetic_diversity_marine_fishes. All sampling
624 metadata are accessible under GEOME at the CoGeDiv Project Homepage: <https://geome-db>.

625 [org/workbench/project-overview?projectId=357](https://www.ncbi.nlm.nih.gov/workbench/project-overview?projectId=357). Sequence reads have been deposited in
626 the GenBank Sequence Read Archive under the accession code BioProject ID PRJNA777424.

627 **References**

- 628 Benvenuto, C., Coscia, I., Chopelet, J., Sala-Bozano, M., and Mariani, S. (2017). Ecological
629 and evolutionary consequences of alternative sex-change pathways in fish. Scientific Reports,
630 7(1):9084.
- 631 Brüniche-Olsen, A., Kellner, K. F., and DeWoody, J. A. (2019). Island area, body size
632 and demographic history shape genomic diversity in Darwin’s finches and related tanagers.
633 Molecular Ecology, 28(22):4914–4925.
- 634 Chen, J., Glémin, S., and Lascoux, M. (2017). Genetic Diversity and the Efficacy of Purifying
635 Selection across Plant and Animal Species. Molecular Biology and Evolution, 34(6):1417–
636 1428.
- 637 Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ
638 preprocessor. Bioinformatics, 34(17):i884–i890.
- 639 Cribari-Neto, F. and Zeileis, A. (2010). Beta Regression in R. Journal of Statistical Software,
640 34(1):1–24.
- 641 Crow, J. F. and Kimura, M. (1970). An Introduction to Population Genetics Theory. Harper
642 & Row.
- 643 Curtis, J. M. R. and Vincent, A. C. J. (2006). Life history of an unusual marine fish: Survival,
644 growth and movement patterns of *Hippocampus guttulatus* Cuvier 1829. Journal of Fish
645 Biology, 68(3):707–733.
- 646 DeWoody, J. A., Harder, A. M., Mathur, S., and Willoughby, J. R. (2021). The long-standing
647 significance of genetic diversity in conservation. Molecular Ecology, 30(17):4147–4154.
- 648 Díez-Del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., and Dalén, L.
649 (2018). Quantifying Temporal Genomic Erosion in Endangered Species. Trends in Ecology
650 & Evolution, 33(3):176–185.
- 651 Domínguez-Seoane, R., Pajuelo, J. G., Lorenzo, J. M., and Ramos, A. G. (2006). Age and
652 growth of the sharpnout seabream *Diplodus puntazzo* (Cetti, 1777) inhabiting the Ca-
653 narian archipelago, estimated by reading otoliths and by backcalculation. Fisheries Research,
654 81(2):142–148.
- 655 Ellegren, H. and Galtier, N. (2016). Determinants of genetic diversity. Nature Reviews Genetics,
656 17(7):422–433.
- 657 Falconer, D. S. (1989). Introduction to Quantitative Genetics. Longman, Scientific & Technical
658 ; Wiley, Burnt Mill, Harlow, Essex, England : New York, 3rd ed edition.
- 659 Feng, C., Pettersson, M., Lamichhaney, S., Rubin, C.-J., Rafati, N., Casini, M., Folkvord, A.,
660 and Andersson, L. (2017). Moderate nucleotide diversity in the Atlantic herring is associated
661 with a low mutation rate. eLife, 6:e23907.
- 662 Frankham, R. (1995). Conservation Genetics. Annual Review of Genetics, 29(1):305–327.
- 663 Froese, R., Pauly, D., and Editors (2000). FishBase 2000: Concepts, design and data sources.
664 page 344.
- 665 Gage, T. B. (2001). Age-specific fecundity of mammalian populations: A test of three mathe-
666 matical models. Zoo Biology, 20(6):487–499.

- 667 Ganas, K., Somarakis, S., Machias, A., and Theodorou, A. J. (2003). Evaluation of spawn-
668 ing frequency in a Mediterranean sardine population (*Sardina pilchardus sardina*). Marine
669 Biology, 142(6):1169–1179.
- 670 Haller, B. C. and Messer, P. W. (2017). SLiM 2: Flexible, Interactive Forward Genetic Simu-
671 lations. Molecular Biology and Evolution, 34(1):230–240.
- 672 Hastings, P. A. (1988). Female choice and male reproductive success in the angel blenny,
673 *Coralliozetus angelica* (Teleostei: Chaenopsidae). Animal Behaviour, 36(1):115–124.
- 674 Hauser, L. and Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly
675 hypotheses slain by beautiful facts. Fish and Fisheries, 9(4):333–362.
- 676 Hedgecock, D. (1994). Does variance in reproductive success limit effective population sizes of
677 marine organisms? In A. Genetics and Evolution of Aquatic Organisms.
- 678 Hedgecock, D. and Pudovkin, A. I. (2011). Sweepstakes Reproductive Success in Highly Fe-
679 cund Marine Fish and Shellfish: A Review and Commentary. Bulletin of Marine Science,
680 87(4):971–1002.
- 681 Hedrick, P. (2005). Large variance in reproductive success and the N_e/N ratio. Evolution,
682 59(7):1596–1599.
- 683 Iglésias, S. (2013). Actinopterygians from the North-Eastern Atlantic and the Mediterranean (A
684 Natural Classification Based on Collection Specimens, with DNA Barcodes and Standardized
685 Photographs), Volume I (Plates), Provisional Version 09.
- 686 Jenkins, T. L., Castilho, R., and Stevens, J. R. (2018). Meta-analysis of northeast Atlantic
687 marine taxa shows contrasting phylogeographic patterns following post-LGM expansions.
688 PeerJ, 6:e5684.
- 689 Jennings, S. and Beverton, R. J. H. (1991). Intraspecific variation in the life history tactics of
690 Atlantic herring (*Clupea harengus* L.) stocks. ICES Journal of Marine Science, 48(1):117–125.
- 691 Kimura, M. (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press,
692 Cambridge.
- 693 Kimura, M. and Crow, J. F. (1964). The Number of Alleles That Can Be Maintained in a
694 Finite Population. Genetics, 49(4):725–738.
- 695 Kraljević, M., Matić-Skoko, S., Dulčić, J., Pallaoro, A., Jardas, I., and Glamuzina, B. (2007).
696 Age and growth of sharpnose seabream *Diplodus puntazzo* (Cetti, 1777) in the eastern
697 Adriatic Sea.
- 698 Lande, R. (1995). Mutation and Conservation. Conservation Biology, 9(4):782–791.
- 699 Lande, R. and Barrowclough, G. F. (1987). Effective population size, genetic variation, and their
700 use in population management. In Soulé, M. E., editor, Viable Populations for Conservation,
701 pages 87–124. Cambridge University Press, Cambridge.
- 702 Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto,
703 P., and Przeworski, M. (2012). Revisiting an Old Riddle: What Determines Genetic Diversity
704 Levels within Species? PLOS Biology, 10(9):e1001388.

- 705 Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A., Scornavacca, C., Carrasco, M. R., Fuchs,
706 J., Illera, J. C., Swardt, D. H. D., Thébaud, C., Milà, B., and Nabholz, B. (2020). Endemic
707 island songbirds as windows into evolution in small effective population sizes. bioRxiv, page
708 2020.04.07.030155.
- 709 Lewontin, R. C. (1974). The Genetic Basis of Evolutionary Change. Number 25 in Columbia
710 Biological Series. Columbia Univ. Pr, New York.
- 711 Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
712 transform. Bioinformatics (Oxford, England), 25(14):1754–1760.
- 713 Louro, B., De Moro, G., Garcia, C., Cox, C. J., Veríssimo, A., Sabatino, S. J., Santos, A. M.,
714 and Canário, A. V. M. (2019). A haplotype-resolved draft genome of the European sardine
715 (*Sardina pilchardus*). GigaScience, 8(5).
- 716 Mackintosh, A., Laetsch, D. R., Hayward, A., Charlesworth, B., Waterfall, M., Vila, R.,
717 and Lohse, K. (2019). The determinants of genetic diversity in butterflies. Nature
718 Communications, 10(1):1–9.
- 719 Marçais, G. and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting
720 of occurrences of k-mers. Bioinformatics, 27(6):764–770.
- 721 Martinez, A. S., Willoughby, J. R., and Christie, M. R. (2018). Genetic diversity in fishes is
722 influenced by habitat type and life-history variation. Ecology and Evolution, 8(23):12022–
723 12031.
- 724 Mccusker, M. R. and Bentzen, P. (2010). Positive relationships between genetic diversity and
725 abundance in fishes. Molecular Ecology, 19(22):4852–4862.
- 726 Milton, P. (1983). Biology of littoral blennioid fishes on the coast of south-west England. Journal
727 of the Marine Biological Association of the United Kingdom, 63(1):223–237.
- 728 Murua, H. and Motos, L. (2006). Reproductive strategy and spawning activity of the European
729 hake *Merluccius merluccius* (L.) in the Bay of Biscay. Journal of Fish Biology, 69(5):1288–
730 1303.
- 731 Naud, M.-J., Curtis, J. M. R., Woodall, L. C., and Gaspar, M. B. (2009). Mate choice,
732 operational sex ratio, and social promiscuity in a wild population of the long-snouted seahorse
733 *Hippocampus guttulatus*. Behavioral Ecology, 20(1):160–164.
- 734 Nei, M., Maruyama, T., and Chakraborty, R. (1975). The Bottleneck Effect and Genetic
735 Variability in Populations. Evolution, 29(1):1–10.
- 736 Nunney, L. (1991). The Influence of Age Structure and Fecundity on Effective Population Size.
737 Proceedings: Biological Sciences, 246(1315):71–76.
- 738 Nunney, L. (1996). The influence of variation in female fecundity on effective population size.
739 Biological Journal of the Linnean Society, 59(4):411–425.
- 740 Pauly, D., Morgan, G. R., International Center for Living Aquatic Resources Management,
741 and Ma'had al-Kuwayt lil-Abḥāth al-'Ilmiyah, editors (1987). Length-Based Methods in
742 Fisheries Research. Number no. 325 in ICLARM Contribution. International Center for
743 Living Aquatic Resources Management ; Kuwait Institute for Scientific Research, Makati,
744 Metro Manila, Philippines : Safat, Kuwait.

- 745 Peart, C. R., Tusso, S., Pophaly, S. D., Botero-Castro, F., Wu, C.-C., Auriolles-Gamboa, D.,
746 Baird, A. B., Bickham, J. W., Forcada, J., Galimberti, F., Gemmell, N. J., Hoffman, J. I.,
747 Kovacs, K. M., Kunnsaranta, M., Lydersen, C., Nyman, T., de Oliveira, L. R., Orr, A. J.,
748 Sanvito, S., Valtonen, M., Shafer, A. B. A., and Wolf, J. B. W. (2020). Determinants of
749 genetic variation across eco-evolutionary scales in pinnipeds. Nature Ecology & Evolution,
750 pages 1–10.
- 751 Pinder, J. E., Wiener, J. G., and Smith, M. H. (1978). The Weibull Distribution: A New
752 Method of Summarizing Survivorship Data. Ecology, 59(1):175–179.
- 753 Pinsky, M. L. and Palumbi, S. R. (2014). Meta-analysis reveals lower genetic diversity in
754 overfished populations. Molecular Ecology, 23(1):29–39.
- 755 Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., der Auwera, G.
756 A. V., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault,
757 J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur,
758 D. G., and Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands
759 of samples. bioRxiv, page 201178.
- 760 Ricklefs, R. E. and Miller, G. L. (1999). Ecology. W.H.Freeman & Co Ltd, New York, 4th
761 edition edition.
- 762 Rolland, J., Schluter, D., and Romiguier, J. (2020). Vulnerability to Fishing and Life History
763 Traits Correlate with the Load of Deleterious Mutations in Teleosts. Molecular Biology and
764 Evolution.
- 765 Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari,
766 Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J. M., Nabholz, B., Roux, C.,
767 Tsagkogeorga, G., a. T. Weber, A., Weinert, L. A., Belkhir, K., Bierne, N., Glémin, S., and
768 Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants
769 of genetic diversity. Nature, 515(7526):261–263.
- 770 Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., and Lynch, M. (2012). Drift-barrier
771 hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences,
772 109(45):18488–18492.
- 773 Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J.,
774 Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B.,
775 Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., Canario, A.
776 V. M., and Reinhardt, R. (2014). European sea bass genome and its variation provide insights
777 into adaptation to euryhalinity and speciation. Nature Communications, 5:5770.
- 778 Tsikliras, A. C. and Stergiou, K. I. (2015). Age at maturity of Mediterranean marine fishes.
779 Mediterranean Marine Science, 16(1):5–20.
- 780 Väli, Ü., Einarsson, A., Waits, L., and Ellegren, H. (2008). To what extent do microsatellite
781 markers reflect genome-wide genetic diversity in natural populations? Molecular Ecology,
782 17(17):3808–3817.
- 783 Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J.,
784 and Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short
785 reads. Bioinformatics, 33(14):2202–2204.
- 786 Waples, R. S. (1991). Heterozygosity and Life-History Variation in Bony Fishes: An Alternative
787 View. Evolution, 45(5):1275–1280.

- 788 Waples, R. S. (2002). Evaluating the effect of stage-specific survivorship on the N_e/N ratio.
789 Molecular Ecology, 11(6):1029–1037.
- 790 Waples, R. S. (2016a). Life-history traits and effective population size in species with overlap-
791 ping generations revisited. Heredity, 117(4):241–250.
- 792 Waples, R. S. (2016b). Tiny estimates of the N_e/N ratio in marine fishes: Are they real?
793 Journal of Fish Biology, 89(6):2479–2504.
- 794 Waples, R. S., Do, C., and Chopelet, J. (2011). Calculating N_e and N_e/N in age-structured
795 populations: A hybrid Felsenstein-Hill approach. Ecology, 92(7):1513–1522.
- 796 Waples, R. S., Luikart, G., Faulkner, J. R., and Tallmon, D. A. (2013). Simple life-history
797 traits explain key effective population size ratios across diverse taxa. Proceedings of the
798 Royal Society B: Biological Sciences, 280(1768):20131339.
- 799 Waples, R. S., Mariani, S., and Benvenuto, C. (2018). Consequences of sex change for effective
800 population size. Proceedings of the Royal Society B: Biological Sciences, 285(1893):20181702.
- 801 White, E. P., Ernest, S. K. M., Kerkhoff, A. J., and Enquist, B. J. (2007). Relationships
802 between body size and abundance in ecology. Trends in Ecology & Evolution, 22(6):323–330.
- 803 Wright, S. (1969). The Theory of Gene Frequencies. Number Sewall Wright ; Vol. 2 in Evolution
804 and the Genetics of Populations. Univ. of Chicago Press, Chicago, Ill., paperback ed edition.
- 805 Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. R News,
806 2(3):7–10.

Legends to figure

Figure 1 : **Sampling and estimation of genetic diversity in 16 marine fish species** - In panels A, B and D, the geographical origin of samples is represented by colors. Atlantic: Bay of Biscay (dark blue), Faro region in Algarve (light blue). Mediterranean: Murcia region in Costa Calida (pink), Gulf of Lion (red). **(A)** Sampling map of all individuals included in this study. Each point represents the coordinates of a sample taken from one of four locations: two in the Atlantic Ocean and two in the Mediterranean Sea. **(B)** Genome-wide diversity in the European pilchard (*S. pilchardus*) and European sea bass (*D.labrax*) estimated after variant calling (orange triangle) or from GenomeScope (orange dot: median; smaller dots: individual estimates) **(C)** Heatmap clustering showing the variance in scaled genetic diversity within species among locations. Each line represents one species, with the corresponding species name written on the right side; every column represents one location. Blue and red colors respectively indicate lower and higher genetic diversity within a location for a given species compared to the average species genetic diversity. **(D)** Individual and median genetic diversity within each species estimated with GenomeScope. Species illustrations were retrieved from Iglésias (2013) with permissions.

Figure 2 : **Relationship between species median genetic diversity (%) and adult lifespan** - Each point represents the median of the individual genetic diversities for a given species. Adult lifespan is defined as the difference between lifespan and age at first maturity in years. Dot points and empty circles represent non-brooding species and brooding species, respectively. Blue and green lines represent the beta regression between adult lifespan and genetic diversity considering either the whole dataset (16 species), or the 11 non-brooding species only, respectively.

Figure 3 : **Variance in reproductive success induced by age-specific vital rates and adult lifespan correlate with observed genetic diversity** - On top, schematic illustration of age-specific fecundity (f_{age} , in orange) and survival ($S_{age \rightarrow age+1}$, blue) for a given species. (A) and (B) represents the relationship between observed genetic diversity on the y -axis and, respectively, $\frac{N_e}{N}$ estimated by AgeNe, and simulated genetic diversity with forward-in-time simulations in SLiM v.3.31 (?), on x -axis. Life tables containing information on age-specific survival, fecundity and lifespan were used for the 16 species. Age at maturity was used only with AgeNe. Dot points represent non-brooding species and empty circles, brooding species. Blue and green lines represent the beta regression between adult lifespan and genetic diversity considering the whole dataset (16 species), and the 11 non-brooding species only, respectively. The p - value and the pseudo- R^2 are represented on the top left for each of the two top panels for the non-brooders model. Panels (C)-(G) represent the relationship between scaled genetic diversity and scaled $\frac{N_e}{N}$ (i.e., divided by the maximum corresponding value) for the 11 non-brooding species. In each panel, the grey points represent scaled $\frac{N_e}{N}$ estimated from life tables including age at maturity, age-specific fecundity and survival and sex-specific differences (as in panel A). Black points are scaled estimates of $\frac{N_e}{N}$ from life tables with only: C) longevity (L), D) longevity (L) and age at maturity (AM), E) longevity (L) and age-specific survival (S), F) longevity (L), age at maturity (AM) and age-specific survival (S) and G) longevity (L) and age-specific fecundity (F). Beta regression models (grey and green lines) that closely overlap the red dotted line indicate that a decrease in $\frac{N_e}{N}$ leads to a similar decrease in genetic diversity.

Figure 4 : **Slope of the linear model between adult lifespan and $\frac{N_e}{N}$ ratio estimated with AgeNe for different combinations of age-specific survival and fecundity** - A) On top, gradient of survivorship curves simulated, ranging from type III (blue, $c < 1$) characterized by high juvenile mortality and low adult mortality; to type II (orange, c around 1), constant mortality and type I (red), low juvenile mortality high adult mortality. At the bottom, simulated fecundity either increases or decreases exponentially with age following $F_{Age} = exp^{f \times Age}$, with f ranging from -1 to 1. 16 simulated life tables were constructed with the same values of age at maturity and lifespan as the 16 studied species, and all possible survivorship curve and fecundity-age relationship shown in Panel A. B) Slope and R^2 of the regression between adult lifespan and $\frac{N_e}{N}$ ratio for the 16 simulated species as a function of c , for constant fecundity with age (thin line) and exponential increase of fecundity with age with $f = 0.142$ (thick line). C) Slope and D) R^2 of the regression between adult lifespan and $\frac{N_e}{N}$ ratio for the 16 simulated species for a gradient of values of c and f . In C), warmer colors indicate steeper negative slopes; in D) higher R^2 .