Life tables shape genetic diversity in marine fishes

Pierre Barry[1], Thomas Broquet[2], Pierre-Alexandre Gagnaire[1]

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

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

Abstract

Genetic diversity varies among species due to a range of eco-evolutionary processes that are not fully understood. The neutral theory predicts that the amount of variation in the genome sequence between different individuals of the same species should increase with its effective population size (N_e). In real populations, multiple factors that modulate the variance in reproductive success among individuals cause N_e to differ from the total number of individuals (N). Among these, age-specific mortality and fecundity rates are known to have a direct impact on the $\frac{N_e}{N}$ ratio. However, the extent to which vital rates account for differences in genetic diversity among species remains unknown. Here, we addressed this question by comparing genome-wide genetic diversity across 16 marine fish species with similar geographic distributions but contrasted lifespan and age-specific survivorship and fecundity curves. We sequenced the whole genome of 300 individuals to high coverage and assessed their genome-wide heterozygosity with a reference-free approach. Individual genome-wide heterozygosity varied from 0.2 to 1.4%, and adult lifespan was by far the most significant predictor of genetic diversity, with a large negative effect ($slope = -0.089$ per additionnal year of lifespan) that was further increased when brooding species providing intense parental care were removed from the dataset ($slope = -0.129$ per additionnal year of lifespan). Using published vital rates for each species, we showed that the $\frac{N_e}{N}$ ratio only generated by life tables predict the observed differences in genetic diversity among species. We further found that the extent of reduction in $\frac{N_e}{N}$ with increasing adult lifespan is particularly strong under Type III survivorship curves (high juvenile and low adult mortality) and increasing fecundity with age, which is typical of marine fish. Our study highlights the importance of vital rates in the evolution of genetic diversity within species in nature.

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

Author Summary

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

Introduction

Genetic diversity, the substrate for evolutionary change, is a key parameter for species' adaptability and vulnerability in conservation and management strategies (Frankham, 1995; Lande, 1995). Understanding the determinants of species' genetic diversity has been, however, a long standing puzzle in evolutionary biology (Lewontin, 1974). Advances in DNA sequencing technologies have allowed describing the range of genetic diversity levels across eukaroyte species and identifying main evolutionary processes underlying that variation (Leffler et al., 2012; Romiguier et al., 2014). But the extent and reasons for which life-history traits, and in particular reproductive strategies, influence genetic diversity remain to be clarified (Ellegren and Galtier, 2016).

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

Demographic variations set aside, the most proximate determinant of N_e is the actual number of individuals (N), also called the census population size. Comparative genomic studies in mammals and birds have showed that current species abundance correlates with the long-term coalescent N_e, despite potential deviation from long-term population stability in several of the species studied (Díez-Del-Molino et al., 2018; Leroy et al., 2020; Peart et al., 2020). General laws in ecology, such as the negative relationship between species abundance and body size (White et al., 2007) have also been used to predict the long-term N_e. Higher genetic diversity in small body size species was found in butterflies and Darwin's finches (Mackintosh et al., 2019; Brüniche-Olsen et al., 2019), while in the latter genetic diversity also positively correlated with island size, another potential proxy for the long-term N_e (Brüniche-Olsen et al., 2019). Suprisingly, however, genetic diversity variation across Metazoans is much better explained by fecundity and propagule size than classical predictors of species abundance such as body size and geographic range (Romiguier et al., 2014). This result has been attributed to differences among species with contrasted reproductive strategies in their long term probability of extinction. Under this hypothesis, species with low fecundity and large propagule size (K-strategists) would be more resilient to low population size episodes compared to species with high fecundity and small propagule size (r-stategists) that would go extinct if they reach such population sizes (Romiguier et al., 2014). By contrast, Mackintosh et al. (2019) found no effect of propagule size on genetic diversity within Papilionidae, a family showing little variation in reproductive strategy. Therefore, the major effect of the r/K gradient on genetic diversity variation across Metazoa probably hides other determinants that act within smaller branches of the tree of life. In particular, the extent and mechanisms by which the complex interplay between demographic and evolutionary processes influence genetic variation remains unclear.

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

configuration (Wright, 1969; Falconer, 1989; Lande and Barrowclough, 1987). A potentially strong effect comes from variance in the number of offspring per parent (V_k), which reduces N following $N_e = \frac{4N-4}{V_k+2}$ (Crow and Kimura, 1970). Variance in reproductive success can naturally emerge from particular age-specific demographic characterists summarized in life tables that contain age- (or stage-) specific survival and fecundity rates (Ricklefs and Miller, 1999). The impact of life tables characteristics on expected $\frac{N_e}{N}$ ratio has been the focus of a large body of theoretical and empirical works (Nunney, 1991, 1996; Waples, 2002, 2016b,a; Waples et al., 2018). Accounting for iteroparity and overlapping generations, a meta-analysis of vital rates in 63 species of plants and animals revealed that half of the variance in $\frac{N_e}{N}$ among species can be explained by just two life-history traits, adult lifespan and age at maturity (Waples et al., 2013). Despite this high predictive power of life tables, there is still no attempt to evaluate the extent to which lifetime variance in reproductive success explains differences in genetic diversity between species with different life table components.

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

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

Material and Methods

Sampling and DNA extraction

We sampled 16 marine teleostean fish species presenting a wide diversity of life-history strategies expected to affect genetic diversity (Table S2). All these species share broadly overlapping distributions across the northeastern Atlantic and Mediterranean regions. Sampling was performed at the same four locations for all species: two in the Atlantic (the Bay of Biscay in southwestern France or northwestern Spain and the Algarve in Portugal), and two in the western Mediterranean Sea (the Costa Calida region around Mar Menor in Spain and the Gulf of Lion in France see Fig 1A). For 12 of these species, 20 individuals were sampled (5 per location). For the 4 other species the total number of samples ranged from 10 to 19 (Table S2). Individuals were either sampled from landings in local fish markets, captured in the field (using hand nets, lure fishing, spearfishing or beach seines) or provided by collaborators. The majority of the sampling was done in 2018 and 2019. Whole-genomic DNA was extracted from fin or tissue clips stored in 95% ethanol using the NucleoSpin Tissue Kit (Macherey-Nagel), and treated with RNAse A to remove residual RNA. Double-stranded nucleic acid concentration was quantified using Qubit2.0 and standardized to 20ng per μl.

Whole-genome sequencing and reads quality control

Individual whole-genome sequencing libraries were prepared following the Illumina TruSeq DNA PCR-Free Protocol and sequenced by Genewiz Inc (USA). Libraries were quantified and multiplexed by groups of 40 individuals and sequenced on two S4 flow cells on a NovaSeq6000 instrument (Illumina) to generate 150 pb paired-end reads, targetting an average read depth of 20X per individual. Raw reads were preprocessed with `fastp` v.0.20.0 (Chen et al., 2018) using default parameters, allowing quality control, filtering by quality, length and complexity, and adapter trimming to be performed in a single step. Base correction was performed using quality comparison between overlapping bases of paired-end reads, and polyG tail trimming was enabled to correct for artefactual G repetitions occuring in Novaseq read tails.

Estimation of genetic diversity

We used `GenomeScope` v.1.0 to estimate individual genome-wide heterozygosity (Vurture et al., 2017). Briefly, this method uses a k-mers based statistical approach to infer overall genome characteristics, including total haploid genome size, percentage of repeat content and genetic diversity from unassembled short read sequencing data. Provided a sufficient average coverage depth (e.g. 20X), `GenomeScope` evaluates the fraction of heterozygous sites from the ratio of the height of the heterozygous to the homozygous k-mer peak, occuring at 50% (i.e. 10X) and 100% (i.e. 20X) of the average coverage depth, respectively. We used `jellyfish` v.2.2.10 to compute the k-mer profile of each individual (Marçais and Kingsford, 2011). The number of different possible k-mers (and thus the precision of the method) increases with k, but so does the runtime and the probability of "wrong" k-mers due to sequencing errors. We set $k = 21$ as recommended by `GenomeScope` and performed a sensitivity analysis by estimating genetic diversity and genome size for one individual of *D. labrax* using k from 17 to 25 (Fig S3). The genetic diversity of each species was determined as the median of the individual genome-wide heterozygosity values. We chose the median instead of the mean diversity since it is less sensitive to the possible presence of individuals with abnormal genetic diversity values (e.g. inbred or hybrid individuals) in our samples.

In order to assess the reliability of `GenomeScope` and detect potential systematic bias, we compared our results with high-standard estimates of genetic diversity obtained after read ali-

gnement against available reference genomes. To perform this test, we used the sea bass (*D. labrax*) and the European pilchard (*S. pilchardus*), two species that represent the lower and upper limits of the range of genetic diversity in our dataset (Table S2, Fig 1D). The 20 re-sequenced genomes for each of these two species were aligned with `bwa-mem` v.0.7.17 (Li and Durbin, 2009) to the reference genomes retrieved from Louro et al. (2019) and Tine et al. (2014) for *S. pilchardus* and *D. labrax*, respectively. We then removed PCR duplicates with the Picard tools `MarkDuplicates` v.2.23.2. We followed the best-practice pipeline in `GATK` v.4.1.6.0 for variant calling (Poplin et al., 2018): we ran `HaplotypeCaller` with default options to generate individual GVCFs files, stored them in a database with `GenomicsDBImport` and finally computed VCF files with `GenotypeGVCFs`. We didn't apply post variant calling filtering steps, such as hard filters on genotype quality scores or Hardy-Weinberg Equilibrium criterion, in order to avoid potential bias in the comparison of genetic diversity between species with very different rates of heterozygosity. However, we assume that possible bias due to the absence of variant filtering should not impact differences among individuals within each species. Finally, the VCF files generated with `GATK` were analyzed with `vcftools` v.0.1.17 (Danecek et al., 2011) to compute individual genome-wide heterozygosity.

Life history traits database

We collected seven simple quantitative variables describing various aspects of the biology and ecology of the 16 species: body size, trophic level, fecundity, propagule size, age at maturity, lifespan and adult lifespan (Table S2 for detailed informations on bibliographic references). There was substantial variability in the values that we found for some of these traits. This variation could have different causes including plasticity, selective pressures or different methodologies. Although we did not take into account this variability, we aimed to take the most representative values reported for each species and each trait, as described below. In addition, we collected two qualitative variables describing the presence/absence of hermaphroditism and brooding behaviour, as revealed by male-pouching of eggs (*H. guttulatus* and *S. typhle*) or nest-guarding (*C. galerita*, *S. cinereus* and *S. cantharus*). As growth is indeterminate in fish, we defined adult body size as the infinite length, L_{inf} determined by the Von Bertalanffy equation ($L_t = L_{inf}[1 - exp^{-K(t-t0)}]$), that links individual body size L_t to age t, with K a parameter defining the shape of this relationship (Pauly et al., 1987). We estimated adult body size as the median of all L_{inf} values reported for each species in the online database Fishbase (Froese et al., 2000). As L_{inf} was not documented in Fishbase for *D. puntazzo* and *C. galerita*, we took the median of the values reported in (Kraljević et al., 2007) and (Domínguez-Seoane et al., 2006) for *D. puntazzo*, and the maximum length observed in (Milton, 1983) for *C. galerita*. Trophic level was retrieved from Fishbase. Fecundity was defined as the absolute fecundity, i.e. the mean number of eggs in an ovary of a female in a single spawning event. Females may spawn several times during one reproductive season (Ganias et al., 2003; Murua and Motos, 2006), so absolute fecundity is not the value most directly relevant to global genetic diversity. However, it is the most commonly reported in the litterature as the number of spawnings events per reproductive season is difficult to measure. Because fecundity is proportional to individual body size, we computed fecundity at infinite length, L_{inf}. Propagule size was determined following Romiguier et al. (2014), as the size of the dispersal stage that becomes independent of the parents. For all species of this study, this corresponded to egg diameter, except for brooders, for which we used hatching size. All propagule size data were retrieved from species-specific references. Age at maturity was defined as the age at which 50% of the population is mature. Values for age at maturity were taken from Tsikliras and Stergiou (2015) for seven species while other values were retrieved from species-specific references. Likewise, lifespan values were taken from (Tsikliras and Stergiou, 2015) for six species and completed with specific references. Finally,

adult lifespan was defined as $Lifespan - Age\ at\ Maturity$ (Waples et al., 2013).

Construction of life tables

Life tables summarize survival rates and fecundities at each age during the life of an individual (Ricklefs and Miller, 1999). Thus, they provide detailed information on vital rates that influence the variance in lifetime reproductive success among individuals. This tool is well designed to describe population structure from the probability of survival to a specific age at which a specific number of offspring are produced. Ideally, age-specific survival is estimated by direct demographic measures, such as mark-recapture. Unfortunately, direct estimates of survival were not available for the 16 studied species. We thus followed Benvenuto et al. (2017) to construct species life tables. Age-specific mortality of species sp, $m_{sp,a}$, is a function of body length at age a, $L_{sp,a}$, asymptotic Von Bertalanffy length L_{inf}, and species Von Bertalanffy growth coefficient, K_{sp} :

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

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

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

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

Variance in reproductive success and the Ne/N ratio

To understand how differences in life tables drive differences in genetic diversity between species, we estimated the variance in lifetime reproductive success, V_k and the ensuing ratio $\frac{N_e}{N}$ using the analytic framework developped in `AgeNe` (Waples et al., 2011). `AgeNe` infers V_k using informations from life tables only. Hence, the variance in fecundity estimated is only generated by inter-individual differences in reproductive success and survival. `AgeNe` assumes constant population size, stable age structure, and no heritability of survival and fecundity. We used the life tables constructed as described above and set the number of new offsprings to 1000 per year. This setting is an arbitrary value which has no influence on the estimation of either V_k nor $\frac{N_e}{N}$ by `AgeNe`. For all species, we set the initial sex-ratio at 0.5 and equal contribution of individuals of the same age (i.e. no sweepstake reproductive success among same-age individuals). We ran `AgeNe` and estimated $\frac{N_e}{N}$ for each species. If genetic diversity differences among species are largely determined by differences in variance in reproductive success, we expect that the correlation between ratios of observed genetic diversity and ratios of $\frac{N_e}{N}$ is well fitted by the line of equation $y = x$. To test this prediction, we fitted a linear model between the two ratios and tested if the estimated slope and intercept are different from 1 and 0, respectively, with a t-test.

Four life-history traits can generate differences in $\frac{N_e}{N}$ between species: age at maturity, age-specific survival rates, age-fecundity relationships and sex-differences in these components. To

determine the role that each parameter plays in the differences in genetic diversity observed between species, we constructed alternative life tables where the effect of each parameter was removed one after the other. For example, to account for the effect of varying age at maturity on $\frac{N_e}{N}$, we constructed similar life tables as previously but with all species being mature at age 1. To test the effect of varying survival rates between ages, we followed Waples (2016b) and calculated constant survival rates in order to have 0.01 percent of individuals remaining at the maximum age. To test the effect of increasing fecundity with age, we set constant fecundity for all ages. Finally, to test the effect of sex-differences in life tables, we constructed identical life tables for both males and females. For each of the 16 alternative life tables, we then tested if the linear relationship between pairwise-ratios of observed genetic diversity and $\frac{N_e}{N}$ values are well-fitted by a model with slope equals to 1 and intercept 0, as previously.

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). Stochastic forward simulations allow a different formalisation compared to the deterministic model implemented in `AgeNe`. Thus, they provide another approach to the problem, which can be more intuitive to understand 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 litterature as desribed above. For each individual, the number of offspring produced per year was determined by a Poisson distribution with mean $\lambda_{S,a}$ specific to each species and each age. Age at first maturity was set to 1 for all simulations. Age and species-specific fecundity was determined as previously and scaled between 0 (age 0) and 100 (maximum age) within each species. In the simulations, each individual first reproduce and then either survive to the next year or die following a probability determined by its age and the corresponding life table. To keep population size constant in these non Wright-Fisher forwards simulations, we introduced a carrying capacity parameter, allowing population size to fluctuate around this capacity (Fig S15 - S22). We arbitrarilly set this parameter to $N = 2000$, and simulated non-recombining 1Mb loci with a mutation rate of $\mu = 1e^{-7}$. Each simulation was run for 25000 years, which was long enough for genetic diversity to reach mutation-drift equilibrium (Fig S23 - S30). For each simulation, we estimated the mean genetic diversity (i.e., the proportion of heterozygous sites along the 1Mb locus) over the last 10000 years after checking that an equilibrium has been reached. For each species, we ran 50 replicates and defined the genetic diversity predicted by a given simulation scenario as the mean genetic diversity at equilibrium averaged over the 50 replicates.

As previously, we evaluated the contribution of each component among 8 alternative life tables by testing the linear relationship between observed and simulated ratios of genetic diversity between pairs of species, and testing if the estimated slope and intercept differed from 1 and 0, respectively.

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 used 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 we introduced variation in survival and fecundity curves to a set of 16 species. First, age-specific mortality was simulated following

Pinder et al. (1978) :

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

where the value of 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 $-\frac{Lifespan}{log(0.01)^{1/c}}$ 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 $F_{Age} = exp^{f \times Age}$, as it is often observed in marine fishes (Curtis and Vincent, 2006). For all simulations, we scaled maximum fecundity to 1 and took values of f ranging from -1 to 1 (Fig 4A).

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 then explored the impact of alternative fecundity-age models on the relationship between adult lifespan and $\frac{N_e}{N}$ using three additional biologically realistic models: linear ($F_{Age} = a \times Age + b$), polynomial ($F_{Age} = [Age - AgeMat][(AgeMat + Lifespan - Age)^2]$, common in mammal) (Gage, 2001) and power-law ($F_{Age} = Age^f$). For the linear and the polynomial model, f describes the maximum fecundity at lifespan and age with the highest fecundity, respectively (i.e. higher absolute values of f correspond to higher differences in fecundity between low and high fecund ages for both models). f lied from -1 to 1 for the linear and the polynomial model. For the power-law model, we took values of f from -5 to 5.

Intraspecific variation in genetic diversity

We addressed the potential effects of population structure, demography and historical contingencies on genetic diversity by examining the extent of spatial variation in genetic diversity between the four populations within each species. First, we evaluated the relative amount of intraspecific compared to interspecific variation in genetic diversity. Then, we applied a z-transformation of individual genetic diversity within each species to put differences in species diversity on the same scale. We finally performed a hierarchical clustering analysis of the matrix of z-transformed genetic diversity values with `pheatmap` function available in `pheatmap v1.0.12` R package.

Statistical analyses

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

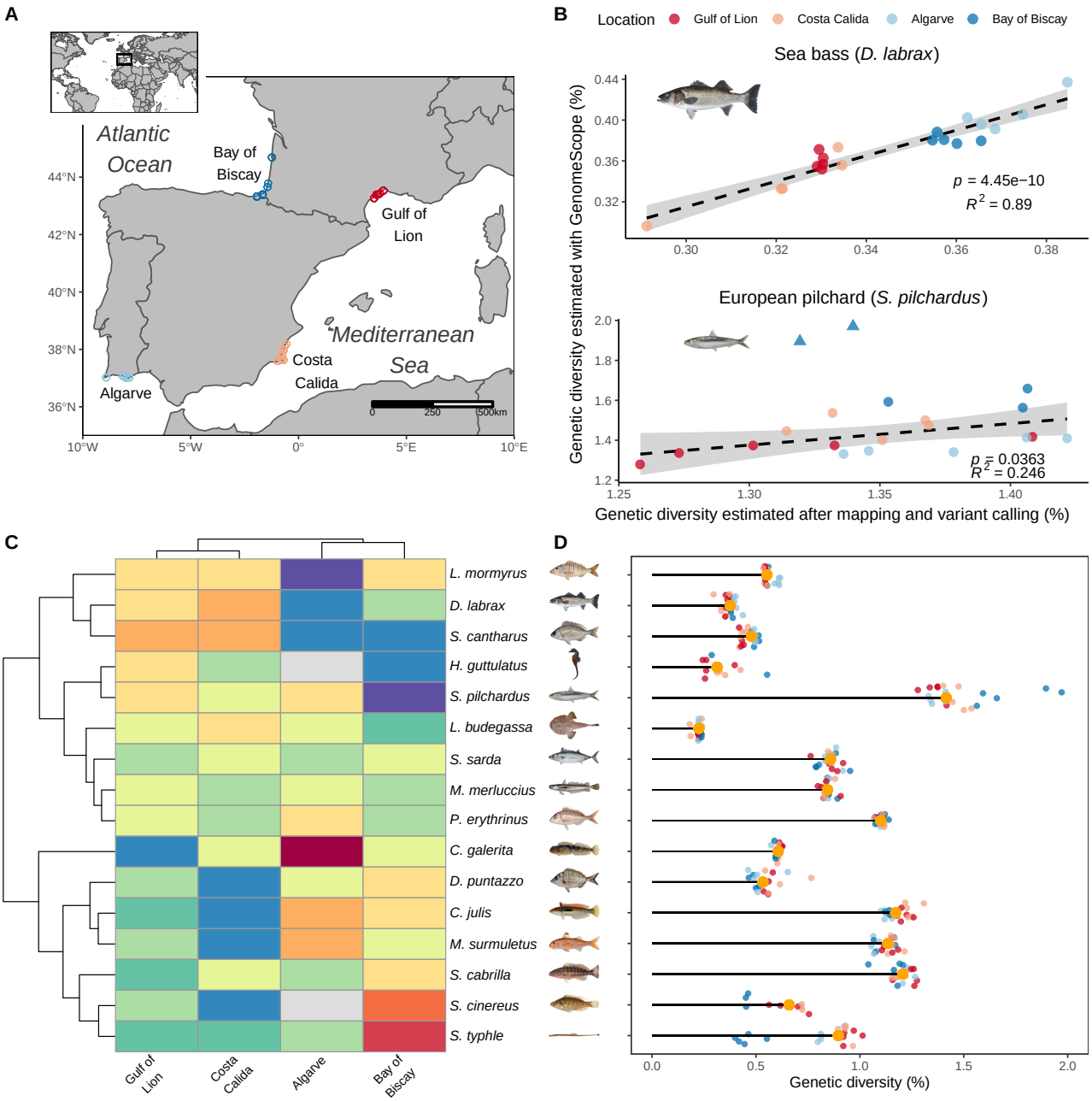

FIGURE 1 – **Sampling and estimation of genetic diversity in 16 marine fish species** - In panels A, B and D, the geographical origin of samples is represented by colors. Atlantic: Bay of Biscay (dark blue), Faro region in Algarve (light blue). Mediterranean: Murcia region in Costa Calida (pink), Gulf 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)** Relationship between individual mean genome-wide heterozygosity estimated with the k-mer based reference-free approach in GenomeScope (y-axis), and the high standard reference-based approach in GATK (x-axis), for european sea bass (*D. labrax*, top) and european pilchard (*S. pilchardus*, bottom, regression made on circle points only, see text). **(C)** Heatmap clustering showing the variance in 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 higher and lower genetic diversity within a location for a given species compared to the average species genetic diversity. **(D)** The range of individual genetic diversity within each species compared to the median genetic diversity represented with an orange dot. Species illustrations were retrieved from Iglésias (2013) with permissions.

Results

Whole-genome resequencing data set

We resequenced 300 individual genomes from 16 marine teleostean species, generating from 59.86×10^6 to 200.92×10^6 reads per individual (mean = 129×10^6, sd = 20×10^6, Fig S1). The read quality score (Q30 rate) ranged between 88% and 94% (mean = 92.4%, sd = 1.1) and the duplication rate lied between 5 and 15% (mean = 10.8%, sd = 2.6) (Fig S1). GC content was moderately variable among species and highly consistent among individuals of the same species, except for one individual of *S. cabrilla*, *D. puntazzo* and *M. surmuletus* that showed a marked discrepancy with the overall GC content of their species (Fig S1). These tree individuals were thus removed from downstream analyses to avoid potential issues due to contamination or poor sequencing quality (see discussion).

Estimation of genetic diversity with `GenomeScope`

The `GenomeScope` model successfully converged for all of the 297 individual genomes retained (Fig S4E). Estimated genome sizes were very consistent within species (Fig S4A-C). Estimated levels of genetic diversity were homogeneous within 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 sensiblity of `GenomeScope` regarding this parameter (Fig S3). 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 S1 ; Fig S2). Individual genetic diversity estimated with `GenomeScope` was significantly positively correlated to that estimated with the `GATK` reference-based variant calling approach for the two control species (Fig 1B). The geographic variance in genome-wide heterozygosity among individuals was very well captured for the sea bass (*D. labrax*, $R^2 = 89\%, p\text{-}value = 4.45e^{-10}$, Fig 1B) but less accurately for *S. pilchardus* ($R^2 = 24.6\%, p\text{-}value = 0.0363$, Fig 1B), in which two individuals suffered from biased estimates. Although this comparison indicated a possibly lower accuracy of `GenomeScope` in the presence of very high heterozygosity values, the range of heterozygosities estimated by `GenomeScope` was highly similar to that estimated with `GATK` for the two control species. Thus, the detection of fine-scale variation among individuals within two species lying at opposite ends of a diversity gradient confirms that `GenomeScope` is well-suited to quantify genetic diversity differences among the 16 species in our dataset.

Adult lifespan is the best predictor of genetic diversity

We evaluated the effect of several key life history traits that potentially affect species genetic diversity (Table 1).

Two widely used predictors of population size, body size and trophic level, were not significantly correlated to genetic diversity ($p\text{-}value = 0.119$ and 0.676 respectively, Fig S6A-B). Although we detected a significant negative relationship between the logarithm of fecundity

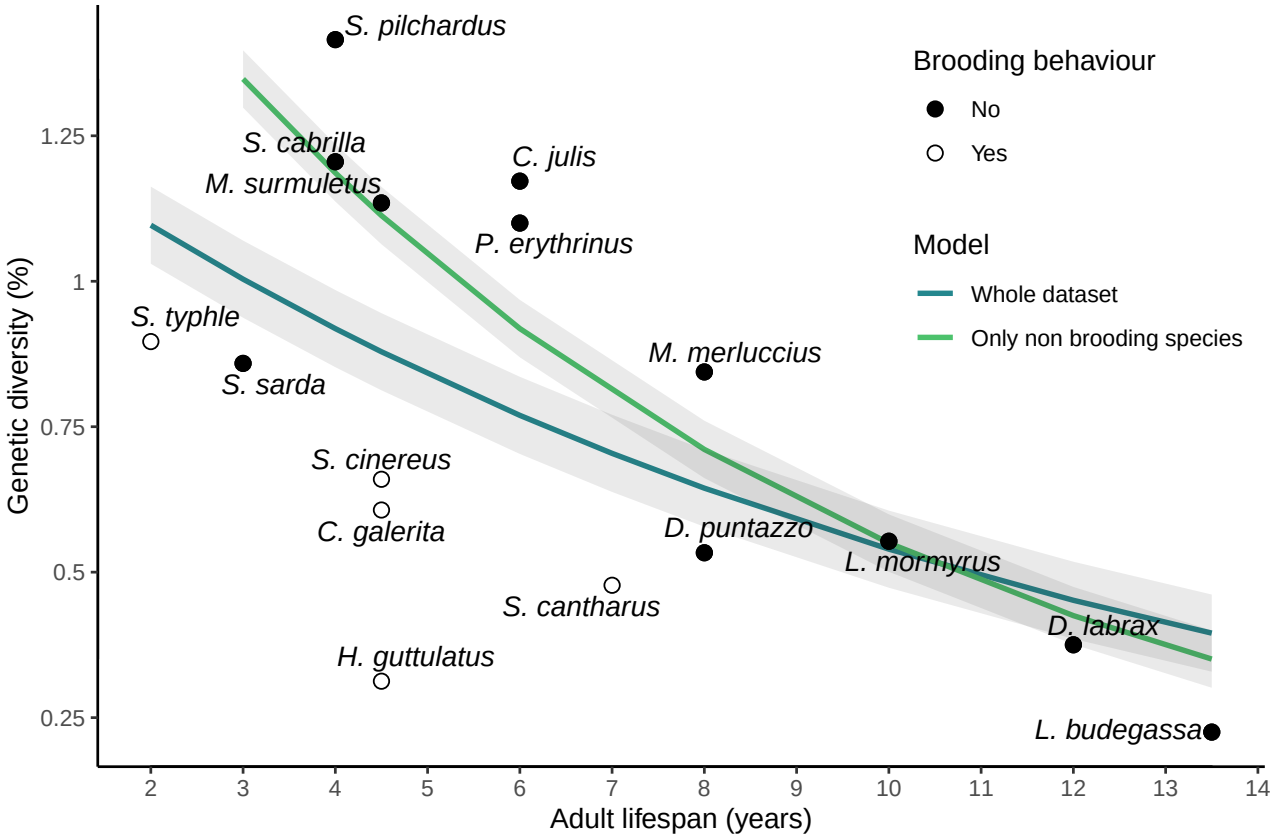

FIGURE 2 – **Relationship between species median genetic diversity (%) and adult lifespan** - Each point represents the median of observed genetic diversity among species' individuals samples. Adult lifespan is defined as the difference between lifespan and age at first maturity in years. Dot points represents non-brooding species, empty circles, brooding species. Dashed blue line and solid green line 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.

and propagule size ($p\text{-}value = 0.00131$, slope $= -0.4385 \pm 0.1076$) as in Romiguier et al. (2014), we found no significant correlation between either propagule size ($p\text{-}value = 0.561$), or the logarithm of fecundity ($p\text{-}value = 0.785$) and genetic diversity (Fig S6C-D).

By contrast, both lifespan ($p\text{-}value = 0.011$) and adult lifespan ($p\text{-}value = 0.007$) were significatively negatively correlated with genetic diversity (Table 1, Fig 2, Fig S6E). The percentage of variance explained by each variable reached 43.8 and 42.9 %, respectively.

We found no significant interaction between hermaphroditism and any of the previous variables on genetic diversity. By contrast, parental care showed a significant intercation with lifespan ($p\text{-}value = 0.0011$), adult lifespan ($p\text{-}value = 0.0008$) and body size ($p\text{-}value = 0.0035$) on genetic diversity. Brooding species (nest protection for *C. galerita*, *S. cinereus* and *S. cantharus* and male-pooch for *H. guttulatus* and *S. typhle*) had systematically lower genetic diversity than non-brooding species with similar lifespan.

When considering only non-brooding species, we found steeper negative correlations and higher percentages of between-species variance in genetic diversity explained by lifespan ($p\text{-}value = 1.017e^{-7}$, pseudo-$R^2 = 0.851$) and adult lifespan ($p\text{-}value = 1.645e^{-7}$, pseudo-$R^2 = 0.829$, Fig 1, Table 1). To test the relevance of considering this sub-dataset, we estimated the slope of the regression and the pseudo-R^2 for all combinations of 11 out of 16 species and compared the distribution of these values to the estimated slope and pseudo-R^2 obtained for the 11 non brooding species (Fig S11). The estimated slope for non brooders lied outside of the 95% confidence interval of the distribution of estimated slopes ($slope = -0.129, 95\%$ CI $= [-0.122, -0.049]$) and pseudo-R^2 (pseudo-$R^2 = 0.829, 95\%$ CI $= [0.073, 0.727]$). Furthermore, considering non-brooding species only, there was still no significant correlation between genetic diversity and trophic level ($p\text{-}value = 0.259$), propagule size ($p\text{-}value = 0.170$), and fecundity ($p\text{-}value = 0.390$), but genetic diversity appeared significantly correlated to body size ($p\text{-}value = 6.602e^{-5}$, pseudo-$R^2 = 0.616$). We did not detect any significant correlation between any trait variable and genetic diversity within the sub-dataset of brooding species. However, this should be taken with caution given the very low number of brooding species ($n = 5$) in our dataset.

Body size and lifespan were highly positively correlated traits in our dataset ($p\text{-}value = 0.00126, R^2 = 0.536$, Fig S5). Thus, using empirical observations only, it was not possible to fully disentangle the impact of each of these traits among the possible determinants of genetic diversity in marine fishes. However, we found important differences in effect sizes for body size ($slope = -0.014$), lifespan (-0.095) and adult lifespan (-0.129), which rule out body size as a major determinant of diversity in our dataset.

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 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 significatively 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 variance of $\frac{N_e}{N}$ 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 S12). 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).

Dataset	Predictor	*p-value*	Pseudo R^2	Slope estimate (± 95% interval)
Whole data set	Body size	0.119	0.192	$-0.006(-0.014; 0.002)$
	Trophic level	0.676	0.012	$-0.091(-0.524; 0.343)$
	Propagule size	0.562	0.015	$-0.014(-0.062; 0.034)$
	Fecundity	0.653	0.013	$-1.22e^{-5}(-6.63e^{-5}; 4.20e^{-5})$
	Lifespan	**0.0107**	0.438	$-0.062(-0.111; -0.013)$
	Adult lifespan	**0.0070**	0.429	$-0.089(-0.156; -0.023)$
	Hermaphroditism	0.434	0.034	$0.1779(-0.278; 0.633)$
	Parental Care	0.274	0.075	$-0.273(-0.772; 0.226)$
No parental care	Body size	$\mathbf{6.60e^{-5}}$	0.616	$-0.014(-0.021; -0.007)$
	Trophic level	0.256	0.093	$-0.326(-0.902; 0.251)$
	Propagule size	0.170	0.175	$-0.518(-1.273; 0.237)$
	Fecundity	0.390	0.056	$-2.51e^{-5}(-8.35e^{-5}; 3.33e^{-5})$
	Lifespan	$\mathbf{1.017e^{-7}}$	0.851	$-0.095(-0.131; -0.060)$
	Adult lifespan	$\mathbf{1.65e^{-7}}$	0.829	$-0.129(-0.179; -0.080)$
	Hermaphroditism	0.454	0.044	$0.206(-0.345; 0.757)$

TABLE 1 – **Statistical relationships between species genetic diversity and life history traits** - Genetic diversity was fitted to 6 quantitative (body size, trophic level, propagule size, fecundity, lifespan and adult lifespan) and two qualitative predictors (hermaphroditism and parental care) with a beta regression model using the `betareg` R package (Zeileis and Hothorn, 2002). In the upper part of the table, regressions were performed with the whole dataset, while in the lower part only the 11 non-brooding species were considered.

To determine whether life tables alone are able to reproduce the observed variability in genetic diversity among species, we compared the regression of pairwise-ratios of $\frac{N_e}{N}$ and observed genetic diversity to the linear model of slope 1 and intercept 0 (i.e., relative differences in $\frac{N_e}{N}$ equal relative differences in genetic diversity). Across all species, the slope of the linear model of the pairwise-ratios of genetic diversity against ratios of $\frac{N_e}{N}$ was not significatively different from 1 (estimated slope $= 0.73$; $sd = 0.31$; $p\text{-}value = 0.38$) but the intercept was different from 0 (estimated intercept $= 0.47$; $sd = 0.20$; $p\text{-}value = 0.016$, Fig 3C). After removing the 5 brooding species, the estimated model was closer to the model with slope equal to 1 (estimated slope $= 0.94$; $sd = 0.22$; $p\text{-}value = 0.80$) and intercept equal to 0 (estimated intercept $= 0.15$; $sd = 0.14$; $p\text{-}value = 0.30$, Fig 3D). Thus, the variance in reproductive success induced by life tables appeared quantitatively sufficient to explain the extent of variation in observed genetic diversity (Fig 3A) and especially so when excluding brooders.

Our next step was to determine the impact of each component of life tables on genetic diversity (Fig S7-S8). Starting from a null model (model 1, Fig S7A-S8A), 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 and maturity (model 2, Fig S7B-S8B) or age-specific survival (model 3, Fig S7C-S8C) did not better predict the ratio of observed genetic diversities. But combining only age at maturity and age-specific survival (model 4, Fig S7D-S8D) and adding age-specific fecundity (model 4-8, Fig S7D-G, Fig S8D-G) resulted in a better fit, which was further improved considering only species with no parental care behaviour. Finally, combining these tree parameters together (age at maturity, age-specific survival, and fecundity, model 8, Fig S7H-S8H) resulted in the best fit for both the slope and the intercept and for both non-brooding species and the whole data set. Adding sex-specific differences in life tables didn't improve the fit, however (models 9 to 16, Fig S7I-P and Fig S8I-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 a mutation rate $\mu = 1e^{-7}$. As expected, including age-specific vital rates decreased the equilibrium level of genetic diversity when compared with that of a Wright-Fisher scenario. Under Wright-Fisher conditions, genetic diversity was equal to 0.08% (in line with theoretical expectations : $\theta = 4N_e\mu$). 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). The pairwise-ratio of simulated genetic diversity was marginally correlated to pairwise-ratio of observed genetic diversity for all species (*p-value* = 0.0476, Fig 3E), but highly significantly correlated considering non-brooders species only (*p-value* = 0.00357, Fig 3F). Comparing the impact of each component of life tables on genetic diversity, we found similar results as previously: life tables with only age-specific survival and/or sex-specific differences could not generate the variance of observed genetic diversities (Fig S9A-B-E-F, Fig S10A-B-E-F), but simulated genetic diversity resulting from life tables with fecundity increasing with age fitted well observed genetic diversities (Fig S9C-D-G-H, Fig S10C-D-G-H).

Life tables drive correlation between lifespan and the Ne/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), under 5 models of fecundity-age relationships.

Considering fecundity constant with age, we found a significant relationship between adult lifesan and $\frac{N_e}{N}$ for species with type III survivorship curves ($c < 1$) but not for species having an age-specific survivorhip 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 empirical estimations for 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 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

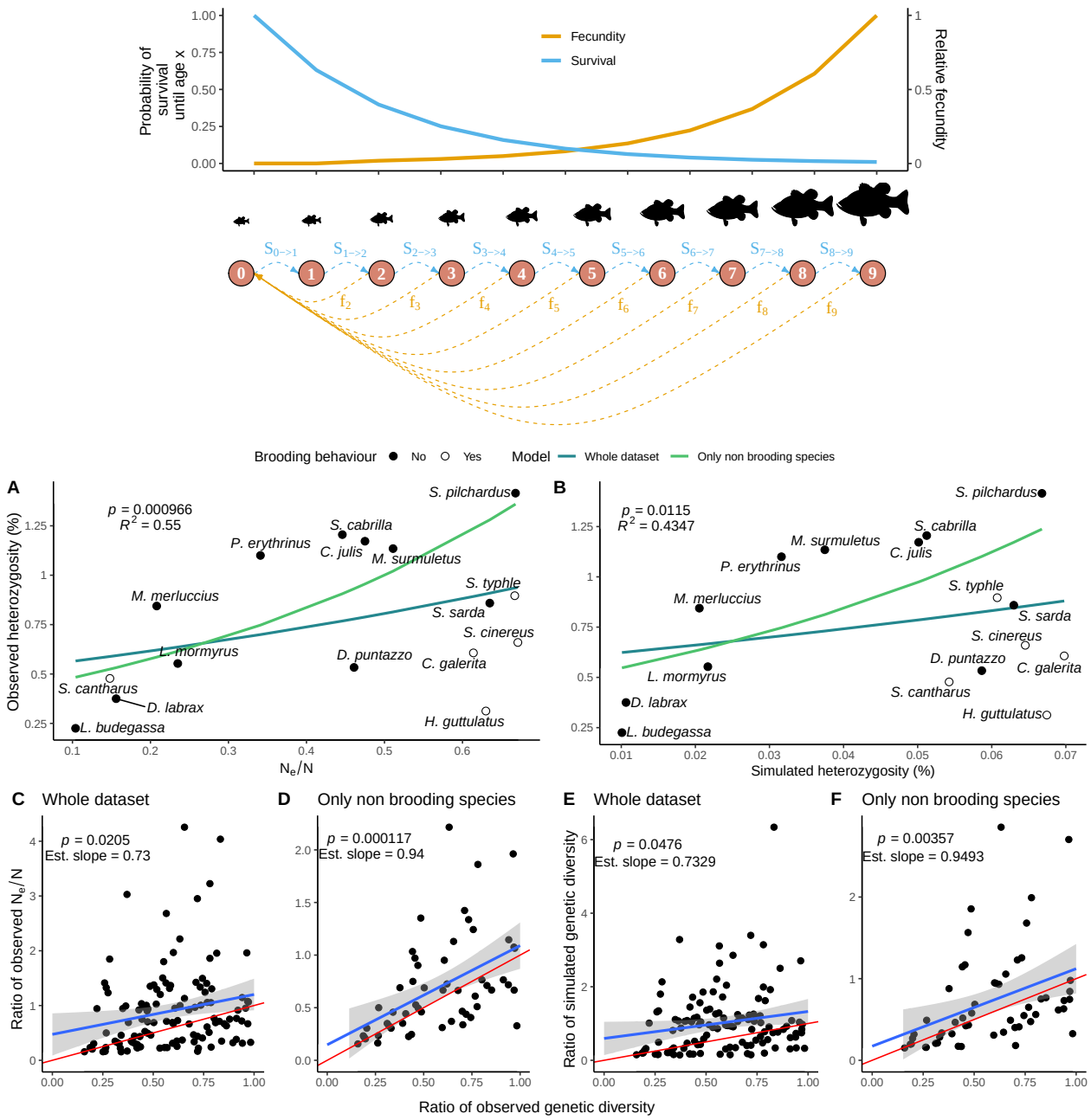

FIGURE 3 – **Variance in reproductive success induced by age-specific vital rates and adult lifespan correlate with observed genetic diversity** - On top, schematic illustration of age-specific fecundity (f_{age}, in blue) and survival ($S_{age->age+1}$, yellow) for a simulated species. (A) and (B) represents 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 (Haller and Messer, 2017), on x-axis. Life tables containing information on age-specific survival and 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), (D), (E) and (F) represent the relationships between ratio of observed genetic on x-axis and, respectively, ratio of $\frac{N_e}{N}$, for panels C and D, and ratio of simulated genetic diversity for panels E and F on y-axis. C and E represents the whole data set, D and F, only the non-brooding species. For each of the four bottom panels, the solid blue lines represents the estimated linear model and the red line, the model with slope 1 and intercept 0. The $p-value$ and slope estimation is represented on the top-left.

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 slighlty less steeper 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. In contrast, we found limited and 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.

Historical contingencies explain intra-specific differences in genetic diversity

The hierarchical clustering distributed the 16 species in two distincts clusters: the first one included 9 species with generally lower genetic diversity in Mediterranean than Atlantic localities, while the opposite was observed in the second cluster (7 species) (Fig 1C). The species of the second cluster are often found in coastal habitats, lagoons, estuaries whereas species of the first cluster are rather pelagic, epi-pelagic or benthic species. The only exception was the presence of *H. guttulatus* in the Atlantic cluster.

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 generationss.

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

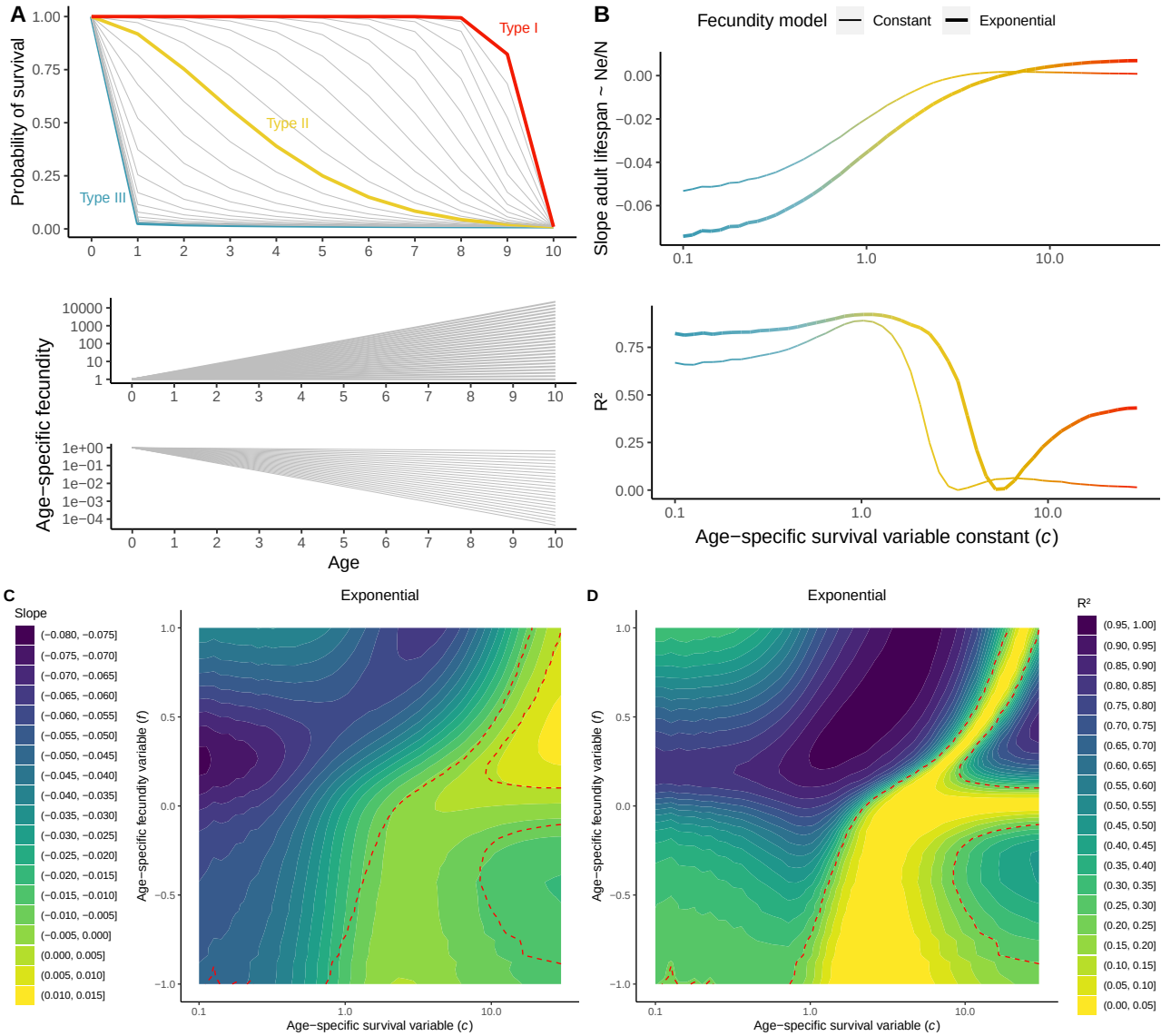

FIGURE 4 – **Slope of the linear model between adult lifespan and $\frac{N_e}{N}$ ratio estimated with `AgeNe` for different combinations of age-specific survival and fecundity** - A) On top, gradient of survivorship curves simulated, ranging from type III (blue, $c < 1$), high juvenile mortality and 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 as $F_{Age} = exp^{f \times Age}$, 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 with the corresponding survivorship curve and fecundity-age relationship. 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 $f = 0.14$ (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), colder colors indicate steeper slopes ; in D) higher R^2.

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 characterised 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 characterised 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 lower variance in reproductive success. Thus, in type III species, higher lifetime variance in reproductive success is expected as lifespan 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, $c > 2$).

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

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

Correlated effects

When relating measures of diversity with the estimates of $\frac{N_e}{N}$ derived from life tables, we did not account for differences in census sizes (N) among species. Population census sizes can be huge and are notoriously difficult to estimate in marine fishes. For that reason, abundance data remain largely unavailable for the 16 species of this study. We nevertheless expect long-lived species to have lower abundance compared to short-lived species, because in marine fishes N is generally negatively correlated to body size, which is itself positively correlated to adult lifespan in our dataset (Fig S5). Hence, while we have demonstrated here that variation in vital rates have a direct effect on long-term genetic diversity, the slope between adult lifespan and genetic diversity may be inflated by uncontrolled variation in N. Recent genome-wide comparative studies found negative correlations between $\frac{N_e}{N}$ and N in Pinnipeds (Peart et al., 2020) as well as between genetic diversity and body size in butterflies and birds (Mackintosh et al., 2019; Brüniche-Olsen et al., 2019). Here, a highly significant negative correlation was found between genetic diversity and body size and the strength of that correlation was comparable to that found in a meta-analysis of microsatellite diversity using catch data and body size as proxies

for fish abundance (Mccusker and Bentzen, 2010). We note, however, that body size was not as good a predictor of genetic diversity as lifespan and adult lifespan for the 11 non-brooding species and it was even not significant in the whole dataset of the 16 species (Table 1).

Another potentially confounding effect is the impact of r/K strategies which are the main determinant of genetic diversity across Metazoans (Romiguier et al., 2014). In our dataset, fecundity and propagule size (proxies for the r/K gradient) showed only little variance compared to their range of variation across Metazoans, and none of them were correlated to adult lifespan. However, we found that the 5 brooding species of our dataset, which are typical K-strategists, displayed lower genetic diversities with respect to their adult lifespan (Fig 2). Most interestingly, when these species were removed from the analysis, the effect of adult lifespan on genetic diversity was amplified, indicating a potentially confounding effect of parental care in marine fishes. Alternatively, low levels of genetic diversity in brooding species can also be explained by underestimated lifetime variance in reproductive success by `AgeNe` due unaccounted variance in reproductive succes within age-class. This effect could be high for species with strong sexual selection and mate choice (Hastings, 1988; Naud et al., 2009). Moreover, most of all these species inhabit lagoons and coastal habitats, corresponding to smaller ecological niches compared to species with no parental care, thus potentially resulting in lower long-term abundances. The discrepancy introduced by brooders in the relationship that we observed here between adult lifespan and genetic diversity may thus involve a variety of effects that remain to be elucidated.

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

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

Using one of the few direct estimates of the per-generation mutation rate in fish, Feng et al. (2017) explained the surprisingly low nucleotide diversity found in the Atlantic herring *Clupea harengus* ($\pi = 0.3\%$) by a very low mutation rate of 2×10^9 estimated from pedigree analysis. Although the herring is one of the most abundant and fecund pelagic species in the North Atlantic Ocean, its genetic diversity appears approximetaly 80% lower than that of

the european pilchard *S. pilchardus*, another member of the *Clupeidae* family that show the highest diversity in our study. Even if *C. harengus* has a larger body size (approximately 30 cm, compared to 20 cm for *S. pilchardus*, Froese et al. (2000)), it has above all a much longer lifespan (between 12 and 25 years) and a later age at maturity (between 2 and 6.5 years) (Jennings and Beverton, 1991). Considering the even lowest estimate of adult lifespan reported for the herring (10 years), the corresponding genetic diversity predicted by our model linking adult lifespan to genetic diversity would be 0.05 %, which is pretty close to the empirical estimate.

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

Conclusion

Here we used a simple approach to generate reference-free genome-wide estimates of diversity with k-mer analysis. Tested on two species with genetic diversities ranging from 0.22 to 1.42% the k-mer approach performed close to the level of a high-standard reference-based method in capturing fine-scale variation in diversity between evolutionary lineages and even populations of the same species. This opens the possibility to address the determinants of genetic diversity in other groups of taxa at limited costs withouth relying on existing genomics resources. Across Metazoans, the level of genetic diversity showed no significant relationship with the species' conservation status (Romiguier et al., 2014). Studies performed at lower phylogenetical scales such as in Darwin's finches and Pinnipeds, however, found reduced contemporary genetic diversity in threatened compared to non-threatened species (Brüniche-Olsen et al., 2019; Peart et al., 2020). Our results complement and extend this literature by showing the importance of taking into account life tables in comparisons of genetic diversity between species. Fish species exposed to overfishing have been showed to exhibit lower microsatellite diversity compared with other fish (Pinsky and Palumbi, 2014). By eliminating preferentially the oldest age classes from fish populations, overfishing may potentially modify life tables towards a decrease in the variance in reproductive success among individuals. Therefore, fishing may shift the mutation-drift balance in a counter-intuitive way that could mitigate the effect of decline in abundance on the loss of polymorphism over the long term.

Acknowledgments

The data used in this work were partly produced with the support of the GenSeq genotyping and sequencing platform, and bioinformatic data analysis benefited from the Montpellier Bioinformatics Biodiversity MBB platform, both platforms being supported by ANR program "Investissements d'avenir" (ANR-10-LABX-04-01). We would like to thank Rémy Dernat and Khalid Belkhir for their invaluable assistance in data storage, management and processing. We are grateful to the colleagues who provided us with samples as well as to those who facilitated or participated in sampling : F. Schlichta, T. Pastor, R. Castilho, R. Cunha, R. Lechuga, D. Pilo, C. Mena, J. Charton, T. Robinet, A. Darnaude, S. Vaz, M. Duranton, N. Bierne, S. Villéger, S. Blouet, as well as the fishermen and employees of fish markets and fish auctions. This work was supported by the ANR grant CoGeDiv ANR-17-CE02-0006-01.

Author contributions

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

Data archiving

Data and scripts used in this study are freely available in the GitHub repository `https://github.com/pierrebarry/life_tables_genetic_diversity_marine_fishes`. All sampling metadata are accessible under GEOME at the CoGeDiv Project Homepage : `https://geome-db.org/workbench/project-overview?projectId=357`. Sequence reads have been deposited in the GenBank Sequence Read Archive under the accession code BioProject PRJNAXXXX.

Références

Benvenuto, C., Coscia, I., Chopelet, J., Sala-Bozano, M., and Mariani, S. (2017). Ecological and evolutionary consequences of alternative sex-change pathways in fish. *Scientific Reports*, 7(1) :9084.

Brüniche-Olsen, A., Kellner, K. F., and DeWoody, J. A. (2019). Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. *Molecular Ecology*, 28(22) :4914–4925.

Chen, J., Glémin, S., and Lascoux, M. (2017). Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Molecular Biology and Evolution*, 34(6) :1417–1428.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp : An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17) :i884–i890.

Cribari-Neto, F. and Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(1) :1–24.

Crow, J. F. and Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Harper & Row.

Curtis, J. M. R. and Vincent, A. C. J. (2006). Life history of an unusual marine fish : Survival, growth and movement patterns of Hippocampus guttulatus Cuvier 1829. *Journal of Fish Biology*, 68(3) :707–733.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., and Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15) :2156–2158.

Díez-Del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., and Dalén, L. (2018). Quantifying Temporal Genomic Erosion in Endangered Species. *Trends in Ecology & Evolution*, 33(3) :176–185.

Domínguez-Seoane, R., Pajuelo, J. G., Lorenzo, J. M., and Ramos, A. G. (2006). Age and growth of the sharpsnout seabream Diplodus puntazzo (Cetti, 1777) inhabiting the Canarian archipelago, estimated by reading otoliths and by backcalculation. *Fisheries Research*, 81(2) :142–148.

Ellegren, H. and Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7) :422–433.

Falconer, D. S. (1989). *Introduction to Quantitative Genetics*. Longman, Scientific & Technical ; Wiley, Burnt Mill, Harlow, Essex, England : New York, 3rd ed edition.

Feng, C., Pettersson, M., Lamichhaney, S., Rubin, C.-J., Rafati, N., Casini, M., Folkvord, A., and Andersson, L. (2017). Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. *eLife*, 6 :e23907.

Frankham, R. (1995). Conservation Genetics. *Annual Review of Genetics*, 29(1) :305–327.

Froese, R., Pauly, D., and Editors (2000). FishBase 2000 : Concepts, design and data sources. page 344.

761 Gage, T. B. (2001). Age-specific fecundity of mammalian populations : A test of three mathe-
762 matical models. *Zoo Biology*, 20(6) :487–499.

763 Ganias, K., Somarakis, S., Machias, A., and Theodorou, A. J. (2003). Evaluation of spaw-
764 ning frequency in a Mediterranean sardine population (Sardina pilchardus sardina). *Marine*
765 *Biology*, 142(6) :1169–1179.

766 Haller, B. C. and Messer, P. W. (2017). SLiM 2 : Flexible, Interactive Forward Genetic Simu-
767 lations. *Molecular Biology and Evolution*, 34(1) :230–240.

768 Hastings, P. A. (1988). Female choice and male reproductive success in the angel blenny,
769 Coralliozetus angelica (Teleostei : Chaenopsidae). *Animal Behaviour*, 36(1) :115–124.

770 Hauser, L. and Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics : Ugly
771 hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4) :333–362.

772 Hedgecock, D. (1994). Does variance in reproductive success limit effective population sizes of
773 marine organisms ? In A. *Genetics and Evolution of Aquatic Organisms*.

774 Hedgecock, D. and Pudovkin, A. I. (2011). Sweepstakes Reproductive Success in Highly Fe-
775 cund Marine Fish and Shellfish : A Review and Commentary. *Bulletin of Marine Science*,
776 87(4) :971–1002.

777 Hedrick, P. (2005). Large variance in reproductive success and the Ne/N ratio. *Evolution*,
778 59(7) :1596–1599.

779 Iglésias, S. (2013). *Actinopterygians from the North-Eastern Atlantic and the Mediterranean (A*
780 *Natural Classification Based on Collection Specimens, with DNA Barcodes and Standardized*
781 *Photographs), Volume I (Plates), Provisional Version 09*.

782 Jenkins, T. L., Castilho, R., and Stevens, J. R. (2018). Meta-analysis of northeast Atlantic
783 marine taxa shows contrasting phylogeographic patterns following post-LGM expansions.
784 *PeerJ*, 6 :e5684.

785 Jennings, S. and Beverton, R. J. H. (1991). Intraspecific variation in the life history tactics of
786 Atlantic herring (Clupea harengus L.) stocks. *ICES Journal of Marine Science*, 48(1) :117–
787 125.

788 Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press,
789 Cambridge.

790 Kimura, M. and Crow, J. F. (1964). The Number of Alleles That Can Be Maintained in a
791 Finite Population. *Genetics*, 49(4) :725–738.

792 Kraljević, M., Matić-Skoko, S., Dul\vcić, J., Pallaoro, A., Jardas, I., and Glamuzina, B. (2007).
793 Age and growth of sharpsnout seabream Diplodus puntazzo (Cetti, 1777) in the eastern
794 Adriatic Sea.

795 Lande, R. (1995). Mutation and Conservation. *Conservation Biology*, 9(4) :782–791.

796 Lande, R. and Barrowclough, G. F. (1987). Effective population size, genetic variation, and their
797 use in population management. In Soulé, M. E., editor, *Viable Populations for Conservation*,
798 pages 87–124. Cambridge University Press, Cambridge.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., and Przeworski, M. (2012). Revisiting an Old Riddle : What Determines Genetic Diversity Levels within Species ? *PLOS Biology*, 10(9) :e1001388.

Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A., Scornavacca, C., Carrasco, M. R., Fuchs, J., Illera, J. C., Swardt, D. H. D., Thébaud, C., Milà, B., and Nabholz, B. (2020). Endemic island songbirds as windows into evolution in small effective population sizes. *bioRxiv*, page 2020.04.07.030155.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. Number 25 in Columbia Biological Series. Columbia Univ. Pr, New York.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14) :1754–1760.

Louro, B., De Moro, G., Garcia, C., Cox, C. J., Veríssimo, A., Sabatino, S. J., Santos, A. M., and Canário, A. V. M. (2019). A haplotype-resolved draft genome of the European sardine (Sardina pilchardus). *GigaScience*, 8(5).

Mackintosh, A., Laetsch, D. R., Hayward, A., Charlesworth, B., Waterfall, M., Vila, R., and Lohse, K. (2019). The determinants of genetic diversity in butterflies. *Nature Communications*, 10(1) :1–9.

Marçais, G. and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6) :764–770.

Martinez, A. S., Willoughby, J. R., and Christie, M. R. (2018). Genetic diversity in fishes is influenced by habitat type and life-history variation. *Ecology and Evolution*, 8(23) :12022–12031.

Mccusker, M. R. and Bentzen, P. (2010). Positive relationships between genetic diversity and abundance in fishes. *Molecular Ecology*, 19(22) :4852–4862.

Milton, P. (1983). Biology of littoral blenniid fishes on the coast of south-west England. *Journal of the Marine Biological Association of the United Kingdom*, 63(1) :223–237.

Murua, H. and Motos, L. (2006). Reproductive strategy and spawning activity of the European hake Merluccius merluccius (L.) in the Bay of Biscay. *Journal of Fish Biology*, 69(5) :1288–1303.

Naud, M.-J., Curtis, J. M. R., Woodall, L. C., and Gaspar, M. B. (2009). Mate choice, operational sex ratio, and social promiscuity in a wild population of the long-snouted seahorse Hippocampus guttulatus. *Behavioral Ecology*, 20(1) :160–164.

Nei, M., Maruyama, T., and Chakraborty, R. (1975). The Bottleneck Effect and Genetic Variability in Populations. *Evolution*, 29(1) :1.

Nunney, L. (1991). The Influence of Age Structure and Fecundity on Effective Population Size. *Proceedings : Biological Sciences*, 246(1315) :71–76.

Nunney, L. (1996). The influence of variation in female fecundity on effective population size. *Biological Journal of the Linnean Society*, 59(4) :411–425.

Pauly, D., Morgan, G. R., International Center for Living Aquatic Resources Management, and Ma'had al-Kuwayt lil-Abḥāth al-'Ilmīyah, editors (1987). *Length-Based Methods in Fisheries Research*. Number no. 325 in ICLARM Contribution. International Center for Living Aquatic Resources Management ; Kuwait Institute for Scientific Research, Makati, Metro Manila, Philippines : Safat, Kuwait.

Peart, C. R., Tusso, S., Pophaly, S. D., Botero-Castro, F., Wu, C.-C., Aurioles-Gamboa, D., Baird, A. B., Bickham, J. W., Forcada, J., Galimberti, F., Gemmell, N. J., Hoffman, J. I., Kovacs, K. M., Kunnasranta, M., Lydersen, C., Nyman, T., de Oliveira, L. R., Orr, A. J., Sanvito, S., Valtonen, M., Shafer, A. B. A., and Wolf, J. B. W. (2020). Determinants of genetic variation across eco-evolutionary scales in pinnipeds. *Nature Ecology & Evolution*, pages 1–10.

Pinder, J. E., Wiener, J. G., and Smith, M. H. (1978). The Weibull Distribution : A New Method of Summarizing Survivorship Data. *Ecology*, 59(1) :175–179.

Pinsky, M. L. and Palumbi, S. R. (2014). Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, 23(1) :29–39.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., der Auwera, G. A. V., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., and Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, page 201178.

Ricklefs, R. E. and Miller, G. L. (1999). *Ecology*. W.H.Freeman & Co Ltd, New York, 4th edition edition.

Rolland, J., Schluter, D., and Romiguier, J. (2020). Vulnerability to Fishing and Life History Traits Correlate with the Load of Deleterious Mutations in Teleosts. *Molecular Biology and Evolution*.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J. M., Nabholz, B., Roux, C., Tsagkogeorga, G., a. T. Weber, A., Weinert, L. A., Belkhir, K., Bierne, N., Glémin, S., and Galtier, N. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515(7526) :261–263.

Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., and Lynch, M. (2012). Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences*, 109(45) :18488–18492.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., Canario, A. V. M., and Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5 :5770.

Tsikliras, A. C. and Stergiou, K. I. (2015). Age at maturity of Mediterranean marine fishes. *Mediterranean Marine Science*, 16(1) :5–20.

Väli, Ü., Einarsson, A., Waits, L., and Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations ? *Molecular Ecology*, 17(17) :3808–3817.

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., and Schatz, M. C. (2017). GenomeScope : Fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14) :2202–2204.

Waples, R. S. (1991). Heterozygosity and Life-History Variation in Bony Fishes : An Alternative View. *Evolution*, 45(5) :1275–1280.

Waples, R. S. (2002). Evaluating the effect of stage-specific survivorship on the Ne/N ratio. *Molecular Ecology*, 11(6) :1029–1037.

Waples, R. S. (2016a). Life-history traits and effective population size in species with overlapping generations revisited. *Heredity*, 117(4) :241–250.

Waples, R. S. (2016b). Tiny estimates of the Ne/N ratio in marine fishes : Are they real ? *Journal of Fish Biology*, 89(6) :2479–2504.

Waples, R. S., Do, C., and Chopelet, J. (2011). Calculating Ne and Ne/N in age-structured populations : A hybrid Felsenstein-Hill approach. *Ecology*, 92(7) :1513–1522.

Waples, R. S., Luikart, G., Faulkner, J. R., and Tallmon, D. A. (2013). Simple life-history traits explain key effective population size ratios across diverse taxa. *Proceedings of the Royal Society B : Biological Sciences*, 280(1768).

Waples, R. S., Mariani, S., and Benvenuto, C. (2018). Consequences of sex change for effective population size. *Proceedings of the Royal Society B : Biological Sciences*, 285(1893) :20181702.

White, E. P., Ernest, S. K. M., Kerkhoff, A. J., and Enquist, B. J. (2007). Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution*, 22(6) :323–330.

Wright, S. (1969). *The Theory of Gene Frequencies*. Number Sewall Wright ; Vol. 2 in Evolution and the Genetics of Populations. Univ. of Chicago Press, Chicago, Ill., paperback ed edition.

Zeileis, A. and Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, 2(3) :7–10.