1

2

3

Detection of genetic purging and predictive value of purging

parameters estimated in pedigreed populations

Eugenio López-Cortegano[1], Diego Bersabé[1], Jinliang Wang[2], and Aurora

García-Dorado[1]*

8

9

[1]Departamento de Genética. Facultad de Biología. Universidad

Complutense. 28040, Madrid. Spain.

[2]Institute of Zoology. Zoological Society of London. Regent's Park, London

NW1 4RY. United Kingdom.

* Corresponding author

15

Running title: purging analysis of pedigreed data

17

Key words: Inbreeding depression; Inbreeding-Purging model; ancestral

inbreeding; PURGd software; Simulation.

20

21

22

ABSTRACT

23

24   The consequences of inbreeding for fitness are important in evolutionary and

25   conservation biology, but can critically depend on genetic purging. However, estimating

26   purging has proven elusive. Using PURGd software, we assess the performance of the

27   Inbreeding-Purging (IP) model and of ancestral inbreeding (F_a) models to detect purging in

28   simulated pedigreed populations, and to estimate parameters that allow reliably predicting

29   the evolution of fitness under inbreeding. The power to detect purging in a single small

30   population of size N is low for both models during the first few generations of inbreeding (t

31   $\approx N/2$), but increases for longer periods of slower inbreeding and is, on average, larger for

32   the IP model. The ancestral inbreeding approach overestimates the rate of inbreeding

33   depression during long inbreeding periods, and produces joint estimates of the effects of

34   inbreeding and purging that lead to unreliable predictions for the evolution of fitness. The IP

35   estimates of the rate of inbreeding depression become downwardly biased when obtained

36   from long inbreeding processes. However, the effect of this bias is canceled out by a

37   coupled downward bias in the estimate of the purging coefficient so that, unless the

38   population is very small, the joint estimate of these two IP parameters yields good

39   predictions of the evolution of mean fitness in populations of different sizes during periods

40   of different lengths. Therefore, our results support the use of the IP model to detect

41   inbreeding depression and purging, and to estimate reliable parameters for predictive

42   purposes.

43

INTRODUCTION

Inbreeding depression is a major threat to the survival of small endangered populations. It is mainly due to the increase in the frequency of homozygous genotypes for deleterious recessive alleles, which leads to fitness decay and increased extinction risk (Lande, 1994; Hedrick and Kalinowski, 2000; O'Grady *et al.*, 2006; Charlesworth and Willis, 2009). However, deleterious recessive alleles that escape selection in non-inbred populations because they are usually in heterozygosis, can be purged under inbreeding as they are exposed in homozygosis. This is expected to result in a reduction of fitness depression and in some fitness recovery, unless the effective population size and the effects of deleterious alleles are so small that drift overwhelms natural selection (García-Dorado, 2012, 2015).

While inbreeding depression is ubiquitously documented (Crnokrak and Roff, 1999; O'Grady *et al.*, 2006), there is far less empirical evidence for the effect of genetic purging. Evidence of purging has often been obtained in situations where inbreeding increases slowly, but many studies have failed to detect purging in both wild and captive populations or have just detected purging effects of small magnitude, particularly under fast inbreeding or during short periods of slow inbreeding (Ballou, 1997; Bryant *et al.*, 1999; Byers and Waller, 1999; Crnokrak and Barrett, 2002; Boakes *et al.*, 2006; Kennedy *et al.*, 2014). This is not surprising, since purging is expected to be less efficient under faster inbreeding, but more delayed under slower inbreeding. Furthermore, purging can be difficult to detect because of lack of experimental power or confounding effects, as concurring adaptive processes (Hedrick and García-Dorado, 2016; López-Cortegano *et al.*, 2016). Thus, failure to detect purging does not mean that purging is irrelevant in actual populations. Developing methods and tools to detect and evaluate purging is of critical importance in conservation, as it may help to improve management policies.

69   The first models aimed to detect purging from pedigreed fitness data were based on

70   different regression approaches that use an ancestral inbreeding coefficient (F_a) to define

71   the independent variable(s) accounting for purging (Ballou, 1997; Boakes et al., 2006).

72   This F_a coefficient, first described by Ballou (1997), represents the average proportion of

73   an individual's genome that has been in homozygosis by descent in at least one

74   ancestor. It is relevant to purging because recessive deleterious alleles can be purged in

75   inbred ancestors, so that individuals with higher F_a are expected to carry fewer such

76   alleles than those with the same level of inbreeding but lower F_a values, and should

77   therefore have higher fitness. Gulisija and Crow (2007) developed a different index to

78   measure the opportunity of purging (O_i) by assuming that, in the same pedigree path, there

79   are no two ancestors that are homozygous for the same deleterious allele. However, the

80   authors noted that, due to this assumption, their approach is appropriate to evaluate the

81   opportunities of purging just for completely recessive and severely deleterious alleles with

82   low initial frequency in shallow pedigrees. Furthermore, they did not develop an explicit

83   model for the dependence of fitness on the opportunity of purging. Therefore, here we do

84   not investigate the properties of this index.

85   More recently, an Inbreeding-Purging (IP) model has been proposed, based on a

86   "purged inbreeding coefficient" (g), that predicts how mean fitness and inbreeding load are

87   expected to evolve in a population undergoing inbreeding. This coefficient g is defined as

88   Wright's inbreeding coefficient (F) adjusted for the reduction in frequency of the

89   deleterious alleles caused by purging, so that it is the coefficient appropriate to predict the

90   actual increase in homozygosis for these alleles. It depends on a purging coefficient (d)

91   that represents the enhancement of selection under inbreeding (García-Dorado, 2012). For

92   each single deleterious allele, d equals the recessive component of the selection coefficient,

93    $i.e.$, the deleterious effect that is concealed in the heterozygous and expressed just in the

94   homozygous condition. Note that d equals the heterozygous value for relative fitness in the

4

95   classical quantitative genetics scale proposed by Falconer (Falconer and Mackay 1996). For

96   overall fitness, which is affected by many alleles with different deleterious effects, reliable IP

97   predictions can be obtained by using a single empirically defined d value. The dependence of

98   g on d is illustrated in Figure 1, and shows that purging is more efficient when inbreeding

99   is slower (i.e., when the effective population size is larger), but also takes longer to

100   become relevant. Therefore, this model predicts that the rate of inbreeding (or the

101   effective population size N) and the number of inbreeding generations (t) critically

102   determine the extent of purging.

103   The purging coefficient d has been estimated from the evolution of mean fitness

104   in *Drosophila* experiments, the IP model providing a much better fit than a model

105   without purging (Bersabé and García-Dorado, 2013; López-Cortegano *et al.*, 2016).

106   Furthermore, equations have been derived to obtain IP predictions for pedigreed

107   individuals and have been implemented in the free software package PURGd. This

108   software analyzes pedigreed fitness data to obtain estimates of the IP parameters,

109   namely the rate of inbreeding depression δ and the purging coefficient d (García-

110   Dorado, 2012; García-Dorado *et al*., 2016). Preliminary analysis of simulated data

111   showed that this software accurately discriminates between situations with and without

112   purging, and that the genealogical IP approach consistently provided a good fit to the data.

113   However, the estimates of δ and d showed some downward bias (García-Dorado *et al*.,

114   2016). Thus, before this method is applied to real data, it is necessary to characterize the bias

115   of (δ, d) estimates obtained under different scenarios and to check how far it affects the

116   reliability of IP predictions of fitness evolution computed using them.

117   Here, we analyze fitness data of simulated pedigreed individuals undergoing inbreeding

118   and purging in order to investigate: i) how often the IP and F_a-based approaches allow to

119   detect purging; ii) the extent to which the estimates of the model's parameters depend on the

120   rate of inbreeding (here determined by the population size N) and on the number of

5

121   inbreeding generations (t); iii) how reliable are the IP and F_a-based predictions for

122   inbreeding scenarios with N and/or t values different from those used to estimate the model's

123   parameters.

124

125

126   **MATERIAL AND METHODS**

127   **The simulated populations**

128   A monoecious panmictic population of size $N = 10^3$ is simulated under a mutation-

129   selection-drift (MSD) scenario over 10^4 generations to obtain a base population that can be

130   assumed to be at the MSD balance. Mutations occur at a rate λ per genome and

131   generation, and have selection coefficient s and degree of dominance h, so that fitness is

132   reduced by $h{\cdot}s$ or s when the mutant allele is in heterozygosis or homozygosis,

133   respectively. According to the standard assumption of non-epistatic models, fitness is

134   multiplicative across loci. In practice, fitness effects can be epistatic to some extent. In

135   particular, the homozygous effect of a deleterious allele may be larger in individuals that are

136   also homozygous for other deleterious alleles, giving reinforcing epistasis that involves

137   recessive components. However, although this could be expected to produce an increase in

138   inbreeding depression, previous simulation results suggest that this increase is canceled out

139   by a parallel excess in purging, so that simple IP predictions not accounting for epistasis still

140   fit the evolution of mean fitness under inbreeding (Pérez-Figueroa *et al*., 2009). The

141   simulation methods are described in detail by Bersabé *et al*. (2016).

142   Two different sets of mutational parameters (CAPTIVE and WILD, summarized in

143   Table 1) are considered. In both cases, a variable selection coefficient is sampled from a

144   gamma distribution with shape parameter $\alpha = 3^{-1}$ and rate parameter $\beta = \alpha / E(s)$,

145   where $E(s)$ stands for the expected s value. Sampled s values larger than 1 are assigned

146   as $s = 1$. The mutation rate and average deleterious effect in the WILD case are twice those

147   of the CAPTIVE one, in order to account for the inbreeding load that has been empirically

148   detected in the wild, which is about fourfold that of captive populations (Ralls *et al.*, 1988;

149   O'Grady *et al.*, 2006; Hedrick and García-Dorado, 2016). For each given *s* value, the

150   degree of dominance *h* is sampled from a uniform distribution ranging between 0 and

151    $e^{-7.5s}$ (García-Dorado, 2003). Note that this gives an average degree of dominance (E(*h*))

152   that is larger in the CAPTIVE than in the WILD case, as the average selection

153   coefficient is lower. The corresponding distributions of homozygous effects are shown

154   in Figure 2.

155   For each case considered, ten base populations are simulated. Populations of reduced

156   size *N*=10, *N*=25, and *N*=50 (lines) are obtained from these base populations at the

157   MSD equilibrium (250, 100 and 50 replicates, respectively, each of the 10 base populations

158   contributing equal numbers of replicates for each size). Effective population sizes are

159   assumed to equal actual population sizes. All lines are continued for 2*N* generations

160   following the same protocol as for the base populations (*i.e.*, under mutation, selection and

161   drift), and pedigrees and individual fitness are recorded.

162

163   **Estimation of inbreeding depression and purging**

164   ***IP Model:*** This model predicts fitness as a function of a purged inbreeding coefficient *g*

165   that is defined as Wright's *F* inbreeding coefficient corrected for the reduction in frequency

166   of deleterious alleles expected from purging. This *g* coefficient can be computed as a

167   function of the purging coefficient *d* (García-Dorado, 2012). For a model with constant

168   effects across loci, *d* equals the per-copy deleterious effect that is expressed in homozygosis

169   but is concealed in heterozygosis (*d*=*s*(1-2*h*)/2). For more realistic models where deleterious

170   effects vary across loci, as in our simulated populations, IP predictions should be averaged

171   over the distribution of deleterious effects. Since this approach is not possible in practical

172   situations, an effective purging coefficient (here referred to just as purging coefficient and

173   denoted by d) has been defined empirically as the d value giving the best predictions when

174   used in the IP model, which has been shown to produce good approximations (García-

175   Dorado, 2012). A simple recurrence equation calculates g each generation as a function of d,

176   the effective population size N, and the F and g values in the previous generation, or from

177   pedigree data. García-Dorado *et al*. (2016) generalized the pedigree recurrence equations to

178   allow for overlapping generations. These equations parallel those classically used to predict

179   the evolution of F using Malecot's coancestry coefficients, introducing an additional term

180   that depends on d. Thus, the model can predict either the average fitness expected at

181   generation t (W_t), or the expected fitness for an individual i with pedigree records (W_i). In

182   the case of individual fitness,

183    $$W_i = W_0 \cdot e^{-\delta \cdot g_i} \quad , \qquad\qquad (1)$$

184   where δ is the rate of inbreeding depression, g_i is the purged inbreeding coefficient of

185   individual i computed using d (Figure 1), and W_0 is the expected fitness in the non-

186   inbred population.

187   Note that, if natural selection can be ignored during the inbreeding period, g can

188   be replaced with F, and δ equals the inbreeding load B in the base population defined

189   as the sum over loci of $2s(1/2\text{-}h)\,q(1\text{-}q)$, as shown by Morton *et al*. (1956), where q is the

190   frequency of the deleterious allele. Thus, the inbreeding load B can be interpreted as the

191   expected rate of inbreeding depression if natural selection is neglected during the

192   inbreeding process. This can be appropriate when very few generations are considered,

193   so that purging has no opportunity to occur, when natural selection is overwhelmed by

194   drift due to a very small effective population size, or when natural selection is relaxed

195   by maintaining a population in benign conditions, as it could occur to some extent in *ex*

196   *situ* conservation programs. Otherwise, purging selection must be taken into account

197   by replacing F with g. Furthermore, non-purging selection (*i.e.*, selection as it would

198   operate in an equilibrium population with stable homozygosis) should also be

considered, at least in not too small populations, as it can compensate for a significant

fraction of the inbreeding depression. To understand this concept, discussed in the

section devoted to the Full Model (FM) in García-Dorado (2012), think of a population

at the MSD equilibrium. This population has finite size N (*i.e.*, inbreeding increases at

a rate $1/2N$) and a given inbreeding load, but it does not experience inbreeding

depression because it is compensated by natural selection. This kind of selection is not

due to a net increase in homozygosis and, therefore, it can be considered part of the

standard selection occurring in populations at the MSD balance and we do not use the

term purging to describe it. According to this Full Model, due to non-purging selection,

the actual expected rate of inbreeding depression as a function of g is $\delta_{FM} = B - B^*$,

where B and B^* are, respectively, the inbreeding loads expected at the MSD balance

for the original non-inbred population and for the new reduced size N. To obtain this

δ_{FM} value, we compute B and B^* using Equations 10 and 13 in García-Dorado (2007),

both averaged over 10^6 (s, h) values sampled from the corresponding joint distribution

(s values larger than 1 were assigned $s=1$ as in the simulation process). Note that δ_{FM}

approaches B for very small populations, but can be substantially smaller when N is

large.

For each pedigree, we estimate the purging coefficient d and the rate of inbreeding

depression δ using the PURGd 2.0 software package (García-Dorado *et al*., 2016; freely

available at https://www.ucm.es/genetica1/mecanismos). These estimates are obtained using

the two methods implemented in PURGd. Results obtained using linear regression for

log-transformed fitness (LR method) are not qualitatively different from those obtained

using the numerical non-linear regression method (NNLR), but give more downwardly

biased estimates of δ and larger standard errors. These LR results are not reported in

the main text, although a summarizing figure is given in the Supplementary Material

(Figure S1). Thus, we only report results from the NNLR method, which fits predictions

225 from Equation 1 by numerically searching for estimates that minimize the residual sums of

226 squares (García-Dorado *et al.*, 2016). The expected fitness value in the non-inbred

227 population, $E(W_0)$, is obtained in a previous step as the mean fitness of non-inbred

228 individuals with non-inbred ancestors ($F = F_a = 0$), as explained in García-Dorado *et*

229 *al*. (2016). Therefore, the program produces estimates of δ and d that are conditional

230 to this estimate of the non-inbred expected fitness. To check for the convergence of the

231 numerical algorithm, we estimate the genetic parameters for each pedigree as the result

232 of a single run, and as the average of five and ten independent runs.

233   A bootstrap method was devised to test the statistical significance of the estimate

234 of d obtained from each replicate line against the null hypothesis $d=0$ and is described

235 in the Supplementary Material.

236   ***Ancestral Inbreeding models:*** Ballou (1997) defined the ancestral inbreeding

237 coefficient (F_a) as the fraction of an individual's genome that has been in

238 homozygosis by descent in at least one ancestor, calculated in terms of the

239 inbreeding coefficient (F) and the ancestral inbreeding coefficient of the individual's

240 parents (sire S and dam D) as

241 $$F_a = \frac{1}{2}\{F_{a\,(D)} + [1 - F_{a\,(D)}] \cdot F_{(D)} + F_{a\,(S)} + [1 - F_{a\,(S)}] \cdot F_{(S)}\} \qquad (2)$$

242   Thus, F_a is related to the purging opportunities in the ancestors of an individual. This

243 equation assumes independence between F and F_a in the same individual, which can lead to

244 some overestimation of ancestral inbreeding. In order to avoid this bias, it has been proposed

245 to estimate ancestral inbreeding by using the so-called gene dropping simulation approach.

246 Therefore, we have also implemented in PURGd this simulation method, which estimates

247 ancestral inbreeding as described by Suwanlee et al (2007) using 10^6 replicates. Results for

248 all the ancestral inbreeding models considered were obtained using F_a calculated both from

249 Equation 2 and from gene dropping. For consistency with our IP method and with

250 previously published F_a based analysis, in the main text we report results obtained using

251 Equation 2, and those obtained using gene dropping are shown in the Supplementary

252 Material.

253   To fit the joint effect of inbreeding and purging on fitness, Ballou proposed the

254 following linear model

255 $$W = W_0 + b_F\, F + b_{FFa}\, F.\, F_a \,,$$

256 where b_F is the partial regression coefficient that gives the decline of fitness with increasing

257 inbreeding (F) for any constant value of the product $F.F_a$. According to Ballou, $-b_F$

258 represents the rate of inbreeding depression, while the coefficient b_{FFa} measures the

259 increase of fitness in inbred individuals due to reduced inbreeding depression caused by

260 purging in their ancestors.

261   Since we use a multiplicative fitness model, we rewrite Ballou's model for individual

262 fitness as

263 $$W_i = W_0 \cdot e^{b_F \cdot F_i + b_{FFa} F_i \cdot F_{ai}} \qquad\qquad (3)$$

264   Two additional linear models have been proposed by Boakes and Wang (2005) to

265 analyze purging using ancestral inbreeding. One of these two models (BW) considers that

266 the effect of purging does not depend on the level of inbreeding, but just on previous

267 purging opportunities. For multiplicative fitness, this model is written as

268 $$W_i = W_0 \cdot e^{b_F \cdot F_i + b_{F_a} F_{ai}} \,, \qquad\qquad (4)$$

269 where the coefficient of the purging term b_{F_a} is the average rate of increase of individual

270 fitness due to the opportunities of purging in the ancestors.

271   The other model proposed by Boakes and Wang (2005) is the mixed "Ballou-Boakes &

272 Wang" model (here B-BW), where the purging term is the sum of those in Ballou and BW

273 models, giving

274 $$W_i = W_0 \cdot e^{b_F \cdot F_i + b_{F_a} F_{ai} + b_{FFa} F_i \cdot F_{ai}}. \qquad\qquad (5)$$

275   Fitness evaluation is often dichotomous by nature (e.g., dead/alive individuals), and

276 both Ballou (1997) and Boakes and Wang (2005) tested their models by fitting dichotomous

11

277   (0, 1) fitness data using logistic regression. To check which is the better approach to handle

278   such data, we generate dichotomous fitness values and analyze them using Ballou's model,

279   with both the NNLR and the Logistic methods (Figure S2; Tables S1 and S2). However, to

280   compare ancestral inbreeding and IP approaches under similarly optimal conditions, in the

281   main text we always report results of NNLR analysis of fitness data simulated as a

282   continuous variable defined in the interval (0, 1). A bootstrap contrast analogous to that

283   performed for the IP analysis is used in each replicate to test the significance of purging in

284   Ballou's analysis (see Supplementary Material).

285   Non-Linear Regression coefficients for F_a-based models, as well as bootstrap

286   errors, are computed using PURGd 2.0. As in the case of the IP model, the intercept is

287   obtained in a previous step as the mean fitness for non-inbred individuals with non-

288   inbred ancestors ($F = F_a = 0$).

289

290

291   **Analysis of the predictive value of the estimates**

292   To evaluate the predictive value of the parameters estimated in the previous

293   section, we use the estimates obtained from different numbers of generations ($t=N/2$,

294    $t=N$, $t=2N$) in lines of different sizes ($N=10$, $N=25$, $N=50$) to predict the evolution of

295   average fitness for lines for each of the three sizes considered (crossed predictions).

296   We check how these predictions fit the corresponding simulated data by graphically

297   comparing the observed and predicted evolution of mean fitness.

298   In the case of the IP model, predictions of the expected fitness at generation t

299   (W_t) are computed using the equation for the evolution of mean fitness, obtained by

300   replacing W_i and g_i in Equation 1 with their expected values at generation t (W_t and

301   g_t). For this purpose, g_t is computed as a function of N using the expression provided

302   in García-Dorado (2012). The neutral prediction of the model by Morton *et al*. (1956)

12

303   is also obtained by replacing g_t with the standard inbreeding coefficient (F_t) into

304   Equation 1 and using the inbreeding load computed in the simulated population ($\delta =$

305   B_{SIM}).

306   In the case of models based on ancestral inbreeding, predictions for mean

307   fitness are obtained by replacing F_i and F_{ai} in Equations 3-5 with their expected

308   values through generations, F_t and F_{at}. Below we derive an expression for the

309   evolution of F_{at} through generations in a panmictic population maintained with

310   effective size N.

311   From Equation 2, assuming a monoecious population, or the same expected F_a

312   value (or F values) for sires and dams, the average ancestral inbreeding at generation

313   t can be computed by iterating the expression

314    $$F_{a\,t} = F_{a\,(t-1)} + \left[1 - F_{a\,(t-1)}\right] \cdot F_{(t-1)} ,$$

315   which, noting that $F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$ and rearranging, can be written as

316    $$F_{a\,t} = 1 - \left(1 - \frac{1}{2N}\right)^{t-1} \cdot \left[1 - F_{a\,(t-1)}\right] \quad . \qquad (6)$$

317   In addition, an expression directly giving the expected ancestral inbreeding after t

318   generations can be derived, so that it is not necessary to iterate expression 6 through

319   generations. For simplicity, we define $x_t = 1 - F_{a\,t}$ and $k = \left(1 - \frac{1}{2N}\right)$, so that Equation 6

320   can be written as $x_t = x_{t-1} \cdot k^{t-1}$. Therefore, since $x_0=1$, the expected value of x_t can be

321   computed as

322    $$x_t = x_o \prod_{i=0}^{t-1} k^i = k^{\sum_{i=0}^{t-1} t} = k^{\,t\,(t-1)/2}$$

323   and, replacing x_t and k into this expression and rearranging, we obtain

324    $$F_{a\,t} = 1 - \left(1 - \frac{1}{2N}\right)^{\frac{1}{2}t(t-1)}$$

325

326

327

RESULTS

IP estimates of the rate of inbreeding depression and the purging coefficient

The inbreeding loads in the simulated base populations (B_{SIM}=0.5828 ± 0.0144 for CAPTIVE; B_{SIM}=2.5370 ± 0.0460 for WILD) are close to their corresponding expectations for the MSD balance (B=0.6266 for CAPTIVE, B=2.5511 for WILD). The estimated rates of inbreeding depression (δ) are close to B for N=10, as usually assumed, but decline for larger sizes, being in good agreement with their expected values (δ_{FM}) when computed from short term data (t=N/2) (Table 2). The estimates of δ based on longer inbreeding periods become downwardly biased.

Estimates of d are large, indicating substantial purging (Tables 2 and S3). There is a trend for a reduction of d when estimated from longer inbreeding periods, which is associated with a parallel reduction in the estimate of δ. As expected, the estimates of this purging parameter are always larger in the WILD case than in the CAPTIVE one. In both cases, the estimates are very similar regardless of the number of runs averaged per replicate (results not shown). Thus, the estimates presented here were obtained from just one run, though more runs might be needed if additional environmental factors were included.

We have also estimated the purging coefficient by using the expected value of the rate of inbreeding depression (δ_{FM}) as a known δ value in PURGd (results shown in Tables 2 and S3). It is interesting to note that this alleviates the underestimation of d with increased number of analyzed generations, compared to the situations where both d and δ are jointly estimated from the data.

350

351

352

14

353 **Estimates of the coefficients in ancestral inbreeding models**

354 Tables 3 and S4-S5 show the estimates of non-linear regression coefficients for F_a-

355 based models. Similar results obtained using gene dropping are shown in the Supplementary

356 Material (Tables S6-S7). In both Ballou's and B-BW models, $-b_F$ estimates obtained from

357 short term data for different population sizes (N) are reasonably close to the expected rate of

358 inbreeding depression (δ_{FM}), although standard errors are larger than in the IP model.

359 However, Ballou's $-b_F$ estimates tend to increase when based on more generations of

360 inbreeding, leading to values well above δ_{FM} in the WILD case.

361 The estimates of the coefficients for terms including F_a are usually positive, indicating

362 purging, but vary depending on N and t in an unpredictable way, particularly for BW and B-

363 BW models where $-b_{Fa}$ can even be negative in some instances.

364 Figure 3 illustrates how different F_a-based models fit the data for lines of different

365 sizes, showing the observed evolution of fitness during $2N$ generations together with the

366 corresponding predictions computed using coefficients estimated from the same data (Figure

367 S3, obtained using gene dropping, gives similar results). BW model fits the data poorly,

368 showing a systematic overestimation of fitness during the first N generations and an

369 increasing underestimation later on, while Ballou's model fits remarkably well. B-BW

370 model does not improve fitting over Ballou's one, which is not surprising as b_{Fa} estimates

371 are usually small. Therefore, hereafter we will use Ballou's model to evaluate the predictive

372 value of F_a-based methods.

373

374

375 **The efficiency of IP and Ballou's models to detect purging**

376 Figure 4 gives the percent of replicates in which a model including purging fitted the

377 data significantly better than a non-purging model, both for IP and Ballou approaches

378 (Figure S4 with Ballou's results obtained using gene dropping gives similar results). For

15

379   both models, purging detection is more likely in larger lines and for larger inbreeding

380   periods, as expected from more efficient purging and larger sample sizes. Detection is also

381   more likely for the WILD than for the CAPTIVE case, as expected.

382   Under both IP and Ballou's models, the proportion of detected cases in the most

383   difficult situation (N=10, t=N/2, CAPTIVE) is very small, indicating that although both

384   approaches detect purging when estimates are averaged over replicates, they may not be able

385   to do so when small replicates are separately considered during short inbreeding periods. The

386   fact that, in that situation, the proportion of detected cases is smaller than 0.05 indicates that

387   the test is conservative. In more favorable situations, both IP and Ballou models give

388   substantial detection rates, usually somewhat larger for the former model.

389

390   **The reliability of predictions based on estimates using IP and Ballou's models**

391   One of the main aims of this work is to check whether each pair of IP parameters (δ,

392   d) estimated by PURGd from pedigree data for each (N, t) situation (Tables 2 and S3) is

393   reliable for predicting the evolution of fitness in lines of different sizes during periods of

394   considerable length (t up to 2N). Thus, Figure 5 gives, for each population size, the crossed

395   IP predictions computed using different (δ, d) estimates obtained from data corresponding

396   to different population sizes and inbreeding periods, together with the prediction

397   computed assuming no selection and using the inbreeding load of the base population (d=0;

398   $\delta = B_{SIM}$), and with the evolution of mean fitness observed in the simulated lines. IP

399   predictions remain quite accurate during the first N generations. In general, there is a slight

400   trend for long-term fitness being better predicted using (δ, d) estimates from long term data.

401   Furthermore, predictions computed using (δ, d) estimates obtained from small lines, where

402   purging is more likely to be overwhelmed by genetic drift, tend to underrate fitness for

403   larger lines. Conversely, IP predictions tend to overestimate fitness in the long term.

404   However, all these biases are usually small, with the exception of those for N=10 lines in the

405   WILD case.

406   In any case, despite the variability observed between the average (δ, d) estimates

407   obtained from different data sets (Tables 2 and S3), IP predictions remain quite accurate

408   and always fit the data much better than a model assuming no selection. The reason is that

409   the reductions in the estimate of δ obtained from longer inbreeding periods are compensated

410   by reductions in the corresponding estimate of d.

411   Figure 6 shows a similar evaluation for the reliability of Ballou's predictions computed

412   using estimates of the corresponding coefficients obtained from different data sets (Tables 3

413   and S4-S5). Figure S5 obtained using gene dropping estimates gives similar results.

414   Predictions obtained using parameters estimated in smaller lines underestimate long-term

415   fitness, while those obtained from larger lines tend to overestimate fitness in the medium-

416   term but can still underestimate fitness in the long term. Fitting also improves when

417   estimates are based on longer inbreeding periods and, of course, when the coefficients used

418   to obtain predictions had been estimated in the same data set for which fitting is tested. In

419   general, predictions are reliable during the first few generations, where purging is irrelevant,

420   but become unreliable later on. Thus, Ballou's predictions of mean fitness are highly

421   dependent on the conditions used to estimate the coefficients of the model, and become

422   very erratic after a few generations. The same analysis was performed for the BW model,

423   giving even less reliable predictions (data not shown).

424   Comparing Figures 5 and 6 shows that IP predictions are more accurate than those

425   of Ballou's F_a-based model, the IP model providing reasonable predictions of the

426   evolution of fitness for any of the population sizes considered using parameters

427   estimated under different conditions.

428

429

DISCUSSION

430

431   Using simulated pedigreed fitness data, we analyze the performance of the

432   Inbreeding-Purging model (IP) and of models based on ancestral inbreeding (F_a) in order

433   to: i) detect purging; ii) estimate genetic parameters that can be used to obtain reliable

434   predictions of the evolution of fitness under inbreeding and purging. The IP model is

435   based on the expected effect of selection against the recessive component of deleterious

436   effects (d) that is exposed in homozygotes due to inbreeding, while the F_a approach is

437   based on the statistical fitting of models including inbreeding (F) and ancestral

438   inbreeding (F_a) terms. To estimate the parameters of these models we have used an

439   updated version of the PURGd software (García-Dorado *et al*., 2016)

440

The statistical estimation approaches

441

442   We have discussed in a previous paper (García-Dorado *et. al.*, 2016) the advantages

443   of the NNLR approach compared to linear regression for log-fitness data (LR), and the

444   analysis of the data presented here confirm those advantages (results not shown).

445   Furthermore, here we compare the performance of our NNLR method with that of the

446   logistic regression approach previously used in the literature to analyze purging for

447   dichotomous data, as those from dead/alive records, (Ballou, 1997; Boakes *et al*., 2007;

448   Ceballos and Álvarez, 2013; Kennedy *et al*., 2014). To do so, we have estimated the

449   parameters of Ballou's model using both approaches for simulated binary fitness data,

450   and we find that the NNLR estimates fit these data as well or slightly better than the

451   logistic ones (Figure S2). Therefore, since the NNLR analysis relies on a model that is

452   consistent with our exponential IP model and has other advantages regarding the

453   estimation of δ, as discussed in García-Dorado (2016), we encourage its use to analyze

454   binary fitness data. Hereafter, we discuss the properties of both IP and F_a models using

455   NNLR estimates obtained from untransformed continuous fitness data.

The mutational models

In order to explore the consequences of purging against the inbreeding load expressed in wild or captive populations, we analyze fitness under two mutational models. The CAPTIVE mutational model corresponds to model II in Pérez-Figueroa *et al*. (2009). This model accounts for the properties of deleterious effects detected in *Drosophila* mutation accumulation experiments, but uses a larger deleterious mutation rate and higher kurtosis to roughly account for the additional rate of mutations that behave as deleterious in molecular evolutionary studies but whose effect is too small to be detected in mutation accumulation experiments (García-Dorado and Caballero, 2000; Ávila and García-Dorado, 2002; García-Dorado *et al*., 2004; Halligan and Keightley, 2009). WILD mutational parameters were obtained by doubling the average deleterious effect and the deleterious mutation rate of the CAPTIVE case to approximately account for the about fourfold inbreeding load expressed in competitive or wild conditions (Ralls *et al*., 1988; O'Grady *et al*., 2006; Yun and Agrawal, 2014; Hedrick and García-Dorado, 2016). Our estimates of the purging coefficient *d* in the CAPTIVE case are larger than those estimated in non-competitive conditions for *Drosophila* (Bersabé and García-Dorado, 2013), but the estimates obtained in the WILD case are similar to those experimentally obtained in competitive conditions (López-Cortegano *et al*., 2016). We find that our CAPTIVE and WILD cases parallel the non-competitive and competitive conditions of those experiments, as the WILD case gives a larger inbreeding load but also a larger purging coefficient than the CAPTIVE one so that, under slow inbreeding, long term inbreeding depression is small in both instances.

Performance of IP and Fa models

The IP estimates of δ obtained using early data of the inbreeding process are in good agreement with their expected value (up to $t=N$ generations in the CAPTIVE case or $t=N/2$

19

in the WILD case; see Table 2). However, they become downwardly biased when based on full data from a long inbreeding process, which is associated with a reduction of the estimates of d. The reason is that, for $t = 2N$, most purging occurs during a small proportion of the period considered and, since the model's predictions are not exact, estimates smaller than the true δ and d values can lead to some overfitting of long-term data. More stable estimates of d were obtained by introducing into the model the expected rate of inbreeding depression (δ_{FM}) as a known δ value. In practice, δ_{FM} is unknown, but δ can be estimated in a previous step by analyzing data of early generations, or by assuming $d=0$ and using fitness data from individuals with no ancestral inbreeding ($F_a=0$; an option incorporated in PURGd 2.0). This δ estimate can then be introduced into PURGd as a known δ value to obtain more stable estimates of d.

A main finding is that, despite the bias for δ and d described above, each joint (δ, d) IP estimate, whether obtained from small or large lines or based on short-term data or on the full long inbreeding process, produces good predictions for the evolution of mean fitness over the whole range of population sizes and during the whole period of inbreeding considered (Figure 5). An exception is that of the smaller lines ($N=10$) for the WILD case, where the observed inbreeding depression is larger than the IP prediction, unless (δ, d) were also estimated from the same data ($N=10$ lines). Furthermore, (δ, d) estimates obtained from $N=10$ lines predict too small fitness in the medium term for larger lines. The reason is that IP is a deterministic model that predicts the consequences of natural selection on homozygous genotypes induced by inbreeding, but does not account for the reduction in the efficiency of natural selection caused by random drift. In fact, it has been found that drift roughly overwhelms purging for $Nd < 1$ (García-Dorado, 2012), so that alleles with $d < 0.1$ should be hardly purged in lines with $N = 10$. In the WILD case, the number of mutations per gamete with an effect small enough to escape purging for $N=10$ is larger than in the CAPTIVE one (see Figure 2 and note that $Ns<2$ implies $Nd<1$ for $h<1$). In fact, the class

20

508   with $d < 0.1$ contributes twice inbreeding load in the WILD than in the CAPTIVE case (0.36

509   vs. 0.18). Thus, in the WILD case, the IP model is less reliable for the smaller lines.

510   Remarkably, even in this $N=10$ case, IP predictions are much more accurate than those

511   computed by ignoring purging.

512   It should be noted that IP predictions (as well as F_a-based ones) do not account for the

513   fitness decline caused by the continuous accumulation of newly arisen mutations. Therefore,

514   they tend to overestimate long-term fitness in small lines where natural selection against the

515   accumulation of new deleterious mutations is relatively inefficient. This bias, although can

516   be corrected in theoretical situations (see the Full Model approach in García-Dorado, 2012),

517   is unknown in practice. In our data, this mutational fitness decline is small for the periods

518   considered, although it could be threatening for very small lines in the long term (García-

519   Dorado *et al.*, 1999; Ávila and García-Dorado, 2002; Caballero *et al.*, 2002; García-Dorado,

520   2003; Halligan and Keightley, 2009).

521   In addition to the IP model, we used three different models to estimate the dependence

522   of individual fitness on F and F_a, where the latter parameter (the ancestral inbreeding) is

523   used as an indirect measure of the purging opportunities in the individual's ancestors. For

524   the three models, we have obtained results using F_a estimates computed using the original

525   Ballou's equation (Equation 2) or the gene dropping simulation approach suggested by

526   Suwanlee *et al.* (2007). We found that, Ballou's original formula produces some upward

527   bias in the estimates of F_a, but the ability of F_a models to detect purging and predict its

528   consequences are very similar regardless how F_a was computed.

529   According to Ballou (1997), when F_a is included into the model, the regression

530   coefficient of fitness on F gives the rate of inbreeding depression ($\delta = -b_F$). This is obviously

531   true for the particular case of $F_a=0$, where b_F estimates the rate of inbreeding depression for

532   fitness in non-purged individuals. However, the meaning of b_F is less clear for $F_a > 0$ since,

533   as shown in the IP approach, the dependence of fitness on F among purged individuals

21

534   varies according to how fast inbreeding has been produced and, therefore, it also depends on

535   F_a. This explains why $-b_F$ is a poor estimator of the expected rate of inbreeding depression

536   unless it is based on early inbreeding periods, otherwise showing important bias of different

537   sign depending on the model used.

538   In Ballou's model, purging is measured by the coefficient corresponding to the

539   interaction effect (b_{FFa}). Thus, this model considers that the role of purging is to reduce

540   inbreeding depression, so that it only affects inbred individuals. Therefore, b_{FFa}

541   measures the rate of reduction of inbreeding depression with increasing F_a. Due to this

542   interaction term, this model has a common feature with the IP approach: the effect of

543   purging increases when inbreeding accumulates, both models predicting an initial

544   fitness decline that is later reversed to some extent, in agreement with the pattern

545   observed in simulated lines.

546   On the contrary, in the BW model purging is measured by the coefficient b_{Fa},

547   which represents the rate of increase in fitness with increasing F_a, averaged over all F

548   values (including individuals with F=0), and does not account for the reversal of the

549   initial depression. Boakes and Wang (2005) found that this BW model was more

550   efficient detecting purging in mutational models with mildly deleterious alleles,

551   probably because those models involved high mutation rates implying larger expressed

552   load in non-inbred individuals, and because those authors detected purging measuring

553   its consequences on the overall load of deleterious alleles per individual. On the

554   contrary, we evaluate the ability of the models to detect the reduction in inbreeding

555   depression, so that Ballou's model is more appropriate than BW. Regarding the B-BW

556   model, it did not outperform Ballou's nor BW models in Boakes & Wang study (2005),

557   nor in the present analysis.

558   Therefore, we consider the performance of Ballou's F_a-based model to detect and

559   predict the consequences of purging on inbreeding depression, and we compare it to

560 that of the IP model. The estimates of the interaction term in Ballou's model (b_{FFa}) are

561 very dependent on both the size of the lines and the number of generations of

562 inbreeding considered. Furthermore, for each population size N, different pairs of joint

563 estimates (b_F, b_{FFa}) produce different predictions for the evolution of fitness, which

564 compromises the reliability of Ballou's method. It is interesting to note that, as F_a

565 approaches 1, ($b_F \cdot F + b_{FFa} F \cdot F_a$) approaches ($b_F + b_{FFa}$)$F$. Thus, after the early fitness

566 recovery ascribed to purging, this method predicts a continuous rate of decline of fitness

567 with increasing F. Since such decline is not a general consequence expected from inbreeding

568 and purging, this prediction can be considered a flaw of the model. However, due to this

569 predicted decline, Ballou's model can spuriously fit the medium-term fitness decline

570 ascribed to the fraction of the inbreeding load caused by deleterious alleles that are not being

571 successfully purged (those with $Nd < 1$), or to the continuous fixation of new deleterious

572 mutations. Overall, due to the erratic nature of Ballou's model predictions, ascribed to the

573 inconsistency of the corresponding estimates, the IP model should be preferred to estimate

574 parameters that can be useful to predict the evolution of fitness under inbreeding.

575 Finally, according to our conservative bootstrap results, the probability of

576 detecting purging in each replicate is higher for IP than for Ballou's analysis and

577 increases for larger lines and longer inbreeding processes. Thus, in the WILD case, at

578 least 20 generations are required to have a good probability (p>0.8) of detecting

579 purging with effective population size 10 or above, while about 10 generations gives a

580 modest detection rate (about 30% for N=10 and 50% for N=25). In the CAPTIVE case,

581 detection chances using data of about 20 generations of inbreeding are modest, unless

582 the effective size is about 50 or larger. Thus, purging can pass undetected because

583 inbreeding is too fast for enough purging to occur, or because, being slow, is tracked

584 for a too short period. In practice, detection rates are likely to be smaller due to the

585 noisy nature of fitness measurements (particularly for binary data), to population

23

586   management partially relaxing fitness, or to concurrent adaptive processes or

587   undetected environmental trends (García-Dorado, 2015; Hedrick and García-Dorado,

588   2016; López-Cortegano et al., 2016).

589   Thus, our results encourage the use of the IP approach to analyze and predict purging,

590   showing that: i) δ estimates are more reliable when based on short periods of inbreeding, so

591   that only small purging has occurred (or on individuals with no ancestral inbreeding); ii)

592   purging is better detected from long inbreeding processes and under slow inbreeding; iii) the

593   estimate of the purging coefficient d is less biased when based on short-term inbreeding, but

594   more reliable estimates can be obtained from longer processes by using a good estimate of δ

595   as a known parameter; iv) joint (δ, d) estimates, even if they are downwardly biased in some

596   cases, usually produce reliable IP predictions for the evolution of mean fitness under

597   inbreeding, unless inbreeding is too fast. We also find that purging detection and

598   measurement are very demanding, which can explain why many analyses have failed to

599   detect purging in individual data sets (Ballou, 1997; Bryant et al., 1999; Byers and Waller,

600   1999; Crnokrak and Barrett, 2002; Boakes et al., 2006; Kennedy et al., 2014). Genomic

601   information can contribute to obtain large samples of data useful to detect and measure

602   inbreeding depression (Kardos et al., 2016; Wang, 2016). Unfortunately, inferring

603   purging using genomic based estimates of inbreeding is not straightforward because

604   the historical information about how present inbreeding has been produced is less

605   explicit in genomic data than in a pedigree. Although it should be possible to infer this

606   historical information from analysis based on the length of the segments that are

607   identical by descent (Keller et al., 2011; Speed and Balding, 2015), no method has so

608   far been developed to obtain estimates of a predictive purging parameter from such

609   data. Another possibility is pedigree reconstruction based on massive molecular

610   markers (Fernández and Toro, 2006; Wang, 2011; Wang et al., 2012; Jiménez-Mena et al.,

611   2016). However, in both instances, detection possibilities may be poor if fitness records are

612 available just for individuals of the present generation. In any case, our understanding of

613 purging can be expected to improve in the future through the accumulation of IP

614 analysis of different sets of available pedigreed data.

615

620

621 **Data Archiving**: An updated version of the PURGd software (PURGd 2.0) published by

622 García-Dorado *et al*. (2016) is available from https://www.ucm.es/genetica1/mecanismos.

623 **References**

624 Ávila V, García-Dorado A (2002). The effects of spontaneous mutation on competitive

625   fitness in *Drosophila melanogaster*. *J Evol Biol* **15:** 561–566.

626 Ballou JD (1997). Ancestral inbreeding only minimally affects inbreeding depression in

627   mammalian populations. *J Hered* **88:** 169–178.

628 Bersabé D, Caballero A, Pérez-Figueroa A, García-Dorado A (2016). On the consequences

629   of purging and linkage on fitness and genetic diversity. *G3 (Bethesda)* **6:** 171–181.

630 Bersabé D, García-Dorado A (2013). On the genetic parameter determining the efficiency of

631   purging: an estimate for *Drosophila* egg-to-pupae viability. *J Evol Biol* **26:** 375–385.

632 Boakes E, Wang J (2005). A simulation study on detecting purging of inbreeding depression

633   in captive populations. *Genet Res* **86:** 139–148.

634 Boakes EH, Wang J, Amos W (2007). An investigation of inbreeding depression and

635   purging in captive pedigreed populations. *Heredity* **98:** 172–182.

636 Bryant EH, Backus VL, Clark ME, Reed DH (1999). Experimental tests of captive breeding

637   for endangered species. *Conserv Biol* **13:** 1487–1496.

638 Byers DL, Waller DM (1999). Do plant populations purge their genetic load? Effects of

639   population size and mating history on inbreeding depression. *Annu Rev Ecol Syst* **30:**

640   479–513.

641 Caballero A, Cusi E, García C, García-Dorado A (2002). Accumulation of deleterious

642   mutations: additional *Drosophila melanogaster* estimates and a simulation of the effects

643   of selection. *Evolution* **56:** 1150–1159.

644 Ceballos FC, Álvarez G (2013). Royal dynasties as human inbreeding laboratories: the

645   Habsburgs. *Heredity* **111:** 114–121.

646 Charlesworth D, Willis JH (2009). The genetics of inbreeding depression. *Nat Rev Genet*

647   **10:** 783–796.

648 Crnokrak P, Barrett SCH (2002). Perspective: purging the genetic load: a review of the

649   experimental evidence. *Evolution* **56:** 2347–2358.

650 Crnokrak P, Roff DA (1999). Inbreeding depression in the wild. Heredity 83: 260–270.

651 Falconer DS, Mackay T (1996) Introduction to Quantitative Genetics, Ed. 4. Longman,

652   Essex, England.

653 Fernández J, Toro MA (2006). A new method to estimate relatedness from molecular

654   markers. *Mol Ecol* **15:** 1657–1667.

655 García-Dorado A (2003). Tolerant versus sensitive genomes: the impact of deleterious

656   mutation on fitness and conservation. *Conserv Genet* **4:** 311–324.

657 García-Dorado A (2007). Shortcut predictions for fitness properties at the mutation–

658   selection–drift balance and for its buildup after size reduction under different

659   management strategies. *Genetics* **176:** 983–997.

660 García-Dorado A (2012). Understanding and predicting the fitness decline of shrunk

661   populations: inbreeding, purging, mutation, and standard selection. *Genetics* **190:** 1461–

662   1476.

663 García-Dorado A (2015). On the consequences of ignoring purging on genetic
664   recommendations for minimum viable population rules. *Heredity* **115:** 185–187.

665 García-Dorado A, Caballero A (2000). On the average coefficient of dominance of
666   deleterious spontaneous mutations. *Genetics* **155:** 1991–2001.

667 García-Dorado A, López-Fanjul C, Caballero A (1999). Properties of spontaneous mutations
668   affecting quantitative traits. *Genet Res* **74:** 341–350.

669 García-Dorado A, López-Fanjul C, Caballero A (2004). Rates and effects of deleterious
670   mutations and their evolutionary consequences. In: Moya A, Font E (eds). *Evolution:*
671   *From Molecules to Ecosystems*. Oxford University Press: Oxford, UK, pp 20–32.

672 García-Dorado A, Wang J, López-Cortegano E (2016). Predictive model and software for
673   inbreeding-purging analysis of pedigreed populations. *G3 (Bethesda)* **6:** 3593–3601.

674 Gulisija D, Crow JF (2007). Inferring purging from pedigree data. *Evolution* **61:** 1043–1051.

675 Halligan DL, Keightley PD (2009). Spontaneous mutation accumulation studies in
676   evolutionary genetics. *Annu Rev Ecol Evol Syst* **40:** 151–172.

677 Hedrick PW, García-Dorado A (2016). Understanding inbreeding depression, purging, and
678   genetic rescue. *Trends Ecol Evol* **31:** 940–952.

679 Hedrick PW, Kalinowski ST (2000). Inbreeding depression in conservation biology. *Annu*
680   *Rev Ecol Syst* **31:** 139–162.

681 Jiménez-Mena B, Schad K, Hanna N, Lacy RC (2016). Pedigree analysis for the genetic
682   management of group-living species. *Ecol Evol* **6:** 3067–3078.

683 Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW (2016). Genomics advances the
684   study of inbreeding depression in the wild. *Evol Appl* **9:** 1205–1218.

685 Keller MC, Visscher PM, Goddard ME, Rosenberg NA (2011). Quantification of inbreeding
686   due to distant ancestors and its detection using dense single nucleotide polymorphism
687   data. *Genetics* **189:** 237–249.

27

688 Kennedy ES, Grueber CE, Duncan RP, Jamieson IG (2014). Severe inbreeding depression

689   and no evidence of purging in an extremely inbred wild species–The Chatham Island

690   black robin. *Evolution* **68:** 987–995.

691 Lande R (1994). Risk of population extinction from fixation of new deleterious mutations.

692   *Evolution* **48:** 1460–1469.

693 López-Cortegano E, Vilas A, Caballero A, García-Dorado A (2016). Estimation of genetic

694   purging under competitive conditions. *Evolution* **70:** 1856–1870.

695 Morton NE, Crow JF, Muller HJ (1956). An estimate of the mutational damage in man from

696   data on consanguineous marriages. *Proc Natl Acad Sci USA* **42:** 855–863.

697 O'Grady JJ, Brook BW, Reed DH, Ballou JD, Tonkyn DW, Frankham R (2006). Realistic

698   levels of inbreeding depression strongly affect extinction risk in wild populations. *Biol*

699   *Conserv* **133:** 42–51.

700 Pérez-Figueroa A, Caballero A, García-Dorado A, López-Fanjul C (2009). The action of

701   purifying selection, mutation and drift on fitness epistatic systems. *Genetics* **183:** 299–

702   313.

703 Ralls K, Ballou JD, Templeton A (1988). Estimates of lethal equivalents and the cost of

704   inbreeding in mammals. *Conserv Biol* **2:** 185–193.

705 Speed D, Balding DJ (2015). Relatedness in the post-genomic era: is it still useful? *Nat Rev*

706   *Genet* **16:** 33–44.

707 Suwanlee S, Baumung R, Sölkner J, Curik I (2007). Evaluation of ancestral inbreeding

708   coefficients: Ballou's formula versus gene dropping. *Conserv Genet* **8:** 489–495.

709 Wang J (2011). COANCESTRY: a program for simulating, estimating and analysing

710   relatedness and inbreeding coefficients. *Mol Ecol Resour* **11:** 141–145.

711 Wang J, El-Kassaby YA, Ritland K (2012). Estimating selfing rates from reconstructed

712   pedigrees using multilocus genotype data. *Mol Ecol* **21:** 100–116.

713   Wang J (2016). Pedigrees or markers: which are better in estimating relatedness and

714   inbreeding coefficient? *Theor Popul Biol* **107:** 4–13.

715   Yun L, Agrawal AF (2014). Variation in the strength of inbreeding depression across

716   environments: effects of stress and density dependence. *Evolution* **68:** 3599–3606.

717

Figure Legends

719   **Figure 1:** Evolution of the expected purged inbreeding coefficient (g) against

720   generation number for different d values, together with the evolution of Wright's inbreeding

721   coefficient (F) for populations of effective size 25 (left) or 100 (right).

722   **Figure 2:** The area below the lines gives the expected number of deleterious mutations

723   with homozygous effects within any interval in the abscissa axis. Dotted line: CAPTIVE

724   mutational model. Dashed line: WILD mutational model. Note that the figure does not

725   show probability density functions, as they do not integrate to 1 but to the mutation rate

726   λ.

727   **Figure 3:** Evolution of mean fitness in simulated lines (red) and the corresponding

728   predictions obtained using different F_a-based models. Predictions are computed for

729   two different cases, CAPTIVE and WILD, and three different population sizes (N=10,

730   N=25 and N=50) over t=2N generations using the coefficients estimated from the same

731   lines and number of generations. Three models based on ancestral inbreeding are used:

732   Ballou's (green), BW (yellow) and B-BW model (black dotted), as well as a prediction

733   without selection (grey).

734   **Figure 4:** Percent of replicates where a model including purging fitted the data

735   significantly better than a non-purging model under the IP or Ballou approaches, both for

736   CAPTIVE and WILD mutational models (bootstrap contrasts with α=0.05).

737   **Figure 5:** Observed fitness for the CAPTIVE (up) and WILD (down) cases, and the

738   corresponding prediction computed using the estimates obtained in the IP model. In each

739   panel, observed and predicted values over t=2N generations correspond to the

740   population size indicated in the column (N=10, N=25 and N=50). Different

741   predictions are plotted using estimates obtained from different data sets, denoted by

742   different colors and strokes as shown in the lateral panel. Neutral predictions,

743   computed assuming no selection and using the inbreeding load observed in the simulated

744   base population (B_{SIM}), are also shown.

745   **Figure 6:** Observed fitness for the CAPTIVE (up) and WILD (down) cases, and the

746   corresponding prediction computed using the estimates obtained in Ballou's model. In

747   each panel, observed and predicted values over $t=2N$ generations correspond to the

748   population size indicated in the column ($N=10$, $N=25$ and $N=50$), and different

749   predictions are plotted using estimates obtained from different data sets, denoted by

750   different colors and strokes as shown in the lateral panel. Neutral predictions,

751   computed assuming no selection and using the inbreeding load observed in the simulated

752   base population (B_{SIM}), are also shown.

753

	E(s)	E(h)	λ
CAPTIVE	0.1	0.337	0.1
WILD	0.2	0.283	0.2

Table 1: Genetic parameters used in simulations for the two different cases (CAPTIVE and WILD): Expected (E) values of the selection coefficient (s, gamma distributed with shape parameter 1/3) and of the degree of dominance (h, uniformly distributed between 0 and $e^{-7.5s}$), and mutation rate (λ).

		CAPTIVE				WILD			
		δ_{FM}	δ	d	$d_{(\delta_{FM})}$	δ_{FM}	δ	d	$d_{(\delta_{FM})}$
N=10	t=N/2		0.5667 ± 0.0185	0.2572 ± 0.0136	0. 2856 ± 0.0144		2.2899 ± 0.0541	0.3233 ± 0.0131	0.3476 ± 0.0130
		0.5540				2.2846			
	t=2N		0.4955 ± 0.0149	0.1981 ± 0.0099	0. 2492 ± 0.0103		1.8043 ± 0.0392	0.2196 ± 0.0076	0.3015 ± 0.0082
N=50	t=N/2		0.5004 ± 0.0266	0.2915 ± 0.0247	0. 2781 ± 0.0281		1.8686 ± 0.0626	0.3954 ± 0.0159	0.4036 ± 0.0152
		0.4448				1.8861			
	t=2N		0.3745 0.0195	0.1499 0.0199	0. 1958 ± 0.0201		1.4010 ± 0.0632	0.2539 ± 0.0218	0.3389 ± 0.0177

Table 2. Estimates of rates of inbreeding depression and purging coefficients from lines of different sizes (*N*) and different numbers of generations (t). Estimates are averaged over replicates, and are given with their empirical standard errors. This table gives the expected rate of inbreeding depression, computed as δ_{FM} (see text for explanation) together with the corresponding PURGd estimates. It also gives the corresponding estimates of the purging coefficient (*d*). An estimate of *d* is also obtained by forcing PURGd to use δ_{FM} as the known rate of inbreeding depression ($d(\delta_{FM})$). Extended results are reported in the Supplementary Material including results from *N*=25 lines and t=*N* (Table S3)

			Ballou		BW		B-BW		
CAPTIVE			b_F	b_{FF_a}	b_F	b_{F_a}	b_F	b_{FF_a}	b_{F_a}
N=10	t=N/2		-0.5529	0.1529	-0.5396	0.0410	-0.5556	0.0562	0.0325
			± 0.0217	± 0.0842	± 0.0185	± 0.0119	± 0.0219	± 0.1167	± 0.0159
	t=2N		-0.6247	0.4040	-0.3536	0.0565	-0.6163	0.3921	0.0010
			± 0.0214	± 0.0222	± 0.0113	± 0.0064	± 0.0212	± 0.0250	± 0.0073
N=50	t=N/2		-0.5506	0.3265	-0.4965	0.0504	-0.6096	0.2434	0.0319
			± 0.0361	± 0.0434	± 0.0206	± 0.0057	± 0.0392	± 0.0569	± 0.0080
	t=2N		-0.7228	0.6377	-0.0575	-0.0176	-0.6363	0.5961	-0.0222
			± 0.0515	± 0.0523	± 0.0105	± 0.0054	± 0.0608	± 0.0595	± 0.0077
WILD			b_F	b_{FF_a}	b_F	b_{F_a}	b_F	b_{FF_a}	b_{F_a}
N=10	t=N/2		-2.4140	2.0244	-2.2974	0.3210	-2.4481	1.3741	0.1763
			± 0.0657	± 0.2515	± 0.0581	± 0.0438	± 0.0678	± 0.2922	± 0.0531
	t=2N		-2.5070	1.9002	-1.2819	0.3079	-2.5667	1.8801	0.0465
			± 0.0663	± 0.0648	± 0.0301	± 0.0154	± 0.0637	± 0.0702	± 0.0232
N=50	t=N/2		-2.1444	1.6447	-1.9312	0.2697	-2.5151	1.1384	0.1994
			± 0.0805	± 0.0899	± 0.0525	± 0.0133	± 0.0929	± 0.1004	± 0.0184
	t=2N		-2.6496	2.4997	-0.0908	-0.0448	-2.4896	2.4214	-0.0421
			± 0.1065	± 0.1066	± 0.0144	± 0.0089	± 0.1217	± 0.1323	± 0.0128

Table 3. Non-linear regression coefficients estimated for Ballou's model, BW model and B-BW model in pedigrees of different populations sizes (N=10 and N=50) and numbers of generations ($t = N/2$ and $t = 2N$). Estimates are averaged over replicates, and are given with their empirical standard errors. Extended results are reported in the Supplementary Material including results from N=25 lines and t=N/2 (Tables S4 and S5).

N=25

N=100

F

g(d=0.05)

g(d=0.1)

g(d=0.5)

Generations

Generations

Homozygous effect