

1 What is the best fitness measure in wild  
2 populations? A case study on the power of  
3 short-term fitness proxies to predict  
4 reproductive value

5

6

7 Vita Živa Alif<sup>1\*</sup>, Jamie Dunning<sup>1</sup>, Heung Ying Janet Chik<sup>2</sup>, Terry Burke<sup>3</sup>, Julia Schroeder<sup>1</sup>

8

9 1 – Department of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire, UK

10 2 – Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, the  
11 Netherlands

12 3 –Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

13

14

15 \*Corresponding author email: [ziva.alif@gmail.com](mailto:ziva.alif@gmail.com)

16

17

## 18 Abstract

19 Fitness is at the core of evolutionary theory, but it is difficult to measure accurately. One way to measure  
20 long-term fitness is by calculating the individual's reproductive value, which represents the expected  
21 number of allele copies an individual passes on to distant future generations. However, this metric of  
22 fitness is scarcely used because the estimation of individual's reproductive value requires long-term  
23 pedigree data, which is rarely available in wild populations where following individuals from birth to  
24 death is often impossible. Wild study systems therefore use short-term fitness metrics as proxies, such  
25 as the number of offspring produced. This study obtained three frequently used short-term proxies for  
26 fitness obtained at different offspring life stages (eggs, hatchlings, fledglings and recruits), and  
27 compared their ability to predict reproductive values derived from the genetic pedigree of a wild  
28 passerine bird population. We used twenty years of precise field observations and a near-complete  
29 genetic pedigree to calculate reproductive success, individual growth rate and de-lifed fitness as lifetime  
30 fitness measures, and as annual de-lifed fitness. We compared the power of these metrics to predict  
31 reproductive values and lineage survival to the end of the study period. The three short-term fitness  
32 proxies predict the reproductive values and lineage survival only when measured at the recruit stage.  
33 There were no significant differences between the different fitness proxies at the same offspring stages  
34 in predicting the reproductive values and lineage survival. Annual fitness at one year old predicted  
35 reproductive values equally well as lifetime de-lifed fitness. However, none of the short-term fitness  
36 proxies was strongly associated with the reproductive values. In summary, the commonly short-term  
37 fitness proxies capture long-term fitness with intermediate accuracy at best, if measured at recruitment  
38 stage. As lifetime fitness measured at recruit stage and annual fitness in the first year of life were the  
39 best proxies of long-term fitness, we encourage their future use.

40

41 Keywords: de-lifing, reproductive values, fitness, individual growth rate, lifetime reproductive success

## 42 Introduction

43 The concept of fitness is central to evolutionary theory (1). Natural selection maximises fitness, which  
44 is therefore a driving force of evolution as well as a measure of evolutionary success (2). One definition  
45 of fitness is how good an individual is at spreading its genes into future generations, relative to all other  
46 individuals in the population (2). A universal definition of fitness in mathematical terms that applies to  
47 all population structures and dynamics is however not agreed on (2–5).

48 Ecological studies measure fitness in diverse ways, often depending on the research question, the  
49 population dynamics, and the ecology of the study species (6,7). While some studies measure fitness  
50 across lifetimes, other studies measure individual annual fitness to examine variation in selection  
51 between years (8). Lifetime fitness is considered more accurate than annual measures, as the latter is  
52 influenced by environmental stochasticity (7,9). Alternative fitness measures have been developed that  
53 account for environmental stochasticity and population dynamics (5,10–12).

54 Some fitness metrics include both survival and fecundity components (8,13), while others focus on only  
55 one component as a proxy, such as lifespan (4), or on only a single life-history trait, such as the age at  
56 first reproduction (13,14). The two most commonly used individual fitness proxies are lifetime  
57 reproductive success (LRS) (15) and individual growth rate (IGR) (13). Both count the number of  
58 offspring produced in the individual's lifetime, but IGR gives more weight to offspring produced at a  
59 younger age (13), therefore results differ (16). As a consequence, different fitness proxies do not  
60 necessarily correlate well (7) and more research is needed to determine which is the most appropriate  
61 measures of fitness (6). Choosing the appropriate fitness proxy is therefore an important consideration  
62 when designing a study (7).

63 The age of an offspring at the time it is included in the fitness measurement of the parent must also be  
64 considered, particularly given that studies count offspring at varying ages or life-history stages.  
65 Offspring survival can be both a part of offspring fitness due to its unique genotype, or a part of parental  
66 fitness in cases where parental phenotype affects offspring mortality, for example through parental care  
67 (17). Counting offspring at higher stages of development assigns a part of offspring fitness into parental

68 fitness, thus potentially affecting the strength and direction of selection. Furthermore, Brommer et al.  
69 (6) show that the age of offspring census directly affects IGR values, while it does not affect LRS,  
70 making the two fitness proxies less comparable at different stages. It is therefore important to  
71 understand what census time best captures parental fitness.

72 Although fitness is considered to be a measure of an individual's gene copy frequency in future  
73 generations, most fitness proxies focus on an individual's direct descendants. Alternatively, the  
74 reproductive value from a single individual, defined as the expected number of copies of each of an  
75 individual's alleles in a future generation conditional on a realised pedigree of descendants, can be used  
76 to measure long-term fitness (18). The reproductive values can be estimated from a genetic pedigree,  
77 following rules of mendelian inheritance to calculate how many allele copies survive on average. The  
78 reproductive values stabilise after  $\log_2 N$  generations, where  $N$  is the population size (4,18,19). While  
79 the ultimate genetic contribution of an individual will only emerge over long timescales ( $>100$   
80 generations), the reproductive values are determined in  $\sim 10$  generations and are a good predictor of the  
81 ultimate genetic contribution (18).

82 The reproductive values closely predict allele survival probability, but not their frequencies (18). Due  
83 to recombination and segregation in meiosis, the actual genetic frequencies, conditional on allele  
84 survival, instead follow a random distribution (18–21). Consequently, not all genealogical ancestors are  
85 also genetic ancestors (22). Despite the difference between actual allele frequencies and the  
86 reproductive values, reproductive value is a practical and relevant measure for evolutionary studies as  
87 it is maximised by natural selection, thus closely corresponding to fitness (18,23).

88 This study examined the correlation between several short-term fitness proxies and reproductive values.  
89 We used data from an isolated house sparrow (*Passer domesticus*) population on Lundy Island (United  
90 Kingdom) with 20 years of life history data, unusually precise measures of survival and reproductive  
91 success, and nearly complete genetic pedigree information (24). We examined the two most commonly  
92 used individual fitness proxies based on fecundity (25): lifetime reproductive success (LRS) and  
93 individual growth rate (IGR) (13). We measured both at four different offspring stages (eggs,  
94 hatchlings, fledglings, and recruits) to investigate which are most accurate. We also used a short-term

- 95 fitness proxy that incorporates survival – de-lifed fitness (8). This is based on individual offspring  
96 production and survival adjusted for population growth.

## 97 Methods

### 98 Study system

99 The house sparrow population on Lundy Island has been continuously monitored since 2000. Lundy is  
100 a small island 19 km off the south-west English coast (51°11N, 4°40W). In 2000, 50 individuals were  
101 brought to Lundy from the mainland for an experiment (26). Due to the distance of the island from the  
102 mainland and the sedentary nature of sparrows, there is minimal dispersal to and from the island (24).  
103 The sparrow population size has fluctuated between 166 and 1242 individuals (juveniles included)  
104 between 1999 and 2019 (Fig 1A).

105 During systematic annual monitoring, each sparrow is ringed with three colour rings and one metal ring  
106 from the British Trust for Ornithology. Since most sparrows are initially caught as nestlings and ringed  
107 as fledglings, we know the identities of the parents attending their nests, and the exact age of all  
108 individuals (27). Over 99% of the population has been ringed since 2000 (27). If an individual is not  
109 seen for two years or more, it is assumed dead, with this assumption based on previous mark–recapture  
110 success data (27,28). Blood samples are collected upon bird capture and genotyped at up to 23  
111 microsatellite loci (24). This allows for the assignment of genetic parentage with 95% confidence (24).  
112 From the genetic pedigree and the social brood information, the reproductive success of individuals is  
113 calculated. Thanks to these data, the study system provides unusually accurate survival, reproduction,  
114 and pedigree data for the complete population (24).

### 115 Pedigree analysis

116 We calculated fitness proxies and the reproductive values for founders and half-founders that were born  
117 between 1999 and 2002, the starting years of the long-term study. Founders and half-founders were  
118 defined as individuals for which both parents, or one parent, respectively, were unknown. To calculate  
119 reproductive values, we used our genetic pedigree containing all reproducing individuals up to 2018.  
120 We removed any individuals from cohorts after 2002 that had at least one unknown parent; thus 8% of  
121 all individuals in the pedigree were removed.

122 Reproductive values were calculated using gene dropping (29). Gene dropping is a computer simulation  
123 in which each individual is assigned two alleles (one paternal and one maternal), and their Mendelian  
124 transmission down the pedigree is simulated. By repeating this simulation many times and calculating  
125 the mean values, robust estimates of reproductive values can be obtained by examining the frequency  
126 of an individual's alleles in subsequent generations. In addition, the allele survival probability can be  
127 calculated by examining in how many simulations the allele survives in present-day individuals. We  
128 ran the simulation 10,000 times using R package *nadiv* (30). Using the results, we derived lineage  
129 longevity, reproductive values, and allele survival probability. We define lineage longevity as the  
130 number of years before a lineage originating from one individual goes extinct, and allele survival  
131 probability is the proportion of gene dropping simulations in which a lineage survives. We explored  
132 whether lineages from the experimentally introduced sparrows differed from native lineages in their  
133 rate of survival to 2018 (last year with complete data) and in their reproductive values. We chose to  
134 work with years rather than generations as a measure of time because sparrows have overlapping  
135 generations.

### 136 **Short-term fitness metrics**

137 We calculated the short-term fitness proxies for the founders and half-founders from cohorts between  
138 1999 and 2002 with complete life-history data. Founders with incomplete life-history data were  
139 removed because this could lead to an underestimation of their reproductive success. The individual  
140 lifetime production of eggs, broods, hatchlings, fledglings, and recruits was then calculated, as well as  
141 IGR at all four offspring stages, and de-lifed fitness. Hatchlings were defined as offspring counted in a  
142 nest two days after hatching, and fledglings were birds that survived until ringing, which is typically 12  
143 days after hatching. Recruits were defined as offspring that successfully reproduced and produced at  
144 least one egg in any subsequent years.

145 The IGR is the dominant eigenvalue of an individual population transition matrix, as described in (13).  
146 In an individual population transition matrix, the sub-diagonal represents survival, and the first row is  
147 filled with the number of offspring produced at each parental age, divided by two to account for parent–  
148 offspring relatedness being  $\frac{1}{2}$ . An example of an individual population transition matrix for an

149 individual that survived three years and had 1, 2 and 1 offspring at ages 1, 2, and 3 respectively, is given  
150 below:

151 
$$\begin{bmatrix} 0.5 & 1 & 0.5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

152 We also calculated annual de-lifed fitness based on the formula (8):

153 
$$p_{ti} = \frac{\xi_{t(i)} - w_{t(i)}}{N_t - 1}$$

154 Where:

- 155 •  $p_{ti}$  is individual fitness
- 156 •  $\xi_{t(i)}$  is the number of individual's surviving offspring at the end of the time step plus one if the  
157 individual survived
- 158 •  $w_{t(i)}$  is the population size in year  $t+1$  divided by population size in year  $t$
- 159 •  $N_t$  is the population size of adults on 1<sup>st</sup> April each year

160 While LRS and IGR are both lifetime fitness measures, de-lifing was designed primarily as a per-  
161 generation fitness proxy and is here calculated annually. However, lifetime de-lifed fitness can be  
162 obtained by summing the annual fitness values for each individual (8,31). We therefore used de-lifed  
163 fitness as both an annual fitness proxy and, after summing, as a lifetime fitness proxy.

164 We calculated Pearson's correlation between each fitness metric and the reproductive values for the  
165 lineages that survived. We ran a binary logistic regression model in R version 4.0.3 (32), using  
166 MCMCglmm (33) with lineage survival to 2018 as the response variable, and each fitness metric as the  
167 explanatory variable. The fitness metrics were z-transformed so that the slopes were not affected by the  
168 variable variances. We used priors with the residual variance fixed at 0.5 and ran the model for 100,000  
169 iterations with the thinning interval set at 70 and the burn-in at 7,000. We examined which fitness metric  
170 had the strongest association with reproductive values based on the slope of the regression.

171



172 [Ethics statement](#)

173 As this was a theoretical study using previously selected data, no ethics approval was required.

## 174 Results

### 175 Reproductive values

176 There were in total 111 lineages arising between 1999 and 2002 used for the analyses. Of these 111  
177 lineages, 18 lineages were founded by sparrows experimentally introduced in 2000 (26) and 93 lineages  
178 stem from native sparrows already present on the island in 2000. Forty-three lineages survived to 2018,  
179 of which 11 were introduced and 32 were native. Hence, at most, 39% of the founders passed genetic  
180 material to 2018, and there was no statistically significant association between a lineage's origin and  
181 survival ( $p = 0.06$ , Fisher's exact test). The mean lineage survival probability was 0.16 (95% CI 0.13–  
182 0.18), and the survival probability for lineages appearing in 2018 was 0.40 (95% CI 0.37–0.43, Fig 1A).  
183 There was variation in the absolute reproductive values (mean = 1.64, 95% CI 1.32–1.95 range: 0.41–  
184 18.89, Fig 1B). The introduced lineages had on average higher reproductive values than native lineages  
185 ( $t = 2.70$ ,  $df = 17.90$ ,  $p = 0.015$ ). Contributions varied over time, but after 2007 fluctuations were more  
186 synchronous among lineages, and the ranking of lineages based on their reproductive values remained  
187 similar (Fig 1E). Population fluctuations closely follow fluctuations in the reproductive values in the  
188 previous year. The change in lineage behaviour after 2007 is visible in lineage longevity, as all lineages  
189 that survived the from 2000 to 2006 also survived until 2018 (Fig 1C). After 2007, the correlation  
190 between annual reproductive value and reproductive values in 2018 also increased, and stabilised  
191 around 2011 subsequent to which the correlation was above 0.95 (Fig 1D).

192

193 *Fig 1: (A) Reproductive values and population size during the study period. Bars represent the population size*  
194 *and lines represent the reproductive values of each of the 43 lineages that survived to 2018. (B) Survival*  
195 *probability of 111 house sparrow lineages on Lundy Island from 1999 to 2018. (C) Reproductive values of the*  
196 *111 lineages. (D) Number of years to lineage extinction for the 111 lineages. Lineages that survived 18, 19, 20*  
197 *or 21 years are those that were still extant in 2018, corresponding to cohorts 1999, 2000, 2001 and 2002*  
198 *respectively. (E) Correlation between the reproductive values in each year and the final year.*

199

## 200 Fitness proxies

201 Fitness proxies were calculated for 86 founders, 44 males and 42 females. We estimated the correlation  
202 with the reproductive values of 42 lineages that survived to 2018 and had no missing fitness data. In  
203 total, individuals included in the analysis produced 2,054 eggs of which 1,746 (85%) survived to  
204 hatching, 881 (43%) to fledging and 294 (14%) recruited into the breeding population.

205 The fitness proxies were all positively associated with the reproductive values (Fig. 2). De-lifed fitness,  
206 IGR and LRS for recruits were all statistically significantly correlated with the reproductive values.  
207 There were no statistically significant differences between the IGR, LRS and de-lifed fitness correlation  
208 coefficients. None of the other correlation estimates were statistically significant (Table 1). De-lifed  
209 fitness at ages 1 and 2 significantly correlated with reproductive value, but in older age classes the  
210 correlation estimate was not statistically significant (Figs 3A and 3B).

211

212

213 *Fig 2: Correlation between each of the fitness proxies at different life stages with reproductive value. Error*  
214 *bars represent 95% confidence intervals. Black bars represent significant results, while light grey bars*  
215 *represent non-significant results.*

216

217 *Fig 3: (A) Correlation between reproductive value and de-lifed fitness by age class (in years), with 95%*  
218 *confidence intervals. N represents the sample size. Older age classes have lower sample sizes because fewer*  
219 *individuals survive to that age. (B) Correlation between de-lifed fitness and reproductive value. The black cross*  
220 *represents mean de-lifed fitness for the respective age. Colours represent the same age class.*

221

222 *Table 1: Mean and standard deviation for short-term fitness proxies at different offspring stages, and de-lifed*  
223 *fitness at different ages.*

	Mean	SD
LRS Eggs	23.78	18.02
LRS Broods	5.80	4.26
LRS Hatchlings	20.30	16.04
LRS Fledglings	10.24	7.85
LRS Recruits	3.42	3.36
De-lifing	0.01	0.03
IGR eggs	4.24	1.53
IGR hatchlings	3.71	1.37
IGR fledglings	2.43	0.99
IGR recruits	1.13	0.78
De-lifing – Age 1	0.0003	0.0198
De-lifing – Age 2	0.0056	0.0170
De-lifing – Age 3	0.0027	0.0115
De-lifing – Age 4	0.0023	0.0087
De-lifing – Age 5	0.0033	0.0048
De-lifing – Age 6	0.0005	0.0046

224

225 There was a significant positive relationship between lineage survival odds and de-lifed fitness, LRS at  
226 recruitment and fledgling stages, and IGR at recruitment (Table 2). The estimated slopes for the de-  
227 lifed fitness and LRS recruits were significantly higher than the slopes of IGR and LRS at fledgling  
228 stage as their 95% confidence intervals are non-overlapping, indicating that fitness measured at the  
229 recruitment stage for these two metrics is a better predictor of lineage survival. There were no  
230 statistically significant differences between IGR, LRS and de-lifed fitness at the same offspring stage  
231 (Table 2).

232

233 *Table 2: Results of binary logistic regressions with lineage survival as a response variable. l-95% CI and u-*  
234 *95% CI are the lower and upper boundaries of the 95% confidence interval for the slope, respectively.*

<b>Variable</b>	<b>Slope</b>	<b>l-95% CI</b>	<b>u-95% CI</b>	<b>p value</b>
LRS eggs	0.42	-0.07	0.95	0.101
LRS hatchlings	0.45	-0.08	0.95	0.063
LRS fledglings	1.06	0.41	1.71	0.001*
LRS recruits	3.26	1.73	4.74	0.001*
LRS broods	0.47	-0.02	0.96	0.056
IGR eggs	0.24	-0.21	0.75	0.310
IGR hatchlings	0.35	-0.09	0.87	0.167
IGR fledglings	1.06	0.49	1.71	0.002
IGR recruits	2.65	1.67	3.70	0.001*
De-lifed fitness	3.01	1.74	4.32	0.001*

235

## 236 Discussion

237 We showed that fitness proxies measured at recruit stage correlates best with long-term reproductive  
238 values and lineage survival, while fitness proxies measured at earlier stages are less useful.

239 Similar to a study by (4) lineage survival is low. While there was no difference in the rate of lineage  
240 survival between native and introduced lineages, the introduced lineages had significantly higher  
241 reproductive values. This indicates that the introduced lineages might have a fitness advantage over the  
242 native ones. For lineages that survived to 2018, there was wide variation in survival probability and  
243 reproductive value. The survival probability of a lineage is associated with its reproductive value in that  
244 year (4,18). While several lineages died out every year prior to 2007, all lineages that survived the  
245 bottleneck in 2008 also survived the next 10 years to 2018. Lineage extinctions are expected to become  
246 less likely over the generations, as all founders with non-zero reproductive values become genealogical  
247 ancestors of all individuals in the future population after only a few generations (19,22). After a founder  
248 becomes an ancestor of all individuals in the current population, its lineage can only go extinct if the  
249 entire population goes extinct. During the 2008 bottleneck, the population size decreased significantly,  
250 shortening the time it took for all founders of persisting lineages to become the common ancestors of  
251 the current population members.

252 There was variation in reproductive value, with most lineages ranging from 0 to 10 but some reaching  
253 contributions of over four times that much. There was also large variation over time as lineages  
254 fluctuated. Lineage stabilisation is also visible in the pattern of lineage fluctuation through time after  
255 2007, as the ranking of lineages based on reproductive value remains similar. The rapidly increasing  
256 correlation between reproductive value in the final year and each of the previous years also shows a  
257 pattern of stabilisation, as found in other studies too (4,18,34). Stabilisation is reached after 12 years,  
258 with the correlation exceeding 0.95 afterwards. Despite stabilisation, reproductive values fluctuated  
259 through time. As we examined reproductive values that are absolute rather than relative to population  
260 size, any change in population size is also reflected in the sum of the reproductive values the year before.  
261 The change in reproductive value occurs one year previously, because the estimates are based on



262 reproducing offspring, which are only recognised in the next year and form the basis of next year's  
263 population.

264 The fitness proxies based on the number of recruits outperformed all other fitness proxies in predicting  
265 reproductive values and lineage survival. Recruits are likely to be the best measure because they are  
266 adult individuals that reproduced, while other proxies include the uncertainty of survival to adulthood  
267 before reproduction even occurs. Given that sparrow offspring experience high rates of mortality, with  
268 only 14% of laid eggs successfully surviving to recruitment, mortality will have a big impact on  
269 reproductive values from short-term metrics measured at early offspring stages. For species with lower  
270 offspring mortality the age at which offspring are counted towards fitness may have less influence on  
271 the predictive power of short-term fitness metrics. While recruits are clearly the best predictor of long-  
272 term fitness, they are the most difficult to measure in most study systems, as it is rarely possible to  
273 monitor all offspring until their first reproduction. This highlights the importance of long-term isolated  
274 island population studies (35), as only in such studies is it possible to accurately estimate the number  
275 of genetic recruits that an individual produced.

276 We found no differences in the performance of de-lifed fitness, IGR or LRS in predicting reproductive  
277 values or lineage survival. A previous study on Ural owls (*Strix uralensis*) and collared flycatchers  
278 (*Ficedula albicollis*) found that LRS performed significantly better than IGR at fledgling stage in  
279 predicting reproductive values, while they both performed similarly at recruitment (10). The correlation  
280 between reproductive value and different fitness proxies at recruit stage was of similar strength as  
281 discovered in previous studies (4,10).

282 In this study, annual de-lifed fitness at ages 1 and 2 were correlated with the reproductive values, but  
283 not at later ages. The correlation at age 1 with reproductive value was similar to that for lifetime de-  
284 lifed fitness, indicating that reproductive success in the first adult year may be sufficient to provide a  
285 good prediction of long-term fitness. Hence, individual reproductive performance in the first year may  
286 be an important proxy for an individual's fitness.

287 There is, however, considerable variation that is not explained by the fitness metrics. A strong  
288 correlation between a short-term fitness metric and the reproductive value measured two decades later,  
289 during which the population has been exposed to varying environmental conditions and population  
290 fluctuations, is unlikely. The strength of the correlation will also depend on the additive genetic variance  
291 and heritability of reproductive success (4). In particular, in our population annual fitness is somewhat  
292 heritable (36), and there has been significant demographic stochasticity in our population for which  
293 LRS and IGR metrics tested here were not designed (37).

294 The underlying theoretical results about reproductive value were derived under the assumption of  
295 diploid Wright-Fisher population and weak selection (18). The Lundy sparrow population might not  
296 meet these assumptions, as there could be undetected strong selection and non-random mating. The  
297 sparrow population can therefore be used to test theoretical predictions on real data but could lead to  
298 erroneous conclusions if assumptions are severely breached. Despite that, reproductive values can lead  
299 to new insights about natural selection and evolutionary outcomes, such as inbreeding, lineage  
300 introgression or cohort effects (2,4,34). Particularly in the presence of environmental, social, or  
301 demographic interactions, such as those occurring in any wild population, studying fitness across an  
302 entire lineage by examining reproductive values can potentially lead to better estimation of  
303 evolutionary outcomes for a certain allele (38).

304 In conclusion, by using reproductive values as a measure of long-term individual fitness we have shown  
305 that recruits, rather than earlier offspring stages, best predict reproductive values. Additionally, annual  
306 fitness measured in the first reproductive season is an equally good predictor of fitness as lifetime fitness  
307 measures. We therefore suggest that future studies should measure short-term fitness at higher offspring  
308 ages to better capture long-term fitness.

309

310

311 [Author Contributions](#)

312 Conceptualization: Vita Živa Alif, Julia Schroeder, Terry Burke

313 Data Curation: Jamie Dunning, Janet Chik, Julia Schroeder, Terry Burke

314 Formal Analysis: Vita Živa Alif

315 Funding acquisition: Terry Burke, Julia Schroeder

316 Investigation: Vita Živa Alif

317 Project administration: Terry Burke, Julia Schroeder

318 Resources: Terry Burke, Julia Schroeder

319 Software: Vita Živa Alif

320 Supervision: Julia Schroeder

321 Validation: Vita Živa Alif

322 Visualisation: Vita Živa Alif

323 Writing – original draft: Vita Živa Alif

324 Writing – review & editing: Vita Živa Alif, Julia Schroeder, Jamie Dunning, Janet Chik, Terry Burke

## 325 References

- 326 1. Fisher RA. The genetical theory of natural selection. [Internet]. Oxford: Clarendon Press; 1930.  
327 302 p. Available from: <https://www.biodiversitylibrary.org/bibliography/27468>
- 328 2. Grafen A. The Price equation and reproductive value. *Philos Trans R Soc B Biol Sci.* 2020 Apr  
329 27;375(1797):20190356.
- 330 3. Orr HA. Fitness and its role in evolutionary genetics. *Nat Rev Genet.* 2009 Aug;10(8):531–9.
- 331 4. Reid JM, Nietlisbach P, Wolak ME, Keller LF, Arcese P. Individuals' expected genetic contributions  
332 to future generations, reproductive value, and short-term metrics of fitness in free-living song  
333 sparrows (*Melospiza melodia*). *Evol Lett.* 2019;3(3):271–85.
- 334 5. Saether B-E, Engen S. The concept of fitness in fluctuating environments. *Trends Ecol Evol.*  
335 2015;30(5):9.
- 336 6. Brommer JE, Merilä J, Kokko H. Reproductive timing and individual fitness. *Ecol Lett.*  
337 2002;5(6):802–10.
- 338 7. Dobson FS, Murie J, Viblanc V. Fitness Estimation for Ecological Studies: An Evaluation in  
339 Columbian Ground Squirrels. *Front Ecol Evol* [Internet]. 2020 [cited 2020 Nov 11];8. Available  
340 from: <https://hal.archives-ouvertes.fr/hal-02903692>
- 341 8. Coulson T, Benton T g, Lundberg P, Dall S r. x, Kendall B e, Gaillard J-M. Estimating individual  
342 contributions to population growth: evolutionary fitness in ecological time. *Proc R Soc B Biol Sci.*  
343 2006 Mar 7;273(1586):547–55.
- 344 9. Dobson FS, Lane JE, Low M, Murie JO. Fitness implications of seasonal climate variation in  
345 Columbian ground squirrels. *Ecol Evol.* 2016;6(16):5614–22.
- 346 10. Brommer JE, Gustafsson L, Pietiäinen H, Merilä J. Single-generation estimates of individual  
347 fitness as proxies for long-term genetic contribution. *Am Nat.* 2004 Apr 1;163(4):505–17.
- 348 11. Moorad JA. Individual fitness and phenotypic selection in age-structured populations with  
349 constant growth rates. *Ecology.* 2014;95(4):1087–95.
- 350 12. Roff DA. Defining fitness in evolutionary models. *J Genet.* 2008 Dec 1;87(4):339–48.
- 351 13. McGraw JB, Caswell H. Estimation of Individual Fitness from Life-History Data. *Am Nat* [Internet].  
352 1996 [cited 2020 Nov 17]; Available from:  
353 <https://www.journals.uchicago.edu/doi/abs/10.1086/285839>
- 354 14. Rubach K, Dobson FS, Zinner B, Murie J, Viblanc V. Comparing fitness measures and the influence  
355 of age of first reproduction in Columbian ground squirrels. *J Mammal* [Internet]. 2020 [cited  
356 2020 Dec 8]; Available from: <https://hal.archives-ouvertes.fr/hal-02916845>
- 357 15. Grafen A. On the uses of data on lifetime reproductive success. In: Clutton-Brock TH, editor.  
358 Reproductive success. Chicago, Illinois, USA: University of Chicago Press; 1988. p. 454–71.
- 359 16. Käär P, Jokela J. Natural selection on age-specific fertilities in human females: comparison of  
360 individual-level fitness measures. *Proc R Soc Lond B Biol Sci.* 1998 Dec 22;265(1413):2415–20.

- 361 17. Wolf JB, Wade MJ. On the assignment of fitness to parents and offspring: whose fitness is it and  
362 when does it matter? *J Evol Biol.* 2001;14(2):347–56.
- 363 18. Barton NH, Etheridge AM. The relation between reproductive value and genetic contribution.  
364 *Genetics.* 2011 Aug;188(4):953–73.
- 365 19. Chang JT. Recent common ancestors of all present-day individuals. *Adv Appl Probab.*  
366 1999;31(4):1002–26.
- 367 20. Chen N, Juric I, Cosgrove EJ, Bowman R, Fitzpatrick JW, Schoech SJ, et al. Allele frequency  
368 dynamics in a pedigreed natural population. *Proc Natl Acad Sci.* 2019 Feb 5;116(6):2158–64.
- 369 21. Hill WG. Variation in genetic identity within kinships. *Heredity.* 1993 Dec;71(6):652–3.
- 370 22. Gravel S, Steel M. The existence and abundance of ghost ancestors in biparental populations.  
371 *Theor Popul Biol.* 2015 May 1;101:47–53.
- 372 23. Grafen A. A theory of Fisher’s reproductive value. *J Math Biol.* 2006 Jul;53(1):15–60.
- 373 24. Schroeder J, Nakagawa S, Rees M, Mannarelli M-E, Burke T. Reduced fitness in progeny from old  
374 parents in a natural population. *Proc Natl Acad Sci.* 2015 Mar 31;112(13):4021–5.
- 375 25. Kozłowski J. Measuring fitness in life-history studies. *Trends Ecol Evol.* 1993 Mar;8(3):84–5.
- 376 26. Ockendon N, Griffith SC, Burke T. Extrapair paternity in an insular population of house sparrows  
377 after the experimental introduction of individuals from the mainland. *Behav Ecol.* 2009 Mar  
378 1;20(2):305–12.
- 379 27. Schroeder J, Cleasby IR, Nakagawa S, Ockendon N, Burke T. No evidence for adverse effects on  
380 fitness of fitting passive integrated transponders (PITs) in wild house sparrows *Passer*  
381 *domesticus*. *J Avian Biol.* 2011;42(3):271–5.
- 382 28. Simons MJP, Winney I, Nakagawa S, Burke T, Schroeder J. Limited catching bias in a wild  
383 population of birds with near-complete census information. *Ecol Evol.* 2015;5(16):3500–6.
- 384 29. MacCluer JW, VandeBerg JL, Read B, Ryder OA. Pedigree analysis by computer simulation. *Zoo*  
385 *Biol.* 1986;5:147–60.
- 386 30. Wolak ME. *nadiv*: an R package to create relatedness matrices for estimating non-additive  
387 genetic variances in animal models. *Methods Ecol Evol.* 2012;3(5):792–6.
- 388 31. Gratten J, Wilson AJ, McRae AF, Beraldi D, Visscher PM, Pemberton JM, et al. A Localized  
389 Negative Genetic Correlation Constrains Microevolution of Coat Color in Wild Sheep. *Science.*  
390 2008 Jan 18;319(5861):318–20.
- 391 32. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R  
392 Foundation for Statistical Computing; 2020.
- 393 33. Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The  
394 MCMCglmm R Package. *J Stat Softw [Internet].* 2010 [cited 2021 Jan 23];33(2). Available from:  
395 <http://www.jstatsoft.org/v33/i02/>

- 396 34. Hunter DC, Pemberton JM, Pilkington JG, Morrissey MB. Pedigree-based estimation of  
397 reproductive value. *J Hered*. 2019 Jul 1;110(4):433–44.
- 398 35. Clutton-Brock T, Sheldon BC. Individuals and populations: the role of long-term, individual-based  
399 studies of animals in ecology and evolutionary biology. *Trends Ecol Evol*. 2010 Oct 1;25(10):562–  
400 73.
- 401 36. Schroeder J, Burke T, Mannarelli M-E, Dawson DA, Nakagawa S. Maternal effects and heritability  
402 of annual productivity: Maternal effects of annual productivity. *J Evol Biol*. 2012 Jan;25(1):149–  
403 56.
- 404 37. Engen S, Lande R, Sæther B, Dobson FS. Reproductive value and the stochastic demography of  
405 age-structured populations. *Am Nat*. 2009 Dec;174(6):795–804.
- 406 38. Graves CJ, Weinreich DM. Variability in Fitness Effects Can Preclude Selection of the Fittest. *Annu*  
407 *Rev Ecol Evol Syst*. 2017;48(1):399–417.
- 408
- 409

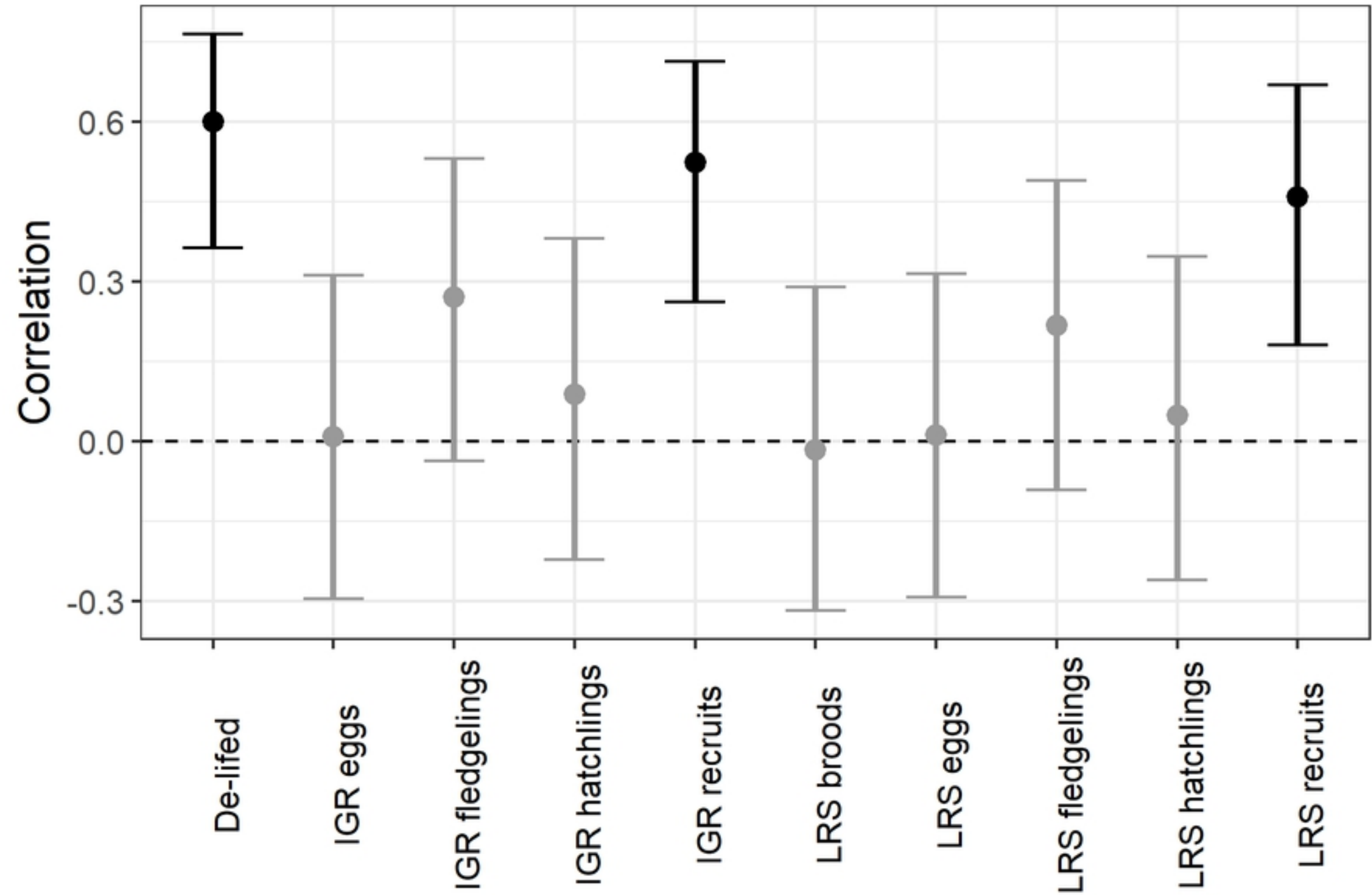


Figure 2

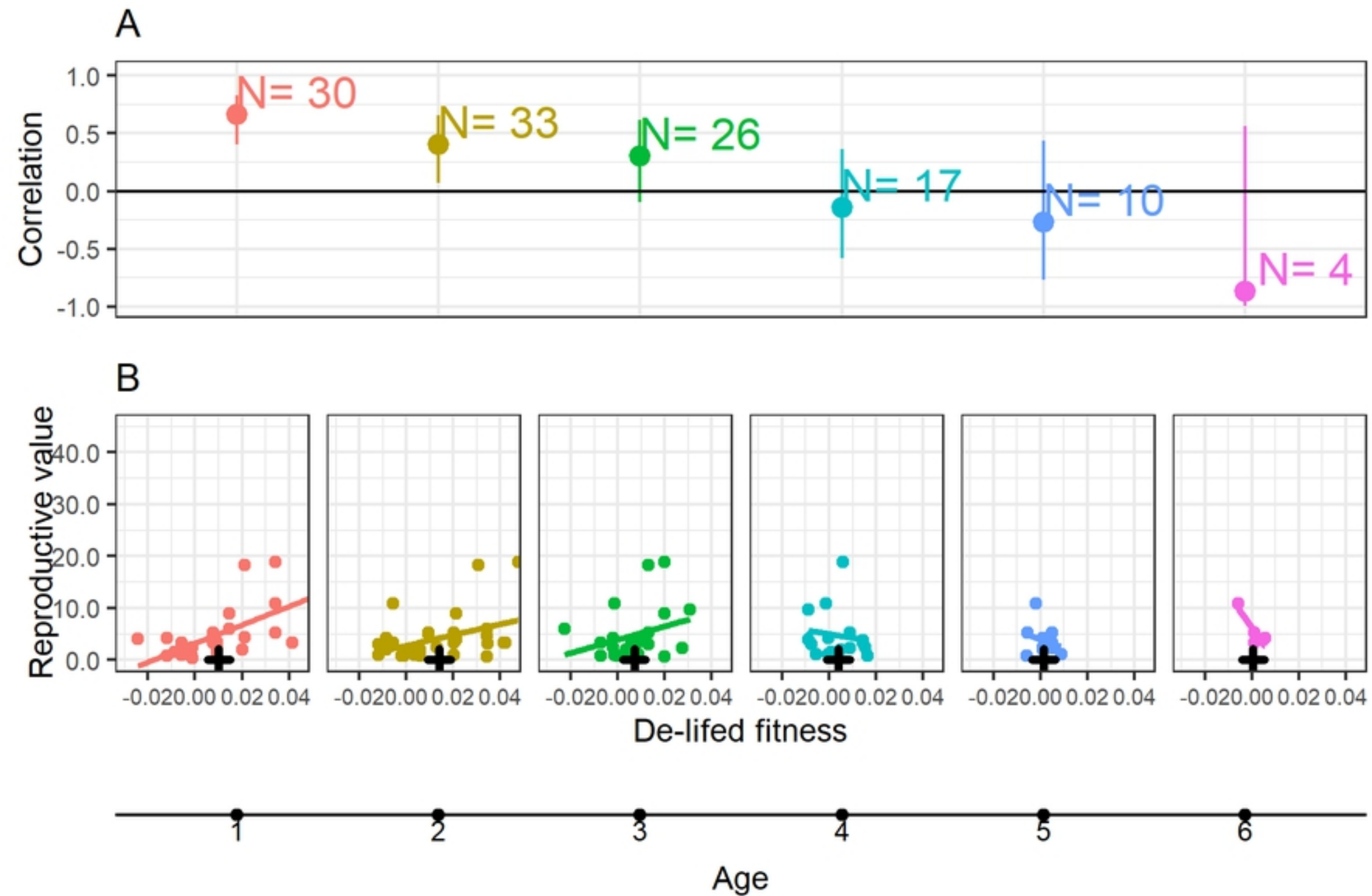


Figure 3



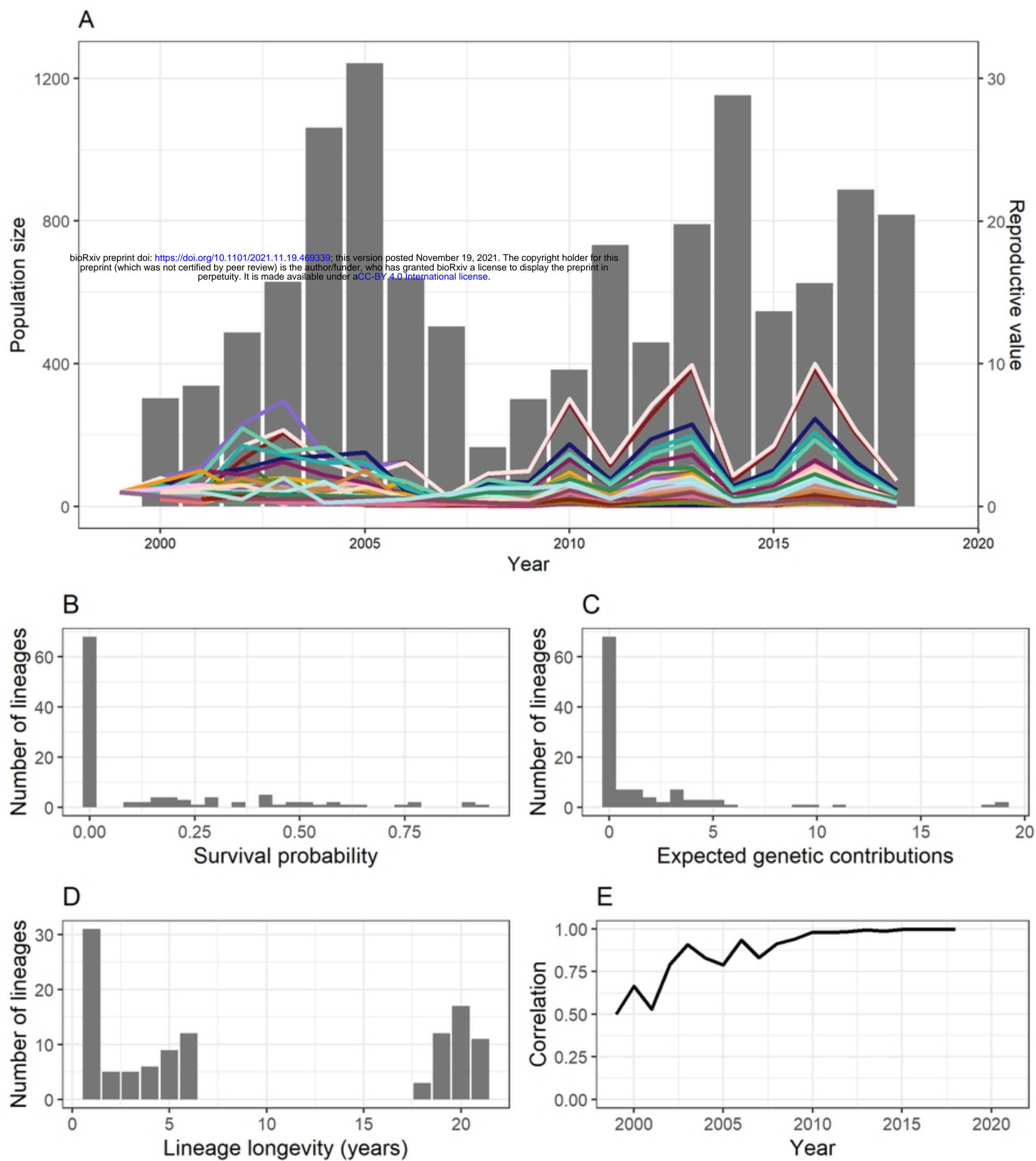


Figure 1